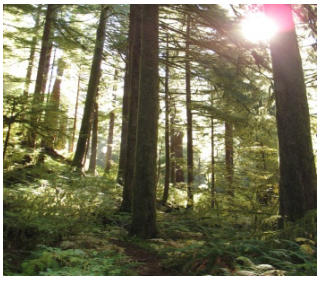
United States Environmental Protection Agency

Office of Policy (1807T)

November 2012
EPA-100-K-X12-009



Measuring the Effects of EPA Compliance Assistance in the Auto Body Sector: A Statistically Valid Pilot Project

Final Report

Promoting Environmental Results Through Evaluation

ACKNOWLEDGEMENTS

This evaluation was developed for the U.S. Environmental Protection Agency's Evaluation Support Division by Industrial Economics, Inc. (IEc) of Cambridge, MA and its subcontractors. The authors of the report included Tracy Dyke-Redmond of IEc, Christopher Leggett (a consultant to IEc) and Michael Crow (a consultant to IEc). A portion of the data presented in this report was collected by staff at Eastern Research Group (ERG), which was a subcontractor to IEc for this project.

The authors would like to thank EPA for the opportunity to participate in this unique evaluation of compliance assistance in the auto body sector. In addition, the authors would like to thank our colleagues at EPA Region 1 and ERG, as well as Charles Fabyonic, for their tireless efforts to collect data from the auto body shops. The authors also thank Michael Battaglia of Abt Associates and Joe Sedransk of Case Western Reserve University for their advice in developing the study design.

TABLE OF CONTENTS

LIST OF APPENDICES:

[1] Two additional materials, a brochure summarizing the surface coating rule requirements and an invitation to attend the workshop and webinars, were not available to include in this document.

EXECUTIVE SUMMARY

BACKGROUND

This report describes the results of a study designed to assess the impact of compliance assistance efforts offered by EPA Region 1 to the auto body sector, prior to the compliance date for a new EPA air regulation. EPA provided the compliance assistance in 2009-2010 through mailings, workshops, webinars, and site visits. The compliance assistance focused primarily on spray coating operations and hazardous waste storage by auto body shops. The study assessed the impacts of EPA compliance assistance in this sector using probability sampling, random assignment (i.e., to treatment and control groups), and on-site observations. The study also assessed the validity of gathering information on the impacts of compliance assistance through phone surveys.

The study grew out of a dialogue between the EPA and the Office of Management and Budget (OMB). In the years leading up to this study, OMB was concerned that EPA had not sufficiently addressed the problems of self-selection bias, non-response bias, and self-reporting bias in assessing the effects of the Agency's compliance assistance. In particular, OMB was concerned that EPA may not have been collecting representative, accurate information about the effectiveness of compliance assistance because the Agency had primarily relied upon (1) gauging effectiveness based on information from entities that voluntarily participated in compliance assistance and (2) collecting information about the effects of compliance assistance through telephone surveys.

The study included two evaluation approaches to assess the effectiveness of compliance assistance in influencing auto body shop behavior: a random assignment experiment focused on the short-term impact of compliance assistance outreach and workshops/webinars, and a quasi-experiment focused on the longer-term impact of a more comprehensive package of compliance assistance activities, including on-site assistance. In addition, the study assessed the validity of performance data obtained through telephone surveys.

The study gathered data on a set of 20 performance measures.[2] These measures related to the use of efficient spray-coating equipment, employee training on the use of spray coating equipment, proper maintenance of particulate filters, and proper hazardous waste

[2] The term "performance" represents facilities' environmental management behaviors; most of the performance measures are related to current regulatory requirements but a few performance measures are not required. The short-term random assignment experiment used the full set of 20 measures, and the long-term quasi-experiment used only 17 measures because not all performance measures were comparable across the comparison groups. The assessment of phone survey validity focused on the 13 measures for which data were collected as a part of the phone survey.

container management. The study also collected information on a number of key facility characteristics that were used to help interpret differences in performance among auto body shops.

Note that the findings for this evaluation are limited by the scope of the study. The study represents a picture of compliance for one window of time, in one sector, testing the impact of one package of compliance assistance. EPA did not intend, nor would it be appropriate, to use the results of the pilot study to draw conclusions about EPA's compliance assistance program as a whole. Moreover, the study does not measure any indirect effects of EPA assistance. For example, the study does not measure the impact of EPA training suppliers and trade associations, which in turn provided information to auto body shops. These and other limitations are discussed in the methodology chapter of the report.

FINDINGS

SHORT-TERM EFFECTIVENESS OF EPA COMPLIANCE ASSISTANCE

The random assignment experiment involved auto body facilities in areas of eastern and central Massachusetts with elevated air toxics risks. These auto body shops were assigned either (1) to a treatment group that was offered compliance assistance by EPA Region 1, or (2) to a control group that was not offered assistance by EPA.[3] The random assignment experiment compared the performance between the treatment and control groups for a single period of time. If the performance of facilities in the treatment group was higher than the performance of facilities in the control group and the difference was statistically significant, this would provide evidence that EPA's compliance assistance was effective in influencing the behavior of the auto body sector over the short term.[4]

However, the random assignment experiment does not provide evidence that EPA assistance to auto body shops affected sector-wide performance. A simple comparison of the groups' performance levels shows statistically significant differences for two performance measures, but the differences were too small to be of *practical* significance.[5] It would be difficult to detect any effects of compliance assistance for the performance measures on which control group performance was quite high.[6] For those measures, there was little room for the treatment group performance to exceed that of the control group. However, even for measures where the control group performance was not high (i.e.,

---

3 Note that the control group did eventually receive an offer of EPA Region 1 compliance assistance, but this offer occurred after the performance measurement for the random assignment experiment.

4 The phrase "higher performance" means that, for the group of performance measures studied, a greater percentage of facilities were observed to be following the performance measure. Thus, in the random assignment experiment, if a greater percentage of facilities in the treatment group was found to be following the performance measures, compared to the control group, this would provide evidence that EPA's compliance assistance was effective.

5 Specifically, the differences between treatment and control groups were less than five percentage points for these two measures where the study detected a statistically significant difference, and for both of these measures more than 95 percent of shops in the treatment and control group were in compliance.

6 For half of the measures studied, more than 90 percent of shops in the control group were in compliance with the measure.

where there was room for improvement), the analysis does not show any significant impact of EPA assistance.

Shops that chose to participate in workshops/webinars (15 percent of the treatment group in the sample) performed significantly better on five measures than the remaining shops in the treatment group that did not avail themselves of those opportunities. On these measures, participants' performance ranged from 6 to 44 percentage points better than that of non-participants. For example, with regard to properly labeling hazardous waste drums, shops that attended a workshop or webinar performed 33 percentage points better than shops that did not attend a workshop or webinar. However, these results only reflect the short-term impact of attending a workshop or webinar. Moreover, it is not possible to separate out the impact of the workshops/webinars relative to the potential effect of self-selection bias. For example, workshop participants may be systematically different from non-participants; their performance may have been superior even if they had not participated in the workshops/webinars.

LONG-TERM EFFECTIVENESS OF EPA COMPLIANCE ASSISTANCE

The quasi-experiment focused on two comparison groups: (1) auto body facilities in areas of eastern and central Massachusetts with elevated air toxics risks and (2) a similar population of auto body facilities in the Piedmont/Tidewater regions of Virginia. EPA Region 1 offered the auto body shops in Massachusetts a full suite of compliance assistance opportunities, while the auto body shops in Virginia did not receive an offer of assistance from EPA or the state of Virginia during the course of the study. The quasi-experiment assessed the impact of compliance assistance by comparing the *change* in performance in Massachusetts over a one year period with the *change* in performance in Virginia over the same time period. This is called a "difference-in-differences" methodology. If the performance the sample of Massachusetts shops improved more over time than performance of the sample of Virginia shops, and the difference-in-differences was statistically significant, this would provide evidence that EPA's compliance assistance was effective in influencing the behavior of the auto body sector over the long term.

The quasi-experiment suggests that overall impact of EPA assistance was minimal for the performance measures evaluated in the long-term experiment. After controlling for shop characteristics that could influence performance, three of the seventeen performance measures showed statistically significant, positive differences-in-differences, indicating a potential impact associated with compliance assistance for these measures. However, the seventeen performance measures were approximately evenly split between negative (larger improvements in Virginia) and positive (larger improvements in Massachusetts) difference-in-differences.

As expected, both Massachusetts and Virginia showed improvements in performance over time. Four out of 17 performance measures showed statistically significant improvements between 2010 and 2011 in Massachusetts, with differences ranging from 10 to 24 percentage points. Similarly, four out of seventeen performance measures

showed statistically significant improvements between 2010 and 2011 in Virginia, with differences ranging from 12 to 28 percentage points. However, the study does not provide strong evidence that the improvement was greater in Massachusetts, where EPA offered assistance.

TELEPHONE SURVEY VALIDITY

The study assessed the validity of data gathered through a telephone survey of auto body shops that were later visited by EPA and contractor personnel. Telephone survey respondents *and* non-respondents were included in the site visits. If there are few statistically significant differences between performance levels assessed using telephone survey data vs. performance levels based on data from on-site visits, this would provide evidence that telephone surveys provide valid data about performance and can be used to measure the impacts of compliance assistance.

This study finds that, while the phone survey results were similar to the site visit results for the majority of the performance measures examined, very large differences were observed for several performance measures. The differences in performance are statistically significant for five of 13 measures. For three of these measures, observed performance during site visits is better than expected based on phone surveys; for two of these measures observed performance during site visits is worse than expected based on phone surveys. The study finds that self-reporting bias was more of a concern than non-response bias. These findings are somewhat different than reported in the literature, and may merit further exploration to better understand the circumstances under which telephone survey results may be relatively reliable.

CONCLUSIONS

This study does not provide evidence that EPA assistance to auto body shops affected sector-wide performance in the short-term. While it appears that EPA assistance may have had a positive effect on sector-wide performance in the long-term for a few measures (3 out of 17 measures), the statistical evidence for an impact is not entirely compelling. Potential explanations for the absence of evidence are listed below, although the study does not demonstrate which, if any of these explanations are correct:

- The direct assistance provided by EPA simply may not have been effective in influencing the targeted population. It is possible that other approaches to providing information to auto body shops would be more effective, although the study does not suggest what, if any, changes to direct assistance should be made.
- The performance of auto body shops appears to have been positively influenced by vendors and suppliers, potentially dampening measurable impacts of EPA assistance provided directly to auto body shops. This study did not measure the indirect effects of information provided by EPA to vendors and suppliers, who in turn may used that information to assist shops. It is possible that the indirect

approach of influencing auto body shops by disseminating information through vendors and suppliers is more effective than direct assistance from EPA.

- Despite considerable outreach efforts by EPA Region 1, fewer than 20 percent of the shops in Massachusetts received interactive assistance during the study (i.e., workshops, webinars, or site visits). Thus, even if the interactive assistance was extremely effective for the shops that received it, the impact may be difficult to detect when this small group of shops is pooled with the remainder of the auto body population.
- For many of the performance measures evaluated, baseline performance was high, leaving little room for performance improvement. The auto body sector in Massachusetts had been exposed to considerable government assistance efforts over the last few decades, which may have limited the impact of additional assistance.

The study findings suggest that several measurement methods might be broadly useful and could be applied in future projects, including (1) obtaining representative data on baseline performance, (2) using phone surveys to assess baseline performance (though further study would be required to better understand the circumstances under which telephone survey results may be relatively reliable); and (3) delaying treatment (e.g., assistance) for a randomly assigned group of entities in order to establish a control group, and then providing treatment to these entities as needed after measurement is complete. However, sector characteristics will influence the transferability of these measurement approaches. For example, it is more difficult to draw statistically-based samples in sectors with a high turnover rate of businesses.

This study suggests a few implications for future compliance assistance efforts. In particular, EPA could consider focusing on outreach to suppliers to disseminate EPA's accurate compliance information

CHAPTER 1 | INTRODUCTION

This report describes the results of a three year pilot study designed to measure the impact of an EPA Region's compliance assistance efforts in the auto body sector. The study took place in eastern Massachusetts between 2009 and 2011, during which time EPA Region 1 had planned compliance assistance to help auto body shops comply with a new EPA air regulation. The study is unique in that it includes robust, quantitative measurement techniques to assess the impact of EPA assistance, which had not previously been attempted.

CONTEXT FOR PILOT MEASUREMENT

This project grew out of a dialogue between the EPA and the Office of Management and Budget (OMB). In the years leading up to this pilot project, OMB had recommended that EPA conduct more rigorous evaluations of the outcomes of its compliance assistance efforts. Compliance assistance typically includes outreach such as mailings and workshops, information posted on the internet, and assistance over the telephone and in site visits. OMB was concerned that in assessing the effects of this assistance, EPA was relying too heavily on information from entities that voluntarily participated in compliance assistance, and that these entities might not be representative of the larger audience EPA was trying to reach. Entities that volunteer to participate in compliance assistance (e.g., by attending workshops) may be more inclined to take action to comply than those entities that don't participate, and thus gathering information about the impact of workshops from voluntary participants may overstate the impact of compliance assistance. This phenomenon is called *self-selection bias*. In addition, EPA frequently relied on telephone surveys to gather information about environmental performance. However, OMB was concerned that entities who agreed to respond to phone surveys might not be representative of the broader population, and might be more likely to be in compliance than those that refused to answer a phone survey; this is termed *non-response bias*. Moreover, OMB noted that self-reported data might not be accurate, and in particular facilities might report over the phone that they were in compliance even if they were not; this is called *self-reporting bias*.[7]

[7] Self-selection bias and non-response bias can also be understood as threats to *external validity*. In other words, these biases limit the extent to which findings can be generalized to other contexts. Self-reporting bias can also be understood as a threat to *measurement validity*, i.e., whether the study is accurately measuring what it intends to measure. For a broader discussion of threats to validity in the context of program evaluation, see Hatry, H. P. and Newcomer, K. E., "Pitfalls of Evaluation," in the Handbook of Practical Program Evaluation, Second Edition, Wholey, J.S., Hatry, H. P., and Newcomer, K. E., eds. Josse-Bass, San Francisco, CA 2004.

In light of these concerns, EPA agreed to develop a statistically valid pilot project that would use representative sampling and a combination of phone surveys and site visits to measure the impact of EPA assistance in a selected sector, while also testing the validity of phone surveys as a data collection approach. The study was designed to correct for the three potential biases inherent in the ways EPA had evaluated its compliance assistance efforts to date (self-selection, non-response, and self-reporting bias). Ultimately, EPA intended that the project would lead to insights about measurement methods that the agency could use going forward.

EPA considered several compliance assistance efforts where it could test the statistically valid measurement approach. As noted earlier, EPA Region 1 was planning a compliance assistance effort in the auto body sector, and volunteered to particulate in the pilot project. EPA Headquarters and the Region agreed that this auto body assistance effort would be a reasonable area to test the measurement approach.

BACKGROUND ON AUTO BODY SHOPS AND APPLICABLE REGULATIONS

Auto body shops pose environmental concerns because of their prevalence, the nature of the materials they work with, and the level of training of their employees. Estimates of the number of auto body shops in the United States range from 35,000 to 80,000. It is common for auto body shops in urban areas to abut residential properties, schools, day care centers, elderly housing, and health clinics. Shops can often be found clustered in minority, immigrant, and/or low income neighborhoods. Fumes from spray painting and dust from sanding can pose risks to workers, neighbors, and the environment. Some of the chemicals used in auto body shop operations are highly toxic, including solvents with volatile organic compounds, paints containing diphenylmethane diisocyanate and toluene disocyanate, sanding dusts containing lead and chromium, and acetylene and metal fumes from welding operations. Despite the risks involved in auto body work, auto body shops are often small businesses with no specialized environmental staff. Without proper training, workers may improperly manage and dispose of chemicals and wastes, and may not take proper precautions to prevent air emissions.

In part due to the risks posed by auto body shops, as well as other businesses that conduct surface coating, EPA promulgated the Subpart HHHHHH National Emission Standards for Hazardous Air Pollutants: Paint Stripping and Miscellaneous Surface Coating Operations at Area Sources in January, 2008. This rule, also known as the Surface Coating Rule or 6H, regulates toxic air emissions from auto body shops and is meant to codify best practices already required by some states.[8] The rule requires that each affected operation must implement management practices to minimize the evaporative

[8] The 6H rule covers 1) paint stripping operations that use methylene chloride-containing paint stripping formulations; 2) spray-applied finishing or refinishing of motor vehicles and mobile equipment (trucks, construction equipment, self-propelled vehicles, and equipment that may be driven on a roadway); and 3) surface coating operations that involve spray-applied coatings that contain metal air toxic compounds to miscellaneous parts and products made of metal, plastic, or a combination of metal and plastic.

toxic emissions of their facility, including properly training staff.[9] Specific requirements for auto body shops outlined in the rule are shown in Exhibit 1-1.

[9] Environmental Protection Agency, FR Vol. 73, No. 6, Wednesday, January 9, 2008. 40 CFR Part 63: National Emission Standards for Hazardous Air Pollutants: Paint Stripping and Miscellaneous Surface Coating Operations at Area Sources. Final Rule. http://www.epa.gov/ttn/atw/area/fr09ja08.pdf.

EXHIBIT 1-1. SUMMARY OF REQUIREMENTS FOR AUTO BODY SHOPS

1) All spray painting must be done in a spray booth.

- Full cars must be painted in a spray booth with four walls, a roof and a ventilation system. (Filters in the booth have to remove at least 98 percent of the particulates.)
- Parts of cars must be painted in a booth with at least three walls or flaps, a roof and a ventilation system that pulls air into the spray booth.
- Spot repairs must be done in an enclosure which prevents any mist from getting out of the enclosure.

2) Painters must use spray guns and techniques which reduce overspray (such as high volume, low pressure, or HVLP, spray guns).

3) All painters must receive training. Owners must keep records of the training of each painter. (Specific training requirements are specified in the rule.)

4) Paint spray gun cleaning cannot create any mist of cleaning solvent to the air. Workers may spray solvent through the gun for cleaning purposes using an enclosed gun cleaner, or they may clean the gun manually.

5) All shops must also send a notification to EPA with some general information by January 2010:

- Location of facility
- Description of spray painting equipment
- Confirmation that shop has necessary equipment and training.

Shops must submit a Compliance Notification to EPA by March 2011 if they did not do so in their Initial Notification.

6) Exemptions to the rule are facility maintenance activities, which include the application of coatings to stationary structures or their appurtenances at the site of installation, to portable buildings at the site of installation, and to pavements and curbs.

Source: Brief Summary New EPA regulations for Auto body Refinishing Shops, 40 CFR Part 63 Subpart HHHHHH, August 2008, online at http://www.epa.gov/ttn/atw/area/autobodybs.doc.

Auto body shops in operation at the time the rule was promulgated were required to comply with the rule by January 2011. In advance of this compliance date, EPA Region 1 offered compliance assistance to auto body shops to help them prepare for the new requirements.

In addition to the new 6H rule, auto body shops are required to comply with other applicable Federal and state regulations. These include regulations governing hazardous waste under the Resource Conservation and Recovery Act (RCRA) and emergency planning under the Emergency Planning & Community Right to Know Act (EPCRA). RCRA requires proper identification, management, and disposal of hazardous waste, such as storing wastes in closed, labeled containers and maintaining record of proper shipment of wastes for disposal.[10] EPCRA requires shops to implement and document emergency procedures, such as posting the current name and telephone number of the emergency coordinator and the location of fire extinguishers and spill control material.[11]

EVALUATION PURPOSE, SCOPE, AND AUDIENCE

The pilot project was designed to accomplish three primary goals:

1) To implement an outcome measurement pilot for compliance assistance activities that uses statistically valid methods and will require the use of an OMB-approved Information Collection Request for data collection;

2) To test whether there is a significant positive correlation between compliance assistance activities and changes in behavior (i.e., improved environmental management practices and reduction/elimination/treatment of pollution), even after controlling for other predictive factors;

3) To assess the accuracy of self-reported environmental performance information obtained through telephone interviews; and

4) To develop a pilot project that has transferable elements so that future, regular activities may be measured using statistically-valid methodologies – but with less rigor than the pilot.

The scope of the pilot project was limited to test the effectiveness of a particular compliance assistance package (i.e., the *treatment*). This assistance included a set of four materials distributed by EPA Region 1: 1) a multimedia guidebook providing a summary of relevant regulatory requirements (e.g., those pertaining to air emissions and hazardous waste handling) impacting auto body shops, 2) a brochure summarizing the Surface Coating Rule requirements, 3) an invitation to attend workshops covering the requirements of the Surface Coating Rule, and 4) a copy of the presentation slides used at the workshops. (Appendix A contains copies of most of these materials.) For shops that opted to attend, the treatment also included participation in a workshop or webinar offered by EPA Region 1. Finally, the treatment included on-site compliance assistance for a randomly selected set of shops.

---

[10] For a more complete description of RCRA requirements, see EPA's website: http://www.epa.gov/compliance/civil/rcra/rcraenfreq.html.

[11] For a more complete description of EPCRA requirements, see EPA's website: http://www.epa.gov/oem/content/epcra/

The study was geographically limited to auto body shops in eastern and central Massachusetts, where Region 1 had a planned compliance assistance campaign, and a comparison group of auto body shops in Virginia, where no EPA assistance campaign was planned.

The primary audiences for the pilot project included EPA Headquarters, Region 1, and the Office of Management and Budget.

EVALUATION QUESTIONS

This evaluation was designed as an integral part of the pilot project, and was designed to answer four questions:

1) Did EPA Region 1's compliance assistance activities contribute to behavior change in the auto body sector?
2) Are the measurement methods employed in the pilot transferable to other assistance activities?
3) What specific characteristics of the auto body sector influence the transferability of the measurement approach in this evaluation?
4) Is the telephone survey a valid and reliable technique for performance measurement and program evaluation?

The next chapter of this report describes the study methodology in detail, and subsequent chapters describe the findings and conclusions of the pilot project.

CHAPTER 2 | METHODOLOGY

This chapter begins with a conceptual overview of the three components of the statistically valid pilot project. The chapter goes on to describe the performance measures and survey instruments and characterize the study populations in Massachusetts and Virginia. The chapter also provides details for each of the study components, including the sampling method, and the analytical approach. The chapter concludes with a summary of study limitations.

CONCEPTUAL OVERVIEW

The pilot project included two evaluation approaches to assess the effectiveness of compliance assistance in influencing auto body shop behavior: a random assignment experiment focused on the short-term impact of compliance assistance outreach and workshops/webinars, and a quasi-experiment focused on the longer-term impact of a more comprehensive package of compliance assistance activities, including on-site assistance. As a part of both of these evaluation approaches, EPA and contractor personnel gathered data about facility performance on key air and waste indicators during site visits at random samples of facilities.[12] These indicators related to the use of efficient spray-coating equipment, employee training on the use of spray coating equipment, proper maintenance of particulate filters, and proper hazardous waste container management.

In addition to the two evaluation approaches designed to assess the impact of EPA's compliance assistance, the pilot project was also designed to assess the validity of performance data obtained through telephone surveys. In particular, the pilot project assessed the validity of a data gathered through a telephone survey of auto body shops that were later visited by EPA and contractor personnel.

Exhibit 2-1 summarizes the three evaluation approaches incorporated in the pilot project design. The remainder of this section provides an overview of each part of the pilot project in turn.

[12] We use the term "performance" throughout this document to represent facilities' environmental management behaviors; some aspects of performance may be related to current regulatory requirements while other aspects of performance may be voluntary.

EXHIBIT 2-1. SUMMARY OF EVALUATION APPROACHES INCLUDED IN THE PILOT PROJECT

PURPOSE	TIME FRAME	EVALUATION APPROACH	COMPLIANCE ASSISTANCE OFFERED
Assess impact of EPA compliance assistance to auto body shops	Short-term	Random-Assignment Experiment	• Workshops/webinars offered to all facilities in the treatment group (attended by subset of facilities) (Interactive assistance) • Compliance assistance materials mailed to all facilities in the treatment group (Static assistance)
	Long-term	Quasi-Experiment	• *Same assistance as for short-term study, plus* • On-site assistance offered to randomly selected facilities in the treatment group (Interactive assistance)
Assess validity of telephone survey data	Short-term	Comparison of phone survey data with site visit data	• *Same assistance as for short-term study*

Part 1: Short-Term Impact of Compliance Assistance Outreach and Workshops

EPA measured the short-term impact of compliance assistance outreach and workshops through a random-assignment experiment involving auto body facilities in areas of eastern and central Massachusetts with elevated air toxics risks. All of these facilities were randomly assigned to either a treatment group (Group A) or a control group (Group B). The random assignment process ensured that the two groups were statistically equivalent with respect to observed and unobserved factors.

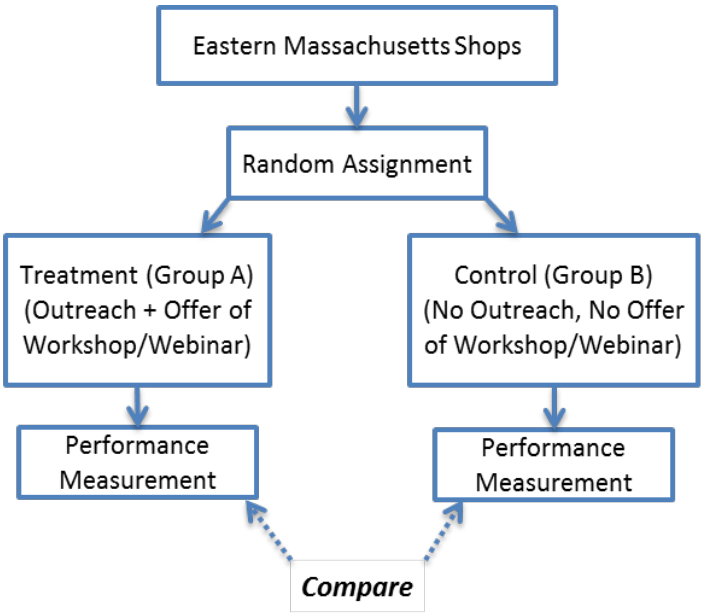
In October 2009, EPA Region 1 sent facilities in the treatment group a package of compliance assistance materials and an invitation to attend workshops and webinars covering existing and pending federal environmental regulatory requirements. (The mailed package of assistance is called "static" assistance in this report.) Between October 2009 and January 2010, 11 percent shops from the treatment group participated in either in a workshop or webinar. (The workshops and webinars are considered "interactive" assistance in this report.) The shops in the control group did not receive the mailings until after the completion of the short-term study.[13]

In spring and summer 2010, after the workshops/webinars had been completed, EPA and contractor personnel visited a random sample of facilities from each of the two groups to assess performance. The impact of compliance assistance was assessed by comparing the

[13] Just prior to the start of the pilot project, EPA Region 1 sent postcards to all auto body shops in Massachusetts notifying them of the surface coating rule and EPA's website, which provides web-based compliance tools. This postcard is not considered part of the treatment.

estimated performance on key indicators between the treatment and control groups (Groups A and B). The diagram in Exhibit 2-2 provides an overview of the short-term experiment.

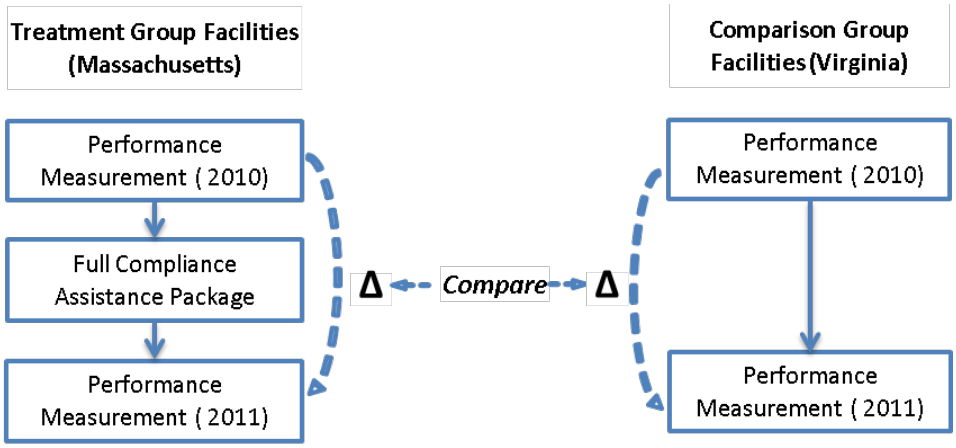
EXHIBIT 2-2. APPROACH TO ASSESSING SHORT-TERM IMPACT OF COMPLIANCE ASSISTANCE



Part 2: Long-Term Impact of Compliance Assistance Package

EPA measured the long-term impact of a more comprehensive compliance assistance through a quasi-experiment involving the full study population of Massachusetts auto body facilities and a similar population of auto body facilities in the Piedmont/Tidewater regions of Virginia. EPA Region 1 offered the facilities in Massachusetts a full suite of compliance assistance activities related to hazardous waste and surface coating requirements, including a static compliance assistance mailing, and interactive workshops, webinars, and (for a sample of facilities) on-site compliance assistance. The facilities in Virginia did not receive compliance assistance from EPA or the state. In each of the two groups, site visits by EPA and contractor personnel at independent random samples of facilities were used to estimate performance before and after compliance assistance was provided. The impact of compliance assistance was assessed primarily by comparing the *change* in performance in Massachusetts with the *change* in performance in Virginia. In other words, the study used a "difference-in-differences" approach to assess the impact of the compliance assistance. Exhibit 2-3, below, provides an overview of the long-term quasi-experiment.

EXHIBIT 2-3. APPROACH TO ASSESSING LONG-TERM IMPACT OF COMPLIANCE ASSISTANCE PACKAGE



Note: Independent random samples of facilities were drawn from each of the two populations (Massachusetts and Virginia) in 2010 and 2011 (i.e., the study did not use a panel design with repeat measurements on the same set of facilities).

Part 3: Assessing the Validity of Telephone Surveys

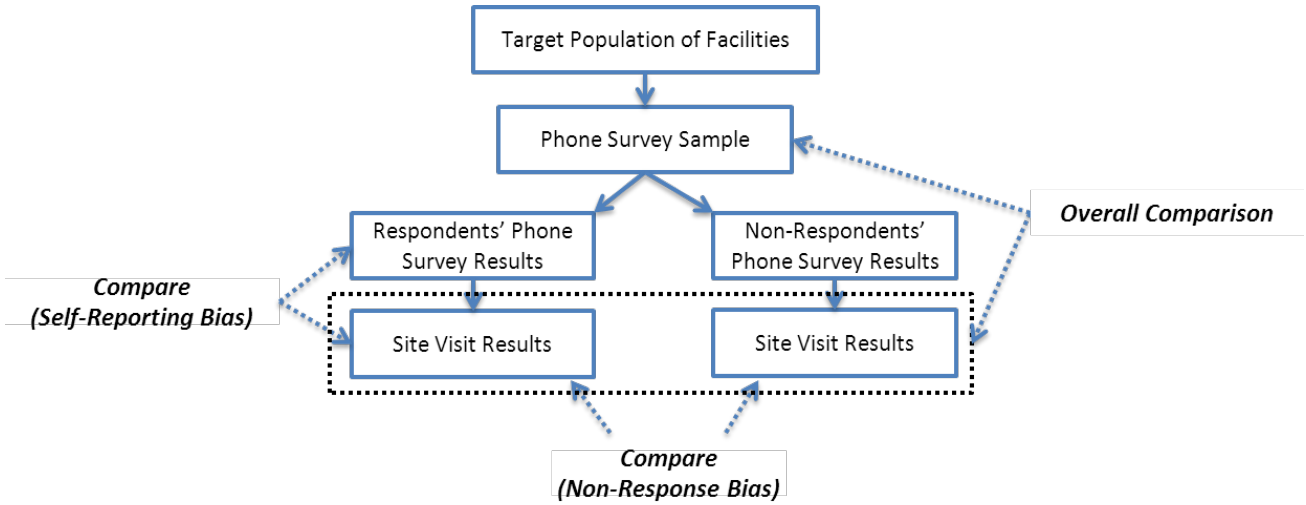
EPA assessed the validity of telephone survey responses using a two-phase sampling approach. In the first phase, EPA and contractor staff conducted telephone surveys at randomly selected samples of facilities in Massachusetts; specifically, the telephone surveys were conducted at samples of facilities drawn from Groups A and B described in the short-term study (see Part 1 above). Questions in the telephone survey were designed to determine facilities' behaviors with regard to key hazardous waste and air indicators.

After the phone surveys, a follow-up measurement verified the accuracy of the telephone surveys. Each sample of facilities that received a phone survey was divided into two sub-groups: 1) facilities that responded to the telephone survey ("respondents") and 2) facilities that did not respond to the telephone survey ("non-respondents"). Random samples were drawn from each of these subgroups, and site visits were conducted at the sampled facilities. The site visits determined facility performance through direct observation by EPA and contractor staff.

The telephone survey validity study was designed to assess two potential sources of bias in telephone survey data. The study assessed potential *self-reporting bias* (i.e., the potential bias associated with facilities reporting inaccurate information over the phone) by comparing site visit data to phone survey data for facilities that responded to the phone survey. The study assessed *non-response bias* (i.e., the potential bias associated with facilities that opted not to respond to the phone survey being systematically different than those that did respond) by comparing site visit data for facilities that responded to the

phone surveys vs. those that did not. The study assessed the overall bias associated with telephone survey data in the auto body sector by comparing overall performance levels estimated from site visit data to overall performance levels estimated telephone survey data. Exhibit 2-4 illustrates how the phone survey validity study worked in Group B. [14]

EXHIBIT 2-4. APPROACH TO ASSESSING PHONE SURVEY VALIDITY FOR GROUP B



PERFORMANCE MEASURES, CONTEXTUAL VARIABLES AND SURVEY QUESTIONNAIRES

In order to measure changes in facilities' environmental management behaviors, EPA gathered data on a set of objective performance measures. In addition, EPA gathered data on contextual variables that may be related to auto body shop performance, e.g. auto body shop characteristics and sources of information shops use to inform their environmental management. This section describes the data gathered as party of the study, and the data collection instruments (on site and phone survey questionnaires) used to gather the data.

Performance Measures

The pilot project analyzed 20 performance measures related to management of air emissions and hazardous waste. As shown in Exhibit 2-5, the short-term analysis used the full set of 20 variables. The long-term analysis used only 17 variables because differing requirements between Massachusetts and Virginia on three waste-related measures made comparison untenable. The phone survey validity analysis focused on the 13 of the 20 variables for which data were collected during the phone survey.

[14] Note that the two-phase sampling approach for Group A is slightly more complex due to stratification, as discussed later in this chapter.

The individual performance measures were derived from the survey questions for the purpose of the analysis. Interviewers verified shop performance on all selected variables through observations during the site visits. For example, interviewers observed the configuration of the shop's spray booth and recorded whether or not the booth was fully enclosed and properly ventilated.

Note that most performance measures related to compliance requirements at the federal, and sometimes state, level. However, two performance measures related to best management practices, and were not required. This report occasionally refers to the percentage of shops "in compliance" with performance measures for simplicity of text, rather than distinguishing between measures that are required and those that are best management practices. This language is intended to convey that the shops met the criteria of the performance measure.

The site visit and phone survey questions asked about a broader range of performance than what was ultimately included in the performance measures. Performance data from the on-site and phone surveys that were not included in the performance measures were excluded because they could not be sufficiently verified on-site, were later determined to be ambiguous in meaning, and/or the sample size was too small to allow for meaningful analysis. For example, variables related to hazardous waste determination and emergency procedures were not included in the performance measures. Summary statistics are provided for all variables in Appendix B.

EXHIBIT 2-5. PERFORMANCE MEASURES USED IN EACH ANALYSIS

MEDIUM	CATEGORY	PERFORMANCE MEASURE	ABBREVIATION	SHORT-TERM	LONG-TERM	PHONE ACCURACY
Air	Spray Booth	Booth exists	Booth_exists	•	•	•
		Spray only in booth	Not_outside	•	•	
		Fully enclosed	Booth_enclosed	•	•	•
		Ventilated with exhaust fan	Booth_ventilated	•	•	•
		Particle filter on exhaust	Filter_exists	•	•	•
		Filter in good condition	Filter_good	•	•	
		Capture efficiency for filter ≥ 98%	Capture98	•	•	•
	Prep Station	Enclosed (3 walls/curtains and roof)	Prep_enclosed	•	•	•
		Ventilated	Prep_vent	•	•	•
	Mixing Room	Enclosed (3 walls/curtains and roof)	Mixroom_enclosed	•	•	•
		Ventilated	Mixroom_vent	•	•	•
	Spray Guns	Only use HVLP/equivalent	Guns_compliant	•	•	•
		Compliant cleaning methods (non-atomized)	Cleaning_compliant	•	•	
		Records of all technicians properly trained	Train_records	•	•	•
	Paint Stripping	Avoid MeCl use*	Avoid_mecl	•	•	
Waste	Management	Used rags/towels stored in closed containers	Rags_closed	•	•	•
		No indication of spills in/near shop*	No_spills	•	•	
		All haz waste drums properly labeled	Drums_labeled	•		•
		All haz waste drums closed	Drums_closed	•		
		Haz waste shipping docs available	Waste_doc	•		

* Performance measures with an asterisk indicate “Best Management Practices” i.e., they are not required.

Contextual Variables

The surveys also collected information on a number of key facility characteristics in order to assist in interpreting the performance data. The data from these variables, and from survey metadata, are used in multivariate regressions and other qualitative and quantitative analytical approaches. Complete summary statistics for these variables can be found in Appendix B. The most important contextual variables for our analysis are shown in Exhibit 2-6, below. Explanatory variables used in the regression analysis are indicated with an asterisk. (The regression analyses are described later in this methodology section.)

EXHIBIT 2-6. KEY CONTEXTUAL VARIABLES

CATEGORY	VARIABLE
Shop Capacity	Number of painting jobs completed per week*
	Hazardous waste generator status (very small quantity generator, or larger)*
External Influences	Whether shop is part of corporate chain*
	Whether shop was recently visited by a non-EPA regulator*
	Timing of awareness of spray-coating regulations*
	Information providers[15] for spray-coating regulations or other regulatory issues*
Survey Metadata	Respondent type (owner, manager, technician, other)
	Interviewer

* Variable utilized in regression analysis.

Survey Instruments

EPA used an on-site survey instrument and a companion telephone survey instrument, provided in Appendices C and D, to gather data. The questions for the survey instruments were approved under the Paperwork Reduction Act as part of Information Collection Request (ICR) number 2344.01, provided in Appendix E. EPA designed the questions to obtain information on (1) environmental performance related to current hazardous waste management and training requirements under the Resource Conservation and Recovery Act (RCRA), (2) environmental performance related to air emissions control requirements associated with the recently promulgated Surface Coating Rule, (3) environmental compliance assistance received by government agencies or other entities, and (4) perceptions regarding the factors that influence shop behaviors related to

[15] Suppliers, consultants, local governments, state, EPA, etc.

environmental performance. EPA designed most of the questions to produce binary (i.e., yes/no) indicators of environmental performance for use as dependent variables in the statistical analysis.

The on-site survey form consisted of over 60 questions – many with multiple parts – and had two distinct sections: an interview component typically conducted in an office at the auto body facility, and a subsequent component conducted during a walk-through of the same facility. During the walk-through component, the interviewer would obtain information on environmental performance through his or her own observations and through targeted questions of shop personnel.

The telephone survey covered a subset of approximately 40 questions included in the on-site survey, making the telephone survey shorter in order to discourage hang-ups. The telephone survey focused mainly on environmental performance measures that could be later verified independently through interviewer observations on site.

Both questionnaires were reviewed by survey experts at Industrial Economics and Abt Associates, and by EPA experts in program evaluation, program review, statistics, and survey design. Both were also pretested on auto body shops in Boston, Massachusetts, which was not included in the sampling frame for the proposed survey. Several of the questions from the two survey modes are identical, so the pretest was limited to a total of nine shops across the two modes: the on-site survey was pretested on five shops, while the telephone instrument was pretested on four shops. The five shops used to pretest the on-site survey were selected from a list provided by the Boston Public Health Commission. The four shops used in the phone survey pretest were selected from a list derived from Dunn & Bradstreet and Reference USA (SIC 7532). The selected shops provided a range of operation sizes (from "mom-and-pop" shops to national chains) and locations within the city. After the pretest, the survey instruments and instructions were revised to address pretest observations regarding question wording, clarity of interviewer instructions, question flow, and survey length. No pilot tests were conducted for the survey.

TRAINING FOR DATA COLLECTORS

Prior to engaging in telephone surveys or site visits, EPA staff and contractors participated in detailed training to discuss the regulatory requirements and how they are applied in auto body shops, how to conduct site visits, and how to record survey data. Site visitor training in 2010 was most extensive. EPA Region 1 staff provided an in person, day-long training session for all site visitors, which included a field visit to a vocational technical school where site visitors were able to go through a "dry run" of the checklist as a group. Site visitors discussed all steps in the site visit process, from receiving the randomly assigned list of shops and planning a site visit schedule for each day, confirming shop locations, identifying the correct shop representative to interview, what to do if a shop location seemed unsafe or a shop was not in operation, and how to conduct the site visits. Site visitors discussed in detail potential differences in

interpretation in the survey questions, and guidance for interpreting each question was incorporated into the survey form itself. In addition to discussing the site visit process, IEc staff provided training for entering site visit data on a paper checklist, and then recording the data in an Access database. All survey question data from the paper checklists was then entered into a separate Access database by a separate data entry staff person, and records from the original and duplicate data entry databases were compared to ensure accuracy for all data entry.[16]

Site visitor training in 2011 was more streamlined, since nearly all site visitors in 2011 had been through the 2010 training. EPA provided a refresher webinar training. For those site visitors who were new to the project in 2011, colleagues who had participated in the 2010 training accompanied them on the first day of site visits in order to ensure they understood how to interpret the questions and enter the data.

IEc provided training to its own and EPA staff conducting telephone surveys in 2010. The training consisted of reviewing how to identify the right person to talk with, how to encourage survey participation, how to interpret questions, and how to enter the data.

Training materials and guidance documents for site visitors and telephone surveyors are included in Appendices C and D.

TARGET POPULATIONS AND SAMPLE FRAMES

The populations of interest for the pilot study are auto body shops subject to the Surface Coating Rule and located in areas with elevated air toxics risks in (1) eastern Massachusetts and (2) the Piedmont and Tidewater regions of Virginia. As discussed earlier, the Massachusetts shops provide a sample frame for the short- and long-term studies, as well as the evaluation of telephone survey validity, whereas the Virginia shops serve as the comparison group for the long-term study. The Virginia population was selected as the comparison group primarily because EPA and the state had no plans for compliance assistance or inspection activity related to the Surface Coating Rule or RCRA for this population, unlike many other parts of the country, and because the population had a sufficient number of shops located in areas with elevated air toxics risks.

For the purpose of this study, areas of elevated air toxics risks are those with elevated cancer and non-cancer risks from air pollution, according to National Air Toxics Assessment (NATA) data.[17] EPA chose to focus on areas with elevated air toxics risks for this study because the Agency expected that there would be greater need for auto body

[16] The only data that was not double entered was open text notes from the site visitor, since variations in spacing and punctuation would make double entry of this narrative data inefficient, and it was not necessary to ensure accuracy of the performance measure data.

[17] Elevated Risk - NATA data on levels of both cancer risk and non-cancer risk was broken into five classes using the Natural Breaks method - http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Natural_breaks_(Jenks). Towns that intersected any of the top four categories for both cancer risk and non-cancer risk were designated as elevated risk areas to be included in the population for this study. (Source: 1999 National Air Toxics Assessment data).

compliance assistance in these areas, given elevated risks to residents from other sources of air pollution.

Sample Frame

The sampling frame in both Massachusetts and Virginia included auto body shops that met the following criteria:

- **Defined as Automotive Body, Paint, and Interior Repair and Maintenance businesses** (North American Industry Classification System code 811121) and listed in either Dunn & Bradstreet or Reference USA.[18] This U.S. industry comprises establishments primarily engaged in repairing or customizing automotive vehicles, such as passenger cars, trucks, and vans, and all trailer bodies and interiors; and/or painting automotive vehicles and trailer bodies; and
- **Part of high-density clusters of shops within elevated-risk areas.** High-density clusters were identified by GIS staff in EPA Region I; most of these clusters were in urban centers but some were in abutting towns.

Auto body shops meeting these criteria were excluded if they were located in an area with regulatory and/or compliance assistance activity quite different from the norm in that part of the state. Specifically, in Massachusetts, auto body shops located in Lawrence and Boston were excluded from the sampling frame because for each of these communities had conducted intensive assistance, outreach, and/or enforcement activities for a number of years. Auto body shops in Worcester were excluded because intensive assistance was planned for the period of the pilot project. In Virginia, auto body shops located in Northern Virginia were excluded because that area of the state has stricter air quality regulations for auto body shops, and the Northern Virginia Regional Office of Virginia DEQ had recently initiated a compliance assistance and self-certification project directed at auto body shops in that area.[19]

Similarities and Differences: Massachusetts and Virginia

While the areas in Massachusetts and Virginia included in the pilot project were similar with regard to elevated air toxics levels and federal requirements, they were different with respect to the state environmental requirements in place prior to the pilot project. As shown in Exhibit 2-7, Massachusetts generally had more stringent requirements in place with regard to limiting air emissions from auto body shops, and with regard to waste management. These regulatory differences likely influenced auto body shop performance. While the pilot project study design made no direct comparisons between performance in Massachusetts and Virginia shops, the difference in regulatory

---

[18] This code replaced SIC code 7532 – "Automotive Body, Paint, and Interior Repair and Maintenance" – which was referred to in the ICR.

[19] This area includes the following localities: Arlington, Fairfax, Loudoun, Prince William, and Stafford counties, and Alexandria, Fairfax, Falls Church, Manassas, and Manassas Park cities.

requirements at the state level provides important context for interpreting the results of the long-term quasi-experiment.[20]

EXHIBIT 2-7. STATE REQUIREMENTS RELATED TO PERFORMANCE MEASURES

MEDIUM	CATEGORY	PERFORMANCE MEASURE	MA REQ?	VA REQ?
Air	Spray Booth	Booth exists	Y	
		Spray only in booth	Y	
		Fully enclosed		
		Ventilated with exhaust fan	Y	
		Particle filter on exhaust	Y	
		Filter in good condition	Y	
		Capture efficiency for filter ≥ 98%		
	Prep Station	Enclosed (3 walls/curtains and roof)		
		Ventilated	Y	
	Mixing Room	Enclosed (3 walls/curtains and roof)		
		Ventilated		
	Spray Guns	Only use HVLP/equivalent	Y	
		Compliant cleaning methods (non-atomized)	Y	
		Records of all technicians properly trained		
	Paint Stripping	Avoid MeCl use*		
Waste	Management	Used rags/towels stored in closed containers	Y	Y
		No indication of spills in/near shop*		
		All haz waste drums properly labeled	SQG**	
		All haz waste drums closed	Y	SQG**
		Haz waste shipping docs available	Y	SQG**

* Performance measures with an asterisk indicate "Best Management Practices" i.e., they are not required.

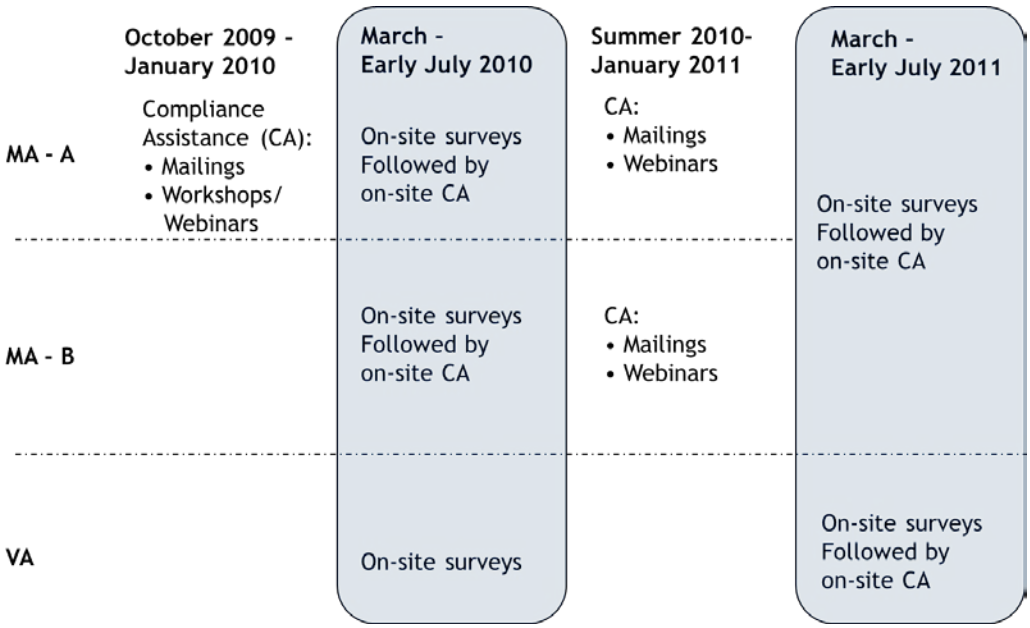
** SQG refers to Small Quantity Generators of Hazardous waste. Where SQG is shown, the requirement is for SQGs, but not for shops with lesser amounts of hazardous waste (e.g. Very Small Quantity Generators or Conditionally Exempt Small Quantity Generators)

[20] Rather than directly comparing performance levels in the two states, the long-term quasi-experiment compares the *change* in performance in Massachusetts (where EPA compliance assistance was provided) with the *change* in performance in Virginia (where no EPA compliance assistance was provided).

DETAILED DESCRIPTION OF STUDY DESIGN AND IMPLEMENTATION

This section describes in more detail each component of the pilot project study, including the treatment approach, sampling and measurement, analytical methods, and minimum detectable effect size anticipated. Exhibit 2-8, below, provides an overview of the timing of measurement and treatment in the Massachusetts and Virginia populations. Discussion sections on each of the study components follow the exhibit.

EXHIBIT 2-8. OVERVIEW OF EVALUATION IMPLEMENTATION



Design Details: Short-Term Impact of Compliance Assistance Outreach and Workshops

EPA measured the short-term impact of compliance assistance outreach and workshops through a random-assignment experiment involving auto body facilities in areas of eastern Massachusetts characterized by high air toxics risks. "Short-term impact" refers to changes in behavior that can be observed within approximately five to nine months of outreach and workshop completion (see schedule in Exhibit 2-8). Prior to study implementation, EPA expected that any detectable effects would most likely be associated with hazardous waste compliance assistance. EPA expected behavior changes related to the Surface Coating Rule to occur over a longer time frame, due in part to the 2011 effective date for the rule.

Treatment Approach

In August 2009, half of the 1,721 auto body shops in the study area in Massachusetts were randomly assigned to the treatment group (Group A); the remainder were assigned to the comparison group (Group B).[21] In October 2009, EPA sent the facilities in the treatment group a compliance assistance package consisting of: (1) a multimedia guidebook (including a DVD) providing a summary of air, water, and RCRA requirements impacting auto body shops in Massachusetts (2) a brochure summarizing the Surface Coating Rule requirements, (3) an invitation to attend workshops and webinars covering the requirements of the Surface Coating Rule and, to a lesser degree, RCRA and other environmental management issues, and (4) a copy of the presentation slides to be used at the workshops/webinars.[22] (Appendix A contains copies of most of these materials.) Between October 2009 and January 2010, 90 shops from the treatment group participated in a workshop or webinar offered by EPA. (EPA initially expected that 150-300 shops would participate in these offerings.) The shops in the control group did not receive the mailings until after the completion of the short-term study.[23,24]

Sampling and Measurement

Shortly after the workshops/webinars had been completed, in April through June 2010, EPA and contractor personnel conducted short (15-20 minute) telephone surveys at samples of shops from the treatment and control groups (A and B). After the phone surveys were complete, EPA and its contractors conducted site visits at a subset of the shops selected for phone surveys. (This approach is referred to as a "two-phase" survey.) Data from the site visits were used to assess shop performance and also gauge the accuracy of the phone survey results.[25]

---

[21] Initially, EPA estimated the size of the population of auto body shops in the study area in Massachusetts to be 1721 shops, but the Agency and its contractors subsequently found that some of these businesses were not actually auto body shops or had gone out of business. The final number of shops in the study area in Massachusetts was 1,636.

[22] These workshops/webinars were organized by EPA together with local partners, and they varied in content, duration, and location. However, at least one hour of all workshops will be dedicated to presenting the new requirements associated with the Surface Coating Rule. A standard PowerPoint presentation was used to cover material related to the Surface Coating Rule.

[23] EPA offered workshops and webinars to the comparison group facilities after the site visits were complete to ensure that all facilities had the opportunity to participate. Any facilities in the comparison group that learned about the earlier series of workshops and indicated to EPA that they would have liked to participate were encouraged to attend workshops at a later date.

[24] Facilities in both the treatment groups and the comparison group received a postcard in March 2009 informing them of pending 6H rule requirements. EPA staff felt this limited outreach was necessary for reasons of fairness, but they did not anticipate it would substantially impact facility performance in the short-term.

[25] EPA had hoped that the phone survey results could be used to improve the accuracy of the on-site performance estimates, but the survey results were not accurate enough for that purpose.

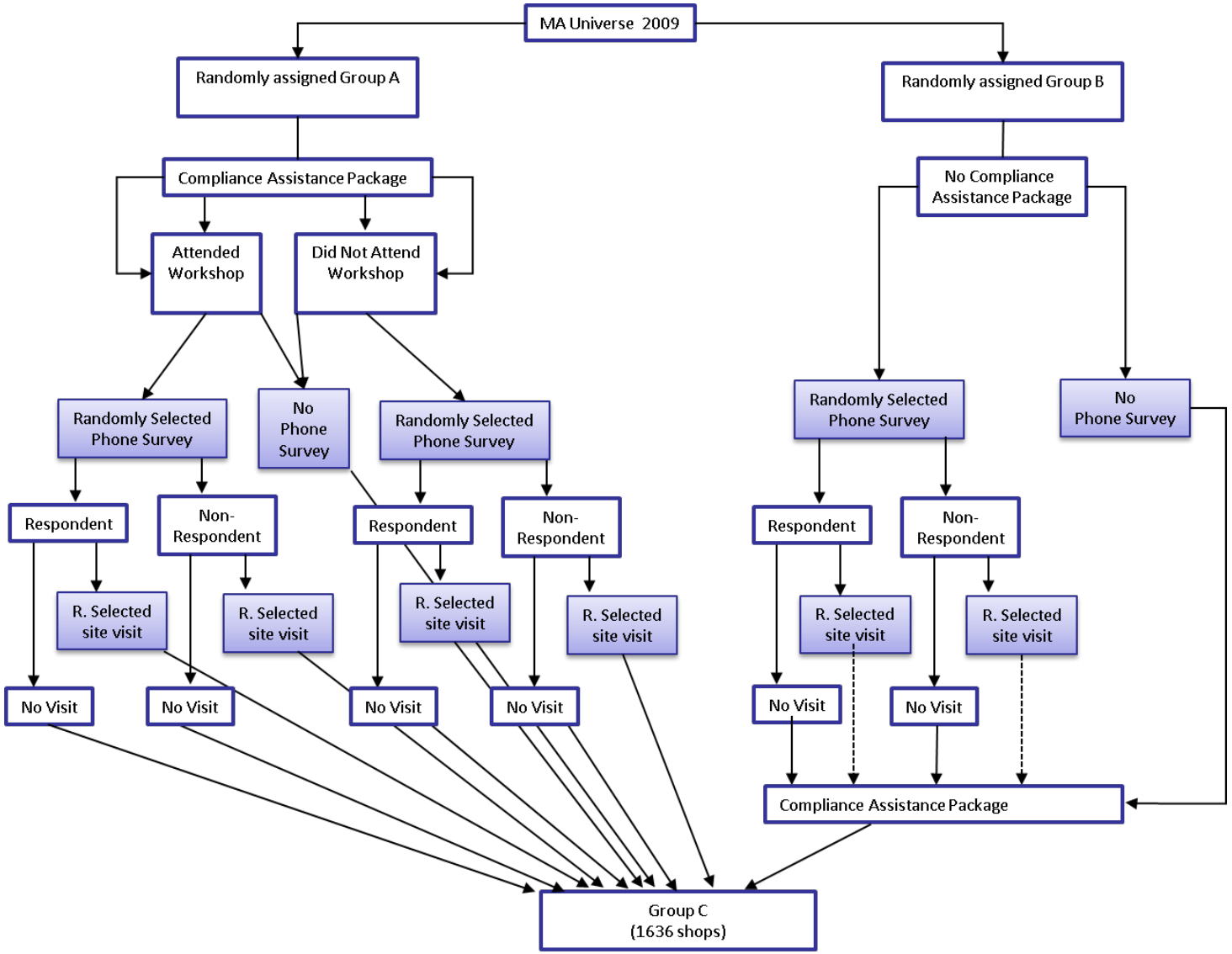
To conduct the site visits, EPA and contractor personnel traveled to the selected shops without advance notice and asked to speak with a shop representative regarding environmental issues.[26] If the shop representative was willing to participate, the site visitor then proceeded to gather data through a brief survey and through observations made during a shop walk-through.[27] In order to avoid potential interviewer bias, EPA and contractor personnel were not aware of the treatment status of the shops that they visited. When a selected shop refused to participate or was outside the target population (i.e., not an auto body shop or out of business), a randomly-selected backup shop from the same stratum was provided to the site visitor.

Exhibit 2-9 illustrates the process used in the short-term experiment.

[26] In some cases, site visitors called in advance to confirm a shop was in operation and/or the address of the shop; however the site visitors did not identify themselves or explain that they were planning to conduct a site visit.

[27] When the appropriate respondent was not present or was too busy to complete the interview, the interviewer made a follow-up appointment, attempting to schedule the appointment within a few days of the original attempt.

EXHIBIT 2-9. PROCESS FLOW FOR SHORT-TERM EXPERIMENT



To select the samples of shops included in the telephone surveys and site visits, EPA used a stratified random sample and proportional allocation. In the first phase (telephone surveys), there were a total of three strata, based on whether a facility was in Group A or Group B, and whether or not facility participated in a workshop/webinar (for Group A only). In the second phase (site visits), there were a total of six strata, based on whether a whether a facility was in Group A or B, whether or not facility participated in a workshop/webinar, and whether or not a facility responded to the telephone survey. Exhibit 2-10 summarizes the number of shops in each stratum for the on-site survey. (Additional details on the phone survey stratification approach are provided later in this chapter in the section "Design Details: Telephone Survey Verification.")

EXHIBIT 2-10. STRATIFICATION FOR SITE VISITS IN SHORT-TERM EXPERIMENT

STRATUM	GROUP	WORKSHOP OR WEBINAR PARTICIPANT	RESPONDED TO TELEPHONE SURVEY	NUMBER OF SHOPS IN STRATUM	NUMBER OF SHOPS SELECTED FOR SITE VISITS	NUMBER OF SITE VISITS COMPLETED
1	A	Yes	Yes	7	4	4
2	A	Yes	No	15	10	8
3	A	No	Yes	39	39	18
4	A	No	No	118	94	49
5	B	No	Yes	37	34	30
6	B	No	No	121	117	60
Total				**337[a]**	**298**	**169**

Note:
[a] The number of shops in all site visit strata (337) is less than the total number of telephone surveys attempted (412) because 75 shops were removed from the list at the time of stratification: 72 shops were removed because they were outside the target population, 2 shops were identified as duplicates, and 1 shop had a language barrier. After stratification was complete and site visit samples were drawn, an additional 3 shops were determined not to be in the target population because they did not conduct spray painting.

Exhibit 2-11 summarizes the site visit survey response outcomes. The overall response rate for the site visits was 81 percent, calculated as the number of respondents divided by the number of valid auto body shops in the site visit sample.[28] (The study design anticipated a response rate of at least 80 percent.) A substantial portion (30 percent) of the shops visited turned out not to be in the target population. This finding suggests the difficulty of generating an accurate list of auto body shops, even after conducting phone surveys which identified and eliminated some invalid shops from the list. The challenges of identifying an accurate list of auto body shops make it difficult to measure performance in this sector, regardless of whether the data source is a phone survey or site visits.

[28] The response rate (81 percent) is equal to 169 completed site visits, divided by the 298 site visits attempted minus 90 shops where visits were attempted but the shop was not in the target population (i.e., not in business or not operating as an auto body shop).

EXHIBIT 2-11. 2010 MASSACHUSETTS SITE VISIT SURVEY RESPONSE OUTCOMES

OUTCOME OF THE SITE VISIT	NUMBER OF SHOPS	PERCENTAGE OF TOTAL SHOPS CONTACTED
Completed survey	169	57%
Site visits not completed due to safety concerns	2	1%
Refused to complete survey	34	11%
Respondent unavailable[a]	3	1%
Not in the target population	90	30%
Total	**298**	**100%**

Note:

[a] EPA staff or contractors attempted to visit these shops but were unable to find someone at the shop to complete the survey.

Analysis

The overall impact of compliance assistance is estimated as the difference in performance between Groups A and B. This difference in performance is analyzed using two approaches. First, for each performance measure, the study compares the estimated proportion of shops in Group A with a positive response to the estimated proportion of shops in Group B in with a positive response (e.g., the percentage of facilities in Group A vs. Group B using appropriate spray booths for painting). Since the estimated proportions are based on samples, not a census of all facilities, the results are expressed with a margin of error for the difference between the proportions in Groups A and B. The analysis identifies performance measures that have statistically significant differences at 1 percent, 5 percent, and 10 percent significance levels. The analysis uses a one-sided hypothesis test, since EPA expected that assistance would improve (not decrease) performance levels and because only positive changes in performance levels would demonstrate effectiveness of the compliance assistance.

The second approach to comparing performance for Groups A and B involves using a multivariate regression analysis, using explanatory variables related to both shop capacity and external influences on shop behavior to control for factors that EPA anticipates may impact performance. The full list of variables used in the regression analysis is provided in the "Contextual Variables" section, above.

Design Details: Long-Term Impact of Compliance Assistance Package

EPA measured the long-term impact of the compliance assistance package (outreach, workshops, and facility visits) using a quasi-experiment involving auto body facilities in Massachusetts and a similar population of auto body facilities in the Tidewater/Piedmont regions of Virginia. The auto body shops in Massachusetts studied in the short-term experiment are identical to those studied in the long-term quasi-experiment. After the short-term impacts were measured, EPA offered Group B the same compliance assistance that had had been previously offered to Group A. The total population of Massachusetts facilities (Groups A and B combined) served as the treatment group for the long-term quasi-experiment (hereafter referred to as Group C). (Exhibit 2-9 above illustrates how Group C is comprised of the shops in Groups A and B). Facilities in Virginia served as the comparison group for measuring long-term impacts of EPA compliance assistance. "Long-term impact" refers to changes in behavior that can be observed up to one year and nine months after outreach and workshop completion (see schedule in Exhibit 2-8). Note that during the course of the long-term quasi-experimental study, all auto body shops were required to come into compliance with the new 6H rule.[29] The compliance date for this rule was in January 2011. Given the new compliance requirement, EPA expected that performance for *all* facilities would improve over the study period. However, given that compliance assistance was offered by EPA in Massachusetts but not in Virginia, EPA expected that the Massachusetts facilities would improve *more than* the facilities in Virginia. The long-term quasi-experiment was designed to test this hypothesis.

Treatment Approach

The Massachusetts facilities received three types of compliance assistance related to the Surface Coating Rule beginning in March 2009:

- **Outreach:** All facilities received a basic postcard in March 2009 (prior to the start of the statistically valid pilot project), identifying the upcoming compliance deadline for the Surface Coating Rule and describing the nature of the requirements. EPA also later sent the facilities in the treatment group a compliance assistance outreach package consisting of: (1) a multimedia guidebook providing a summary of air, water, and RCRA requirements impacting auto body shops in Massachusetts, (2) a brochure summarizing the Surface Coating Rule requirements, and (3) a copy of the presentation slides to be used at workshops/webinars. EPA sent this compliance assistance package to facilities in Groups A and B in October 2009 and August 2010, respectively. (Appendix A contains copies of most of these materials.)
- **Workshops/webinars:** All facilities were also offered an opportunity to participate in compliance assistance workshops/webinars led by EPA personnel.

[29] This statement refers to all auto body shops in existence at the time that the 6H rule was promulgated in January 2008. Auto body shops that began operations after that date were required to comply with the 6H rule when they began operations.

EPA conducted workshops and webinars covering the requirements of the Surface Coating Rule and, to a lesser degree, RCRA and other environmental management issues. Overall, 12 percent of Group C facilities participated in the workshops/webinars.30

- **On-site assistance:** During the course of the short-term experiment (described in the section above), EPA and contractor personnel conducted site visits at a sample of 169 facilities between May and July 2010. After measuring facility performance as part of the short-term experiment, EPA and contractor staff provided customized, on-site assistance to each auto body shop to help them understand their compliance requirements. Additionally, EPA Region 1 staff provided on-site assistance visits to facilities that requested such assistance. Overall, approximately 14 percent of the population received on-site assistance.

The Virginia facilities received no EPA assistance prior to measurement.

Sampling and Measurement

EPA measured performance at two points in time in both the treatment and comparison groups (via independent samples rather than panels), resulting in four separate estimates of performance:

1. Massachusetts 2010 (Pre-treatment: estimate obtained from Group B)
2. Massachusetts 2011 (Post-treatment; estimate obtained from all of Group C)
3. Virginia 2010
4. Virginia 2011

EPA estimated the performance for the Massachusetts shops in 2010, prior to receiving assistance, using the data from Group B in 2010 from the short-term experiment. Recall that in the short-term experiment, facilities in Massachusetts were randomly divided into two groups, A and B. At the time of the short-term measurement survey in 2010, Group B had not yet received workshop/webinar offers, the compliance assistance outreach package, or on-site assistance. Thus, because Group B was randomly selected from all of the shops in the study area in Massachusetts (Group C), and because Group B had not received treatment at the time of measurement, Group B served as a baseline measurement for Group C.

Performance estimates for all four groups (Massachusetts 2010, Massachusetts 2011, Virginia 2010, and Virginia 2011) are based on data gathered by EPA and contractor personnel through on-site observations. As with the short-term experiment, interviewers traveled to the selected shops without advance notice and asked to speak with a shop representative regarding environmental issues. If the shop representative was willing to participate, the interviewer then proceeded to gather data through a brief survey and

[30] Facilities in Groups A and B were offered this opportunity beginning in October 2009 and August 2010, respectively.

through observations made during a shop walk-through.[31] When a selected shop refused to participate or was outside the target population (i.e., not an auto body shop or out of business), a randomly-selected backup shop from the same stratum was provided to the interviewer.

In Group C, surveys were completed at 90 facilities in 2010 and at 101 facilities in 2011; sample size targets had been 100 and 100, respectively. The 2010 sample was drawn as described earlier for Group B in the short-term experiment. The 2011 sample was stratified based on whether or not a shop had received interactive EPA assistance (including workshops, webinars, or on-site assistance). Within each stratum, a simple random sample was selected for site visits.

In Virginia, EPA and contractor staff completed site visits at 93 facilities in 2010 and 86 facilities in 2011; sample size targets had been 100 and 100, respectively. The sampling method for Virginia in both 2010 and 2011 was simple random sampling from the population of shops. In Virginia, the 2010 sample was excluded from the sampling frame prior to drawing the sample for 2011, because the process of conducting the site visits and any assistance provided by the site visitor could have affected shop behavior, and therefore shops that had been sampled were no longer a true control group.

Exhibit 2-12 summarizes the stratification for site visits for the long-term quasi-experiment.

EXHIBIT 2-12. STRATIFICATION FOR SITE VISITS IN LONG-TERM QUASI-EXPERIMENT

YEAR	STRATUM	GROUP	RECEIVED INTERACTIVE ASSISTANCE?	RESPONDED TO TELEPHONE SURVEY	NUMBER OF SHOPS IN STRATUM	NUMBER OF SHOPS SELECTED FOR SITE VISITS	NUMBER OF SITE VISITS COMPLETED
2010	5	MA - B	No	Yes	37	34	30
2010	6	MA - B	No	No	121	117	60
2011	7	MA - C	Yes	N/A	279	20	18
2011	8	MA - C	No	N/A	1,190	132	83
2010	9	VA	No	N/A	443	172	91
2011	10	VA	No	N/A	231	226	86
	Total						368

Exhibits 2-13 and 2-14 summarize the site visit survey response outcomes from Massachusetts and Virginia, respectively. The overall response rate for the site visits for the long-term study was 83.8 percent, calculated as the number of respondents divided by

[31] When the appropriate respondent was not present or was too busy to complete the interview, the interviewer made a follow-up appointment, attempting to schedule the appointment within a few days of the original attempt.

the number of valid auto body shops in the site visit sample. (The study design anticipated a response rate of at least 80 percent.) The Massachusetts response rate was 84.9 percent, while the Virginia response rate was 82.7 percent.

A substantial portion (37 percent) of the shops visited turned out not to be in the target population. The figure was substantially higher in Virginia than in Group C; in Group C, returned outreach mailings and phone surveys helped to identify and remove a number of invalid shops from the list prior to site visits. The high dropout rate is consistent with that observed for Group A in the short-term experiment. Altogether, this finding suggests that efforts to remove the invalid auto body shops from the sample frame can substantially improve the efficiency of on-site survey methods. However, even after such methods, the list was still inaccurate enough to make it difficult to measure performance in this sector.

EXHIBIT 2-13. GROUP C SITE VISIT SURVEY RESPONSE OUTCOMES

OUTCOME OF THE SITE VISIT	MASSACHUSETTS 2010 (GROUP B)		MASSACHUSETTS 2011	
	NUMBER OF SHOPS	PERCENTAGE OF TOTAL SHOPS CONTACTED	NUMBER OF SHOPS	PERCENTAGE OF TOTAL SHOPS CONTACTED
Completed survey	90	60%	101	66%
Site visits not completed due to safety concerns	1	1%	1	1%
Refused to complete survey	19	13%	6	4%
Respondent unavailable[a]	1	1%	6	4%
Not in the target population	40	26%	38	25%
Total	**151**	**100%**	**152**	**100%**

Note:

[a] EPA staff or contractors attempted a visit, but were unable to find someone at the shop to complete the survey.

EXHIBIT 2-14. VIRGINIA SITE VISIT SURVEY RESPONSE OUTCOMES

OUTCOME OF THE SITE VISIT	VIRGINIA 2010		VIRGINIA 2011	
	NUMBER OF SHOPS	PERCENTAGE OF TOTAL SHOPS CONTACTED	NUMBER OF SHOPS	PERCENTAGE OF TOTAL SHOPS CONTACTED
Completed survey	91	53%	86	38%
Site visits not completed due to safety concerns	0	0%	1	0%
Refused to complete survey	12	7%	7	3%
Unable to reach shop[a]	5	3%	12	5%
Not in the target population	64	37%	120	53%
Total	**172**	**100%**	**226**	**100%**

Note:

[a] EPA staff or contractors attempted a visit, but were unable to find someone at the shop to complete the survey.

Analysis

The impact of EPA Region 1's two-year compliance assistance effort (outreach, workshops, and facility visits) is estimated as the difference-in-differences between Massachusetts and Virginia shops measured in 2010 and 2011. Specifically, the study measures the difference between: 1) the change in performance over time for Massachusetts shops that received EPA compliance assistance and 2) the change in performance over time for Virginia shops that did not receive compliance assistance.

Two different approaches are used to estimate the difference-in-differences. The first approach compares the estimated change over time in the proportion of shops in with a positive response for each measure for each group (Massachusetts vs. Virginia). Since the estimated proportions are based on samples, not a census of all facilities, the results are expressed with a margin of error for the difference-in-differences. The analysis identifies performance measures that have statistically significant difference-in-differences at 1 percent, 5 percent, and 10 percent significance levels. The analysis uses a one-sided hypothesis test, since EPA expected that assistance would improve (not decrease) performance levels and because only positive changes in performance levels would demonstrate effectiveness of the compliance assistance.

The second approach involves estimating the difference-in-differences in the context of multiple regression analysis, using explanatory variables related to both shop capacity and external influences on shop behavior to control for factors that EPA anticipates may impact performance. (Exhibit 2-15 provides background on multivariate regression.) The independent variables used in the regression for the long-term analysis are the same as those used for the short-term analysis, described earlier.

EXHIBIT 2-15. MULTIPLE REGRESSION ANALYSIS

Multiple regression analysis is a statistical technique used to quantify the relationship between several "independent" variables and a single "dependent" variable. For example, suppose one wanted to determine how various characteristics impact the selling price of single-family homes. Multiple regression analysis could be applied to a dataset that includes numerous housing transactions in a single housing market. The dependent variable would be the selling price. The independent variables would be characteristics such as square feet of living area, age, lot size, and distance to the nearest city. Suppose we describe the relationship between selling price and these four characteristics using the following simple equation:

PRICE = a + b X SQUAREFEET + c X AGE + d X ACRES + e X DISTANCE

Here, the lower-case letters (a,b,c,d, and e) represent unknowns that are referred to as "coefficients." Multiple regression analysis uses the data on housing transactions to develop estimates for these coefficients. Once the coefficients have been estimated, they can be used to quantify the impact of any individual characteristic on PRICE, holding all other characteristics constant. For example, if the estimated value of d is 6,000, then each additional acre is predicted to increase selling price by $6,000. The results can also be used to determine whether or not the impact of any given independent variable on PRICE is "statistically significant." That is, is the estimated value of the coefficient large enough that it is unlikely to have occurred by chance?

In the above example, the dependent variable (PRICE) is continuous and a linear regression is used. When the dependent variable is binary (e.g., a shop is either in compliance or not in compliance), then linear regression is no longer appropriate, but logistic regression analysis can be applied. Logistic regression analysis is a form of regression analysis that accounts for the binary nature of the dependent variable.

Design Details: Telephone Survey Verification

In assessing the impact of compliance assistance activities, EPA typically relies on the results of telephone surveys of regulated facilities. OMB has questioned the accuracy of these telephone survey efforts, given the potential for self-reporting bias and non-response bias. The two-phase sampling approach used in the pilot project gathered data to quantify the accuracy of telephone surveys for auto body shops in eastern Massachusetts in 2010. The study does this by conducting telephone surveys followed by site visits with telephone survey respondents (to assess self-reporting bias) and non-respondents (to assess non-response bias).

Sampling and Measurement

EPA measured environmental performance at Massachusetts auto body shops (Groups A and B) shortly after offering a compliance assistance package to Group A. In the first phase, the workgroup attempted telephone interviews with a stratified random sample of 412 Massachusetts shops between April and June 2010. The telephone survey sample was stratified by whether or not a shop was offered the compliance assistance package (Group A or B) and, for Group A, by whether or not the shop attended a workshop or webinar offered by EPA. A total of 80 facilities were contacted for the telephone survey. Exhibit 2-16 summarizes the stratification approach for the telephone surveys.

EXHIBIT 2-16. STRATIFICATION FOR TELEPHONE SURVEY

STRATUM	GROUP	WORKSHOP OR WEBINAR PARTICIPANT	NUMBER OF SHOPS IN STRATUM	NUMBER OF SHOPS SELECTED FOR TELEPHONE SURVEY	NUMBER OF TELEPHONE SURVEYS COMPLETED OR PARTIALLY COMPLETED
1	A	Yes	90	22	7
2	A	No	720	190	35
3	B	No	826	200	38
Total			**1,636[a]**	**412**	**80**

Note:

[a] Although 1,721 shops were initially assigned to treatment and control groups, 85 of these were removed prior to the telephone survey because they were identified during the treatment period as being outside of the target population (i.e., they were not auto body shops or they were out of business).

Exhibit 2-17 shows the response outcomes of the telephone surveys. The overall response rate for the telephone survey was 28 percent, calculated as number of completed or partially completed surveys divided by the estimated number of valid auto body shops in the telephone sample. Only shops that were in business and operating as auto body shops were classified as "valid" in calculating the response rate.[32] The response rate of 28 percent was within the range of 20 - 40 percent anticipated in the study design.

EXHIBIT 2-17. 2010 TELEPHONE SURVEY RESPONSE OUTCOMES

CATEGORY	OUTCOME OF THE CALL	NUMBER OF SHOPS	PERCENTAGE OF TOTAL SHOPS CONTACTED
Telephone Respondent	Completed survey	69	17%
	Partially completed survey	11	3%
Telephone Non-Respondent	Refused to complete survey	79	19%
	Unable to reach shop	177[a]	43%
	Language barrier	1	0%
Not in Target Population	Not an auto body shop or out of business	75	18%
	Total	**412**	**100%**

Note:
[a] Telephone interviewers attempted to contact these shops up to three times, calling on different days and at different times of day. Three contact attempts were made at 90 percent of the shops in this group, while one or two contact attempts were made at the remaining 10 percent of the shops.

In the second phase, the workgroup completed site visits in May, June and July 2010 at a stratified random sample of 169 of the shops selected for the telephone survey. The site visit sample was stratified by group (A or B), by response to the telephone survey (respondent or non-respondent), and, for Group A, by attendance at a workshop/webinar offered by EPA. The stratification and response outcomes related to the site visits are shown in Exhibits 2-12 and 2-11, earlier in this chapter.

[32] The response rate (28 percent) is equal to 80 shops that completed or partially completed telephone surveys, divided by the estimated number of valid shops: 412 shops selected for the telephone survey, minus 75 shops that phone surveyors found were invalid, minus an estimated 62 shops that were likely invalid, but that telephone surveyors were unable to reach. The number 62 is based on the proportion of shops in this stratum that on-site surveyors did reach and that turned out to be invalid (35.3 percent), multiplied by the 177 shops that telephone interviewers were unable to reach.

Analysis

The telephone survey bias can be expressed as the difference between the estimated performance from the telephone survey respondents (P_{Phone}) and the estimated performance obtained from the site visit survey respondents ($\hat{P}_{site}$), for matching survey questions on which performance could be verified on-site:

$$\hat{B} = \hat{P}_{phone} - \hat{P}_{site}$$

Chapter 5 of this report presents findings of this comparison.

STUDY LIMITATIONS

As with any evaluation, the findings of this evaluation are limited by a variety of constraints. The most important limitations are described below:

- **Limited scope of the study.** The study represents a picture of compliance for one window of time, in one sector, testing the impact of one package of compliance assistance. The results of the analyses are relevant for the population of auto body shops located in risk-based clusters in Massachusetts. EPA did not intend, nor would it be appropriate, to use the results of the pilot study to draw conclusions about the compliance assistance program as a whole.
- **Indirect effects of EPA assistance.** The study does not measure any indirect effects of EPA assistance. For example, the study does not measure the impact of EPA training suppliers and trade associations, who in turn provide information to auto body shops. If there were substantial indirect impacts, especially within the control group, the impact of direct compliance assistance would be harder to detect.
- **Compliance assistance from non-EPA sources:** A large proportion of shops in both the treatment and control groups received compliance assistance from suppliers (between 88 percent and 96 percent across all shops sampled in Massachusetts and Virginia for both study years). Trade associations also provided compliance assistance to a number of shops (15 percent of shops sampled in Massachusetts in 2011, and 27 percent of shops sampled in Virginia in 2011). While suppliers and trade associations may have assisted slightly more shops in Virginia than in Massachusetts in 2011, which would reduce the detectible difference-in-differences between the treatment and comparison groups in the long-term quasi-experiment, these effects are likely to be small.
- **Minimum detectable effect:** Due to resource constraints, the sample sizes for both the short- and long-term comparisons are somewhat limited, leading to higher-than-ideal minimum detectable effects for each of the two comparisons. As a result, limited compliance assistance impacts are unlikely to be detected. For instance, we would not expect to be able to statistically confirm behavioral

changes that occur at less than 9 to 15 percent of facilities in the treatment group for the short-term experiment.

- **Diluted treatment effects.** At the outset of the study, EPA anticipated that its most effective compliance assistance strategy is on-site assistance, followed by workshops/webinars. However, only between 15 and 18 percent of shops sampled received interactive assistance (workshops, webinars, and/or site visits).[33] The impacts of this assistance are diluted in the results, given the small proportions of facilities receiving this assistance.
- **Measurement effects:** It is possible that the process of measuring the effects of compliance assistance, through telephone surveys and site visits, may have influenced performance. To counteract this potential measurement effect, shops visited in 2010 were excluded from the 2011 sample, but it's possible that other shops may have learned about EPA's increased attention in the sector, and consequently improved their performance. For example, site visitors in Virginia reported that some shops – particularly larger ones – that refused to participate seemed to be aware of EPA's new presence in the sector, and possibly aware of the study itself. If there are in fact strong measurement effects from our data collection efforts, our estimates may not fully reflect the true impacts of compliance assistance.
- **Economic recession/downturn:** The economic conditions existing during the study may compromise the degree to which EPA can generalize the study results. If facilities are currently more reluctant to invest in compliance-related purchases or training than they would be under more favorable economic conditions, or if short staffing makes complying with operational requirements more difficult, then the impact of compliance assistance may be unusually small.

[33] In the short term study, about 15 percent of the shops in the treatment group sample in Massachusetts in 2010 attended these workshops/webinars; in the long-term study about 18 percent of the Massachusetts 2011 sample received interactive EPA compliance assistance.

CHAPTER 3 | SHORT-TERM EFFECTIVENESS OF EPA COMPLIANCE ASSISTANCE

This chapter describes findings from the short-term experiment, which was designed to evaluate the near-term effectiveness of a compliance assistance package that EPA Region 1 offered to auto body shops in Massachusetts. The compliance assistance was related to hazardous waste regulations and to EPA's new spray coating rule and it consisted of mailings and workshop/webinar opportunities. (See Chapter 2 for a detailed description of the compliance assistance that was offered.)

Half of the auto body shops in the target population were randomly assigned to a treatment group and half to a control group. EPA offered compliance assistance to all shops in the treatment group and delayed assistance for shops in the control group until after the short-term experiment had ended.

Several months after offering compliance assistance to the treatment group, EPA staff and contractors conducted site visits at a random sample of shops from each group in order to measure environmental performance. Because the shops from each group were randomly assigned, they can be considered statistically equivalent. The control group is used to estimate what shop performance would be without treatment, and the difference in performance between the treatment and control group is used to estimate the incremental impact of EPA compliance assistance. (See Chapter 2 for a detailed description of the measurement and analytic approach.)

The findings from the analysis of the short-term experiment are as follows:

- This study does not provide evidence that EPA assistance to auto body shops affected sector-wide performance. A simple comparison of the groups' performance levels shows statistically significant differences for two performance measures, but the differences were too small to be of *practical* significance. It would be difficult to detect any effects of compliance assistance for the 10 of 20 performance measures on which control group performance was quite high (greater than 90 percent of shops were in compliance with the measure). For those measures, there was little room for the treatment group performance to exceed that of the control group. However, even for measures where the control group performance was not high (i.e., where there was room for improvement), the analysis does not show any significant impact of EPA assistance.
- Shops that chose to participate in workshops/webinars (15 percent of the treatment group in the sample) performed significantly better on five measures than the remaining shops in the treatment group that did not avail themselves of those

opportunities. On these measures, participants' performance ranged from 6 to 44 percentage points better than that of non-participants. For example, with regard to properly labeling hazardous waste drums, shops that attended a workshop or webinar performed 33 percentage points better than shops that did not attend a workshop or webinar. However, these results only reflect the short-term impact of attending a workshop or webinar. (Chapter 4 discusses the longer-term impact of participating in workshops/webinars, as well as receiving on-site assistance.) Moreover, it is not possible to separate out the impact of the workshops/webinars relative to the potential effect of self-selection bias. For example, workshop participants may be systematically different from non-participants; their performance may have been superior even if they had not participated in the workshops/webinars.

The remainder of this chapter discusses the findings in detail, including characteristics of the auto body shops in short-term experiment, a comparison of environmental performance of treatment and control groups, and a comparison of Massachusetts auto body shops in the treatment group in 2010 that did versus did not attend a workshop/webinar.

CHARACTERISTICS OF AUTO BODY SHOPS IN THE TREATMENT AND CONTROL GROUPS

As expected, the vast majority of the auto body shops in the sample were small, independent operations. Sampled shops reported completing an average of only 7.3 paint jobs per week. Only 11 percent of shops sampled in Massachusetts in 2010 indicated that they were part of a corporate chain. Approximately 50 percent of sampled shops classified themselves as very small quantity generators of hazardous waste (VSQGs), 50 percent classified themselves as small quantity generators (SQGs), and only one shop classified itself as a large quantity generator (LQG). When asked how they obtain information about how to comply with state and federal regulations, 90 percent of the shops cited suppliers, 20 percent cited EPA, and 13 percent cited trade associations (multiple responses were allowed, so percentages do not sum to 100). As expected with random assignment to treatment/control groups, the characteristics of the treatment group shops are generally similar to the characteristics of the control group shops.

Exhibit 3-1 summarizes characteristics of shops in the treatment and control groups in Massachusetts in 2010.

EXHIBIT 3-1. CHARACTERISTICS OF AUTO BODY SHOPS IN MASSACHUSETTS IN 2010

CHARACTERISITIC	PERCENTAGE OF SHOPS IN GROUP A	PERCENTAGE OF SHOPS IN GROUP B	TOTAL PERCENTAGE OF SHOPS SAMPLED IN MASSACHUSETTS IN 2010
Part of a corporate chain	14%	9%	11%
VSQG	53%	46%	50%
SQG	46%	54%	50%
LQG	1%	0%	1%
Receive information on how to comply with federal and state environmental regulations from:			
• Suppliers	93%	87%	90%
• Corporate environmental division	0%	1%	1%
• Educational institutions (e.g., vocational technical school)	1%	2%	2%
• Environmental consultant	4%	3%	4%
• Other auto body shops	1%	2%	2%
• Trade association	11%	16%	14%
• Local government	3%	1%	2%
• Occupational Safety and Health Administration (OSHA)	0%	0%	0%
• State environmental agency	3%	6%	4%
• U.S. EPA	20%	20%	20%
• Other sources	0%	9%	5%

ENVIRONMENTAL PERFORMANCE OF TREATMENT AND CONTROL GROUPS

This section compares the environmental performance of the treatment and control groups using the performance measures discussed in Chapter 2. The performance measures are a set of shop characteristics that (1) were potentially impacted by compliance assistance efforts and (2) could be independently verified through site visits. The treatment and control groups are compared first through a simple comparison of group means and then through a regression-based approach that estimates the treatment effect while controlling for shop characteristics.

Simple Comparison of Means

The mean performance levels for the treatment and control groups are presented in Exhibit 3-2. Of the 20 performance measures evaluated, the performance of the treatment group was higher than the performance of the control group for 8 measures, the performance of the control group was higher for 11 measures, and the two groups had identical performance for one measure. The performance difference between the treatment and control groups was highest for prep_enclosed (control group 25.6 percentage points higher), mixroom_vent (control group 11.1 percentage points higher), and train_records (treatment group 8.9 percentage points higher).[34] Using a one-sided hypothesis test, only two of the 20 differences between the two groups were statistically significant.[35] Specifically, the difference for booth_exists was significant at the 5 percent level (treatment group performance was 4.3 percentage points higher) and the difference for not_outside was significant at the 10 percent level (treatment group performance was 2.2 percentage points higher). While the difference between the treatment and control groups is statistically significant for these two performance measures, it is not practically significant, since more than 95 percent of shops in the treatment and control group were in compliance with the measure. It is also important to recognize that when conducting multiple comparisons, i.e., comparing treatment and control groups for 20 measures, the relatively large number of comparisons increases the likelihood that a statistically significant difference will be detected just by chance.

[34] See Exhibit 2-5 for definitions of performance measures.

[35] One-sided hypothesis tests are appropriate when the researcher has a strong *a priori* expectation that the result can only go in one direction and when a result in the opposite direction is considered functionally equivalent to no difference at all. The practical impacts of using a one-sided hypothesis test are (1) outcomes where the performance of the control group is better than the performance of the treatment group are never considered statistically significant and (2) outcomes where the performance of the treatment group is better are more likely to be classified as statistically significant than under a two-sided test. As discussed in the Information Collection Request submission to the Office of Management and Budget (Appendix E), the research team chose to use one-sided hypothesis tests *before* conducting the experiment.

EXHIBIT 3-2. COMPARISON OF AVERAGE PERFORMANCE FOR TREATMENT AND CONTROL GROUPS

CATEGORY	PERFORMANCE MEASURE[A]	TYPE OF MEASURE	TREATMENT GROUP		CONTROL GROUP		DIFFERENCE[B]	MARGIN OF ERROR FOR DIFFERENCE	REGRESSION-ADJUSTED DIFFERENCE[C]
			PERCENT	N	PERCENT	N			
SPRAY BOOTH	**booth_exists**	**Air - Spray booth**	**100.0%**	**79**	**95.7%**	**90**	**4.3%****	**±3.5%**	
	not_outside	**Air - Spray booth**	**100.0%**	**79**	**97.8%**	**88**	**2.2%***	**±2.6%**	
	booth_enclosed	Air – Spray booth	98.8%	77	100.0%	86	-1.2%	±1.9%	
	booth_ventilated	Air – Spray booth	95.8%	76	98.9%	85	-3.1%	±4.1%	
	filter_exists	Air – Spray booth	97.1%	78	100.0%	87	-2.9%	±3.1%	
	filter_good	Air – Spray booth	60.1%	75	59.3%	86	0.8%	±12.5%	0.9%
	capture98	Air – Spray booth	83.9%	13	83.0%	20	0.9%	±22.4%	
PREP STATION	**prep_enclosed**	Air – Prep Station	69.8%	13	95.4%	19	-25.6%	±9.8%	
	prep_vent	Air – Prep Station	94.6%	11	100.0%	16	-5.4%	±11.9%	
MIXING ROOM	**mixroom_enclosed**	Air – Mixing Room	93.2%	54	93.0%	58	0.2%	±7.8%	0.7%
	mixroom_vent	Air – Mixing Room	72.4%	49	83.5%	55	-11.1%	±13.0%	-10.8%
PAINT STRIP	**avoid_mecl**	Air – Paint Stripping	94.9%	79	91.1%	88	3.8%	±6.3%	3.4%
SPRAY GUNS	**guns_compliant**	Air – Spray Guns	100.0%	78	100.0%	90	0.0%	±0.0%	
	cleaning_compliant	Air – Spray Guns	75.9%	79	78.8%	89	-2.8%	±10.6%	-4.5%
	train_records	Air – Spray Guns	54.6%	79	45.7%	90	8.9%	±12.6%	10.0%
WASTE MGMT	**drums_labeled**	Waste Management	30.2%	77	34.3%	90	-4.2%	±11.7%	-0.4%
	drums_closed	Waste Management	55.5%	78	60.5%	90	-4.9%	±12.1%	-4.2%
	rags_closed	Waste Management	24.6%	59	29.3%	69	-4.7%	±13.0%	-2.1%
	no_spills	Waste Management	81.0%	79	86.5%	90	-5.5%	±9.3%	-3.7%
	waste_doc	Waste Management	74.6%	79	73.9%	87	0.7%	±11.1%	5.2%

Notes:

[A] See Exhibit 2-5 for definitions of performance measures.

[B] ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (one-sided test). These measures are in bold text.

[C] As discussed in the text, regression models were not estimated for performance measures with small sample sizes and/or with average performance levels near 100%.

Regression Analysis

While random assignment helps to ensure that the treatment and control groups are similar with respect to all characteristics other than the treatment itself, it is possible that differences between the two groups will arise by chance, particularly with relatively small sample sizes for the two groups. As a result, and in order to potentially achieve greater precision in estimating the magnitude of the treatment effect, the differences between the treatment and control groups were also evaluated within a multivariate regression context.

For example, suppose that shops that are part of corporate chains receive better training and access to environmental consultants than other shops, and they therefore have better environmental performance. In addition, suppose that the treatment group sample happens (by chance) to have more of these shops than the control group sample. In this situation, a simple comparison of means would overestimate the impact of compliance assistance, as it would attribute the entire difference between the two groups to compliance assistance, when in reality a portion of the difference can be attributed to the fact that the treatment group has more shops that are part of corporate chains. When analyzed within a multivariate regression context, the portion of the treatment/control difference that can be attributed to a difference in the number of shops that are part of a corporate chain is implicitly subtracted out, and only the residual difference is attributed to compliance assistance.

As all of the performance measures are binary (0/1) variables, logistic regression analysis was applied. The independent variables are summarized in Exhibit 3-3. The "treatment" variable is a binary variable intended to capture the impact of EPA compliance assistance. The remaining independent variables are shop characteristics that may be related to environmental performance, but that were unlikely to have been influenced by the treatment (i.e., by EPA compliance assistance).

Several performance measures were not included in the regression analysis due to inadequate sample sizes and/or average performance levels that were close to 100 percent. Having a preponderance of observations where the dependent variable is either zero or one frequently leads to estimation problems with logistic regression, particularly when sample sizes are small and explanatory variables are binary rather than continuous.

In addition to the individual performance measures, a "rollup" measure was developed that summarizes performance across all 20 measures. The rollup measure is defined as the percentage of performance measures that are equal to one, and it is analyzed using standard linear regression techniques.

EXHIBIT 3-3. DEFINITIONS OF REGRESSION VARIABLES

VARIABLE NAME	TYPE	MEAN	DEFINITION
Treatment	binary	0.47	= 1 if the shop is in the treatment group; = 0 if the shop is in the control group
corp_chain	binary	0.11	= 1 if the shop is part of a corporate chain; = 0 otherwise
num_jobs	continuous	7.26	Number of paint jobs per week (estimated by respondent)
SQG	binary	0.50	= 1 if the shop is a small or large quantity generator of hazardous waste; = 0 otherwise.
aware_b2009	binary	0.06	= 1 if the shop learned about the EPA spray regulations before 2009; = 0 otherwise
non_EPAvisit	binary	0.15	= 1 if the shop was inspected or visited by a non-EPA government environmental or health and safety official within the last six months; = 0 otherwise
pvt_assist	binary	0.89	= 1 if the shop indicated that it obtains information about environmental compliance from coating manufacturers, suppliers, consultants, or a trade association; = 0 otherwise

Exhibit 3-4 shows the results of the regression analysis. For the linear regression of the rollup measure, the coefficients provide the estimated impact on the dependent variable (i.e., rollup of environmental performance) of a one-unit increase in the independent variable. For example, the coefficient of 0.049 associated with SQG indicates that the rollup performance measure is approximately 5 percentage points higher for SQGs than for VSQGs.

In the linear regression model with the rollup measure as the dependent variable, three variables were statistically significant in addition to the constant term. Specifically, the coefficients associated with num_jobs and SQG were positive and significant at the 1 percent level and the coefficient associated with pvt_assist was positive and significant at the 5 percent level. These results indicate that overall environmental performance is better in larger shops (as measured by number of jobs and hazardous waste generator status) and in shops that say they obtain information on environmental compliance from coating manufacturers, suppliers, consultants, or a trade association.

In the logistic regression models with individual performance measures as dependent variables, only two of the independent variables (num_jobs and SQG) had coefficients that were statistically significant across multiple performance measures with significance

levels lower than 10 percent. The coefficient associated with num_jobs was positive and significant (at the 1percent level) in the models with filter_good and mixroom_vent as dependent variables. The coefficient associated with SQG was positive and significant (at the 1 percent level) in the models with train_records and waste_doc as dependent variables, and it was positive and significant (at the 10 percent level) in the model with drums_closed as the dependent variable. These two variables (num_jobs and SQG) could be conceived as approximate measures of shop size or volume of work.

The coefficient associated with the treatment variable was not statistically significant in any of the estimated models. The treatment coefficients were used to estimate the regression-adjusted difference between the treatment and control groups (holding all other variables at their means), and these differences are reported in the final column of Exhibit 3-2. The regression-adjusted estimates of the differences were generally similar to the estimates obtained through comparisons of group means. Thus, although several different confounding factors did have statistically significant impacts on environmental performance, the similarity of the regression-based estimates to the group mean estimates indicates that these factors were likely fairly well balanced between the treatment and control groups.[36]

[36] Models were also estimated with binary indicator variables for each interviewer in order to control for potential interviewer effects. The estimated treatment effect with interviewer variables was generally similar to the estimated treatment effect without interviewer variables, and the interviewer variables were excluded from the final regressions for simplicity of presentation. The absence of an impact was likely due to the fact that the interviewers were fairly well balanced between the treatment and control groups (i.e., each interviewer had a similar number of shops from each group).

EXHIBIT 3-4. REGRESSION COEFFICIENTS (Z-STATISTICS IN PARENTHESES) [a, b, c]

INDEPENDENT VARIABLE	DEPENDENT VARIABLE [d]									
	ROLLUP	FILTER GOOD	MIXROOM ENCLOSED	MIXROOM VENT	CLEANING COMPLIANT	TRAIN RECORDS	DRUMS LABELED	DRUMS CLOSED	NO SPILLS	WASTE DOC
treatment	-0.003	0.04	0.14	-0.77	-0.26	0.40	-0.02	-0.17	-0.30	0.32
	(-0.17)	(0.12)	(0.15)	(-1.35)	(-0.68)	(1.18)	(-0.05)	(-0.52)	(-0.70)	(0.81)
corp_chain	-0.030	0.21	-1.13	1.05	0.80	-0.05	-0.30	**-1.01***	-0.83	-0.76
	(-1.13)	(0.31)	(-0.99)	(0.74	(0.90)	(-0.08)	(-0.50)	(-1.70)	(-1.38)	(-1.10)
num_jobs	**0.001*****	**0.12*****	0.02	**0.17*****	0.02	0.02	-0.01	0.01	0.01	0.00
	(2.33)	(3.23)	(1.27)	(3.21)	(0.77)	(0.67)	(-0.37)	(0.78)	(0.33)	(0.24)
SQG	**0.049*****	-0.35	1.02	-0.19	-0.35	**0.90*****	0.03	**0.61***	-0.71	**2.18*****
	(2.60)	(-0.97)	(1.01)	(-0.30)	(-0.88)	(2.62)	(0.07)	(1.77)	(-1.47)	(4.59)
aware_b2009	0.024	**1.97***	**-2.35***	-1.68	-0.33	1.03	**1.67***	0.05	0.76	-0.52
	(0.73)	(1.78)	(-1.80)	(-1.50)	(-0.38)	(1.13)	(1.77)	(0.07)	(0.57)	(-0.71)
non_EPAvisit	-0.004	-0.35	-0.91	0.13	0.02	-0.45	0.42	-0.35	1.40	-0.43
	(-0.15)	(-0.64)	(-0.83)	(0.17)	(0.04)	(-0.86)	(0.90)	(-0.75)	(1.28)	(-0.79)
pvt_assist	**0.02****	-0.23	0.77	-0.02	-0.46	**0.99***	0.84	0.38	0.59	0.85
	(1.97)	(-0.41)	(0.93)	(-0.03)	(-0.65)	(1.66)	(1.30)	(0.69)	(0.82)	(1.31)
constant	**0.691*****	0.07	1.80	0.84	**1.77****	**-1.68*****	**-1.62****	-0.12	**1.60****	-0.54
	(22.58)	(0.12)	(1.56)	(0.92)	(2.32)	(-2.73)	(-2.35)	(-0.22)	(2.46)	(-0.80)
n	163	155	108	100	162	163	161	162	163	162

Notes:

[a] ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (one-sided test for "treatment" and two-sided test for all other variables). These measures are in bold text.

[b] The first column presents results from a linear regression with a continuous dependent variable (rollup), while the remaining columns present results from logistic regressions with binary (0/1) dependent variables.

[c] Design-based weights were used in estimation.

[d] See Exhibit 2-5 for definitions of performance measures.

COMPARISON OF SHOPS THAT DID VERSUS DID NOT ATTEND A WORKSHOP/WEBINAR

An important part of the compliance assistance EPA offered to shops in the treatment group was the opportunity to attend EPA-led workshops and/or webinars that provided information about the spray coating rule and hazardous waste management. Only about 15 percent of the shops in the treatment group sample in Massachusetts in 2010 attended these workshops/webinars (12 of the 79 shops selected).

Exhibit 3-5 compares the environmental performance of shops that chose to attend the workshops/webinars with shops that did not. Of the 18 performance measures with sufficient data for comparison, five of the differences were statistically significant (one-sided test), and the magnitude of several of the differences was quite large. Specifically, the difference for drums_closed was significant at the 1 percent level (workshop/webinar group performance was 44.1 percentage points higher), the difference for drums_labeled was significant at the 5 percent level (workshop/webinar group performance was 32.7 percentage points higher), the difference for filter_good was significant at the 1percent level (workshop/webinar group performance was 31.9 percentage points higher), the difference for cleaning_compliant was significant at the 5 percent level (workshop/webinar group performance was 16.9 percentage points higher), and the difference for avoid_mecl was significant at the 5 percent level (workshop/webinar group performance was 6.1 percentage points higher). As the workshops/webinars were voluntary, impacts associated with compliance assistance efforts may be conflated with self-selection bias. That is, the shops that chose to participate in the workshops/webinars may have been systematically more inclined to improve their performance than those that did not choose to participate, with or without the compliance assistance.

EXHIBIT 3-5. AVERAGE PERFORMANCE FOR TREATMENT GROUP SHOPS THAT ATTENDED WORKSHOP/WEBINAR VERSUS THOSE THAT DID NOT

CATEGORY	PERFORMANCE MEASURE [A]	TYPE OF MEASURE	ATTENDED WORKSHOP/WEBINAR		DID NOT ATTEND WORKSHOP/WEBINAR		DIFFERENCE [B]	MARGIN OF ERROR FOR DIFFERENCE
			PERCENT	N	PERCENT	N		
SPRAY BOOTH	**booth_exists**	Air – Spray Booth	100.0%	12	100.0%	67	0.0%	±0.0%
	not_outside	Air – Spray booth	100.0%	12	100.0%	67	0.0%	±0.0%
	booth_enclosed	Air – Spray booth	100.0%	10	98.6%	67	1.4%	±2.3%
	booth_ventilated	Air – Spray booth	90.1%	12	96.9%	64	-6.8%	±14.6%
	filter_exists	Air – Spray booth	90.1%	12	98.5%	66	-8.4%	±14.4%
	filter_good	**Air - Spray booth**	**86.8%**	**11**	**54.9%**	**64**	**31.9%*** **	**±20.5%**
	capture98	Air – Spray booth	65.1%	6	87.6%	7	-22.5%	±36.6%
PREP STATION	**prep_enclosed**	Air – Prep Station	--[C]	4	63.9%	9	--[C]	--[C]
	prep_vent	Air – Prep Station	--[C]	4	93.6%	7	--[C]	--[C]
MIXING ROOM	**mixroom_enclosed**	Air – Mixing Room	80.1%	7	95.7%	47	-15.6%	±23.6%
	mixroom_vent	Air – Mixing Room	80.1%	7	70.9%	42	9.2%	±25.7%
PAINT STRIP	**avoid_mecl**	**Air - Paint Stripping**	**100.0%**	**12**	**93.9%**	**67**	**6.1%** **	**±4.8%**
SPRAY GUNS	**guns_compliant**	Air – Spray Guns	100.0%	12	100.0%	66	0.0%	±0.0%
	cleaning_compliant	**Air - Spray Guns**	**90.1%**	**12**	**73.1%**	**67**	**16.9%** **	**±16.7%**
	train_records	Air – Spray Guns	57.5%	12	54.0%	67	3.6%	±25.5%
WASTE MGMT	**drums_labeled**	**Waste Management**	**57.5%**	**12**	**24.8%**	**65**	**32.7%** **	**±25.1%**
	drums_closed	**Waste Management**	**92.5%**	**12**	**48.3%**	**66**	**44.1%*** **	**±15.4%**
	rags_closed	Waste Management	40.0%	8	21.7%	51	18.3%	±31.4%
	no_spills	Waste Management	82.5%	12	80.7%	67	1.8%	±19.9%
	waste_doc	Waste Management	75.0%	12	74.5%	67	0.5%	±22.5%

Notes:

[a] See Exhibit 2-5 for definitions of performance measures.

[b] ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (one-sided test). These measures are in bold text.

[c] Stratified estimates could not be developed due to the existence of sampling strata without any shops.

CHAPTER 4 | LONG-TERM EFFECTIVENESS OF EPA COMPLIANCE ASSISTANCE

OVERVIEW

This chapter describes findings from the long-term quasi-experiment which was designed to evaluate the effectiveness of a compliance assistance package that EPA Region 1 offered to auto body shops in Massachusetts. While the short-term experiment described in Chapter 3 was designed to assess the impact of compliance assistance after several months, the long-term experiment was designed to assess changes in behavior that can be observed up to one year and nine months after receiving assistance.

EPA measured the long-term impact of the compliance assistance through a quasi-experiment involving the full study population of Massachusetts auto body facilities and a similar population of auto body facilities in the Piedmont/Tidewater regions of Virginia. EPA Region 1 offered the facilities in Massachusetts a package of compliance assistance, including of mailings, workshop/webinar opportunities, and on-site assistance. The facilities in Virginia did not receive compliance assistance from EPA or the state. The impact of compliance assistance is assessed primarily by comparing the *change* in performance in Massachusetts (where a portion of the change is potentially due to EPA compliance assistance) with the *change* in performance in Virginia (where no EPA compliance assistance was provided). In other words, the study uses a "difference-in-differences" approach to assess the impact of the compliance assistance. To establish the change in performance, site visits were conducted at a random sample of auto body shops in each state in 2010 and again in 2011.

The findings from the analysis of the long-term experiment are as follows:

- This study suggests that overall impact of EPA assistance was minimal for the performance measures evaluated in the long-term experiment. After controlling for shop characteristics that could influence performance, three of the seventeen performance measures showed statistically significant, positive differences-in-differences, indicating a potential impact associated with compliance assistance for these measures. However, the seventeen performance measures were approximately evenly split between negative (larger improvements in Virginia) and positive (larger improvements in Massachusetts) difference-in-differences. In addition, a one-tailed hypothesis test was used, which lowers the threshold for detecting a statistically significant impact.

- As expected, both Massachusetts and Virginia showed improvements in performance over time. Four out of 17 performance measures showed statistically significant improvements between 2010 and 2011 in Massachusetts, with differences ranging from 10 to 24 percentage points. Similarly, four out of seventeen performance measures showed statistically significant improvements between 2010 and 2011 in Virginia, with differences ranging from 12 to 28 percentage points. However, the study does not provide strong evidence that the improvement was greater in Massachusetts, where EPA offered assistance.

The remainder of this chapter discusses the findings in detail, including a summary of characteristics of the auto body shops in long-term treatment and comparison groups, environmental performance trends in Massachusetts and Virginia, and a comparison of comparison of shops that received interactive compliance assistance versus shops that did not.

CHARACTERISICS OF AUTO BODY SHOPS IN THE TREATMENT AND COMPARISON GROUPS

Important characteristics of the visited auto body shops are summarized in Exhibit 4-1 for Massachusetts and Virginia in 2010 and 2011. The vast majority of the shops were small, independent operations, with approximately 7 percent to 13 percent reporting that they were part of a corporate chain and with an average of only approximately 5 to 10 paint jobs completed per week. In general, the Virginia shops appear to be somewhat larger than the Massachusetts shops and perform a greater number of paint jobs. For example, approximately 12 percent of Virginia shops reported that they were part of a corporate chain, compared to approximately 8 percent of Massachusetts shops that were part of a corporate chain. In addition, Virginia shops reported performing an average of approximately 10 paint jobs per week versus only 6 jobs per week in Massachusetts. On the other hand, approximately 80 percent to 100 percent of Virginia shops classified themselves as Very Small Quantity Generators (VSQGs) of hazardous waste, versus only approximately 45 percent to 65 percent for Massachusetts shops. These differences may indicate that the Virginia shops are more specialized than the Massachusetts shops, with a larger number of spray coating jobs but with less work in other auto body areas that generate hazardous wastes. It may also indicate that more Virginia shops have shifted to waterborne paints than in Massachusetts.

When asked how they obtain information about how to comply with state and federal regulations, the three most frequent responses were suppliers, EPA, and trade associations. Suppliers were cited far more often than any other source, with approximately 85 percent of the shops in both states citing suppliers. In Massachusetts, approximately 23 percent of the shops cited EPA (19 percent in 2010 and 26 percent in 2011) while only about 6 percent cited EPA in Virginia. Finally, approximately 15 percent of the Massachusetts shops and 12 percent of the Virginia shops cited trade associations. Multiple responses were allowed, so the percentages for the various information sources do not sum to 100.

It is important to note that the long-term experiment attempts to make inferences about the impact of compliance assistance by examining differences in the *rates of change* between the two states, so that minor differences between the two states with regard to shop characteristics would not compromise the study findings. However, if the characteristics of the sampled shops changed substantially from one year to the next *within* a given state, then the experiment could potentially conflate this change with compliance assistance impacts. Thus, for example, the rather large changes in the percentage of shops that are VSQGs may be cause for concern (Massachusetts shops increased from 46.3 percent in 2010 to 66.8 percent in 2011, while Virginia shops decreased from 98.9 percent in 2010 to 81.4 percent in 2011). These differences may be the result of actual shifts in the composition of auto body shops in the two states, sampling variability, or interviewer effects. The regression analysis controls for these differences and also for the less substantial changes observed in the number of paint jobs per week and in the percentage of shops that are part of a corporate chain.

EXHIBIT 4-1. CHARACTERISTICS OF AUTO BODY SHOPS IN MASSACHUSETTS AND VIRGINIA (2010/2011)

CHARACTERISITIC	PERCENTAGE OF SHOPS IN MASSACHUSETTS		PERCENTAGE OF SHOPS IN VIRGINIA	
	2010	2011	2010	2011
Part of a corporate chain	8.8%	6.8%	13.2%	10.5%
VSQG	46.3%	66.8%	98.9%	81.4%
SQG	53.7%	31.9%	1.1%	17.4%
LQG	0.0%	1.4%	0.0%	1.2%
Receive information on how to comply with federal and state environmental regulations from:				
• Suppliers	86.9%	85.2%	84.6%	83.8%
• Corporate environmental division	1.2%	0.0%	11.0%	0.0%
• Educational institutions (e.g., vocational technical school)	2.4%	6.5%	1.1%	0.0%
• Environmental consultant	3.3%	7.9%	6.6%	2.5%
• Other auto body shops	2.3%	9.3%	4.4%	6.3%
• Trade association	16.1%	13.4%	1.1%	22.5%
• Local government	1.1%	14.4%	5.5%	3.8%

CHARACTERISITIC	PERCENTAGE OF SHOPS IN MASSACHUSETTS		PERCENTAGE OF SHOPS IN VIRGINIA	
	2010	2011	2010	2011
• Occupational Safety and Health Administration (OSHA)	0.0%	5.1%	5.5%	5.0%
• State environmental agency	5.8%	6.0%	4.4%	0.0%
• U.S. EPA	19.7%	26.3%	6.6%	6.3%
• Other sources	9.4%	10.7%	22.0%	20.0%

ENVIRONMENTAL PERFORMANCE TRENDS IN MASSACHUSETTS AND VIRGINIA

This section compares the environmental performance trends in Massachusetts and Virginia using the performance measures described in Chapter 2. The performance measures are a set of shop characteristics that (1) were potentially impacted by compliance assistance efforts and (2) could be independently verified through site visits. Three performance measures that were included in the short-term experiment were dropped from the long-term quasi-experiment due to differences in state regulations related to hazardous wastes (drums_closed, drums_labeled, waste_doc). The performance trends (i.e., differences between 2010 and 2011 performance) are compared first through a simple assessment of group means and then through a regression-based approach that compares trends while controlling for shop characteristics.

Comparison of Group Means

The mean performance levels for Massachusetts and Virginia in 2010 and 2011 are presented in Exhibit 4-2. The baseline (i.e., 2010) performance levels were generally quite similar in Massachusetts and Virginia for the majority of the 17 performance measures. However, there were four measures where baseline performance levels in Virginia were substantially lower than in Massachusetts: filter_good (40.3 percent in Virginia versus 59.3 percent in Massachusetts), prep_enclosed (81.3 percent in Virginia versus 95.4 percent in Massachusetts), avoid_mecl (72.4 percent in Virginia versus 91.1 percent in Massachusetts), and cleaning_compliant (45.1 percent in Virginia versus 78.8 percent in Massachusetts). [37]

[37] See Exhibit 2-5 for definitions of performance measures.

Of the 17 performance measures evaluated, four of the measures showed statistically significant improvements from 2010 to 2011 in Massachusetts using a one-sided hypothesis test[38]: filter_good improved by 9.7 percentage points, capture98 improved by 17.0 percentage points, train_records improved by 12.5 percentage points, and rags_closed improved by 23.5 percentage points. For three of these measures, statistically significant improvements were also observed in the Virginia samples: filter_good (27.8 percentage point improvement), capture98 (12.0 percentage point improvement), and train_records (14.1 percentage point improvement). In addition, a statistically significant improvement in avoid_mecl was observed in Virginia (15.6 percentage point improvement).

When the Virginia performance change is subtracted from the Massachusetts performance change, the resulting difference-in-differences is only statistically significant for one of the performance measures, mixroom_enclosed. For mixroom_enclosed, there was a 4.0 percentage point increase in Massachusetts and a 6.6 percentage point decline in Virginia, leading to a difference-in-differences of 10.6 percentage points.

[38] One-sided hypothesis tests are appropriate when the researcher has a strong *a priori* expectation that the result can only go in one direction and when a result in the opposite direction is considered functionally equivalent to no difference at all. The practical impacts of using a one-sided hypothesis test are (1) outcomes where Virginia performance improvement was greater than Massachusetts performance improvement are never considered statistically significant and (2) outcomes where Massachusetts performance improvement is greater than Virginia performance improvement are more likely to be classified as statistically significant than under a two-sided test. As discussed in the Information Collection Request submission to the Office of Management and Budget (Appendix E), the research team chose to use one-sided hypothesis tests *before* conducting the experiment.

EXHIBIT 4-2. COMPARISON OF AVERAGE PERFORMANCE FOR MASSACHUSETTS AND VIRGINIA IN 2010 AND 2011

CATEGORY	PERFORMANCE MEASURE [A]	MASSACHUSETTS 2010	MASSACHUSETTS N	MASSACHUSETTS 2011	MASSACHUSETTS N	MASSACHUSETTS CHANGE [B]	VIRGINIA 2010	VIRGINIA N	VIRGINIA 2011	VIRGINIA N	VIRGINIA CHANGE [B]	MASSACHUSETTS CHANGE MINUS VIRGINIA CHANGE [B]	MARGIN OF ERROR	REGRESSION-ADJUSTED DIFFERENCE IN DIFFERENCES [C,D]
SPRAY BOOTH	booth_exists	95.7%	90	95.0%	101	-0.7%	97.8%	91	96.5%	86	-1.3%	0.6%	6.5%	--
	not_outside	97.8%	88	95.2%	97	-2.5%	97.8%	89	97.6%	83	-0.2%	-2.4%	6.0%	--
	booth_enclosed	100.0%	86	98.1%	95	-1.9%	100.0%	89	100.0%	80	0.0%	-1.9%	2.2%	--
	booth_ventilated	98.9%	85	99.0%	93	0.1%	100.0%	89	98.7%	74	-1.4%	1.5%	3.2%	--
	filter_exists	100.0%	87	96.1%	94	-3.9%	94.4%	89	96.3%	80	1.9%	-5.8%	6.2%	--
	filter_good	59.3%	86	68.9%	90	**9.7%***	40.3%	77	68.1%	72	**27.8%*****	-18.2%	17.5%	-14.8%
	capture98	83.0%	20	100.0%	37	**17.0%****	88.0%	25	100.0%	20	**12.0%****	5.0%	18.6%	--
PREP STATION	prep_enclosed	95.4%	19	82.4%	18	-12.9%	81.3%	16	81.3%	16	0.0%	-12.9%	27.3%	--
	prep_vent	100.0%	16	100.0%	18	0.0%	93.8%	16	93.3%	15	-0.4%	0.4%	14.5%	--
MIXING ROOM	mixroom_enclosed	93.0%	58	97.0%	60	4.0%	98.3%	59	91.7%	60	-6.6%	**10.6%****	9.1%	**12.1%****
	mixroom_vent	83.5%	55	80.3%	58	-3.2%	79.7%	59	67.2%	58	-12.4%	9.2%	17.8%	**17.7%***
PAINT STRIP	avoid_mecl	91.1%	88	90.3%	94	-0.8%	72.4%	87	88.0%	75	**15.6%*****	-16.4%	12.3%	-13.7%
SPRAY GUNS	guns_compliant	100.0%	90	100.0%	100	0.0%	96.7%	91	96.5%	85	-0.2%	0.2%	4.5%	--
	cleaning_compliant	78.8%	89	75.3%	100	-3.5%	45.1%	91	45.2%	84	0.2%	-3.7%	16.0%	-0.4%
	train_records	45.7%	90	58.2%	101	**12.5%****	42.9%	91	57.0%	86	**14.1%****	-1.6%	17.1%	5.6%
WASTE MGMT	rags_closed	29.3%	69	52.8%	51	**23.5%*****	31.4%	35	40.0%	35	8.6%	14.9%	23.8%	**21.7%***
	no_spills	86.5%	90	92.2%	100	5.7%	92.3%	91	96.4%	84	4.1%	1.6%	9.3%	-0.7%

Notes:
[A] See Exhibit 2-5 for definitions of performance measures.
[B] ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (one-sided test).
[C] As discussed in the text, regression models were not estimated for performance measures with small sample sizes and/or with average performance levels near 100%.
[D] The regression-adjusted estimate was calculating by using the estimated regression coefficients to predict performance levels for Massachusetts in 2010/2011 and Virginia in 2010/2011, holding all other explanatory variables at their mean values.

Regression Analysis

While focusing on differences in performance *trends* between Massachusetts and Virginia helps to control for baseline differences between the two states, the impact of compliance assistance can be obscured if shop characteristics change between the two years due to (1) random variation associated with sampling or (2) changes in the population of shops. In order to control for these influences, the trend differences were evaluated within a multivariate regression context.

As all of the performance measures are binary (0/1) variables, logistic regression analysis was applied rather than standard linear regression techniques. The independent variables are summarized in Exhibit 4-3. The first three variables are used to assess the difference in performance trends between Massachusetts and Virginia. First, the "state_MA" variable captures the difference between Massachusetts and Virginia in 2010 (i.e., the baseline difference between the two states). Second, the "year_2011" variable captures the trend in Virginia between 2010 and 2011 (i.e., the improvement that the difference-in-differences approach assumes would have occurred in Massachusetts in the absence of compliance assistance). Finally, "MA_X_2011" is an interaction variable that captures the difference between Massachusetts and Virginia trends (i.e., the difference-in-differences). The remaining independent variables are shop characteristics that were expected to potentially be related to environmental performance, but that were unlikely to have been influenced by the treatment (i.e., by EPA compliance assistance).

Several performance measures were not included in the regression analysis due to inadequate sample sizes and/or average performance levels that were close to 100 percent. Having a preponderance of observations where the dependent variable is either zero or one frequently leads to estimation problems with logistic regression, particularly when sample sizes are small and explanatory variables are binary rather than continuous.

EXHIBIT 4-3. DEFINITIONS OF INDEPENDENT VARIABLES INCLUDED IN REGRESSIONS

VARIABLE NAME	TYPE	MEAN	DEFINITION
state_MA	binary	0.52	= 1 if the shop is in Massachusetts; = 0 if the shop is in Virginia
year_2011	binary	0.51	= 1 for interviews conducted in 2011; = 0 for interviews conducted in 2010
MA_X_2011	binary	0.27	= 1 if the shop is in Massachusetts and the interview was conducted in 2011; = 0 otherwise.
corp_chain	binary	0.10	= 1 if the shop is part of a corporate chain; = 0 otherwise
num_jobs	continuous	7.91	Number of paint jobs per week (estimated by respondent)
SQG	binary	0.27	= 1 if the shop is a small or large quantity generator of hazardous waste; = 0 otherwise.
aware_b2009	binary	0.05	= 1 if the shop learned about the EPA spray regulations before 2009; = 0 otherwise
non_EPAvisit	binary	0.26	= 1 if the shop was inspected or visited by a non-EPA government environmental or health and safety official within the last six months; = 0 otherwise
pvt_assist	binary	0.85	= 1 if the shop indicated that it obtains information about environmental compliance from coating manufacturers, suppliers, consultants, or a trade association; = 0 otherwise [A]

Notes:

[A] This variable is an indirect measure of private compliance assistance, as it only indicates that the shop generally obtains information about how to comply with environmental regulations from private sources. It does not indicate that the shop actually received compliance assistance recently from a private source.

Estimated regression coefficients are presented in Exhibit 4-4. Focusing first on the results associated with potential confounding factors (corp_chain, num_jobs, SQG, aware_b2009, nonEPA_visit, and pvt_assist), only two of the independent variables (num_jobs and SQG) had coefficients that were statistically significant for at least two of the performance measures. These two variables could be conceived as approximate measures of shop size or volume of work. The coefficient associated with num_jobs

was positive and significant at the 1 percent level in the model with filter_good as the dependent variable, and it was positive and significant at the 5 percent level in the model with mixroom_vent as the dependent variable. The coefficient associated with SQG was positive and significant at the 5 percent level in the model with cleaning_compliant as the dependent variable, and it was positive and significant at the 10 percent level in the models with mixroom_enclosed and train_records as the dependent variables.

The first three independent variables listed in Exhibit 4-4 (state_MA, year_2011, and MA_X_2011) were used to assess the difference in performance trends between Massachusetts and Virginia. The coefficient associated with state_MA was *positive* and significant in the models with cleaning_compliant and avoid_MeCl as dependent variables (1 percent and 5 percent significance levels, respectively). This indicates that baseline (i.e., 2010) performance was significantly higher in Massachusetts than in Virginia for these two performance measures. On the other hand, the coefficient associated with state_MA was *negative* and significant (at the 5 percent level) in the model with mixroom_enclosed as a dependent variable. This means that baseline performance was lower in Massachusetts than in Virginia for this performance measure. The coefficient associated with year_2011 was *positive* and significant (at the 10% significance level or lower) in three of the seven models, indicating that for these performance measures there was a general trend toward improved performance in Virginia between 2010 and 2011. However, the coefficient associated with year_2011 was negative and significant (at the 5% level) for mixroom_vent, indicating that performance declined for this performance measure in Virginia.

The results for the difference-in-differences variable (MA_X_2011) were mixed. The coefficient was positive and significant in two of the eight models: the models with mixroom_enclosed (5 percent significant level) and rags_closed (10 percent significance level) as dependent variables. This indicates that EPA compliance assistance in Massachusetts may have had a positive impact for these performance measures. However, it is important to note that a one-sided hypothesis test was used for the difference-in-differences variable, and positive differences are more likely to be classified as statistically significant when using a one-sided test rather than a two-sided test. The estimated coefficient was negative in three of the eight models estimated.

The difference-in-differences coefficients were used to estimate the regression-adjusted difference between the Massachusetts and Virginia performance trends (holding all other variables at their means), and these differences are reported in the final column of Exhibit 4-2. The regression-adjusted estimates of the impact of compliance assistance show that after controlling for shop characteristics that might influence performance, the Massachusetts performance improvement was significantly greater than the Virginia performance improvement for three measures: mixroom_enclosed (5 percent significance level), mixroom_vent (10 percent significance level), and rags_closed (10 percent significance level). Note that for only one of these measures did the pattern in performance trends match expectations: for rags_closed, both Massachusetts and Virginia shops improved, but Massachusetts shops improved more than Virginia shops.

In the case of mixroom_enclosed, Massachusetts shops' performance improved slightly, while Virginia shop performance declined. For mixroom_vent, both Massachusetts and Virginia shop performance actually *declined* between 2010 and 2011, but Massachusetts shops' performance did not decline as much as Virginia shops' performance.

The regression-adjusted estimates of the impact of compliance assistance were substantially higher than estimates based on group means for three of the performance measures (mixroom_vent, train_records, and rags_closed). For two of these measures (mixroom_vent and rags_closed) the regression-based estimates were statistically significant whereas the estimates based on group means were not. These results are likely due to the ability of the regression analysis to control for changes in the composition of the sampled shops between the two years. Specifically, in Massachusetts, there was a shift towards smaller shops between the 2010 and 2011 sample (fewer weekly paint jobs and more VSQGs), while in Virginia the opposite trend was observed. This shift was not deliberate on the part of EPA, rather it was coincidental since shops were randomly selected for assessment via site visit/phone. As performance was generally better for larger shops, these trends tended to diminish the estimated impact of compliance assistance in the simple comparison of group means.[39]

[39] For the long-term experiment, we cannot control for any potential interviewer effects because all of the 2010 shop visits in Virginia were conducted by the same individual, and this individual did not conduct any shop visits in either 2011 or in Massachusetts. Thus, the dataset cannot be used to assess the extent to which this individual's interpretations of shop conditions may have differed from those of other interviewers.

EXHIBIT 4-4. LOGISTIC REGRESSION COEFFICIENTS (Z-STATISTICS IN PARENTHESES)[A,B]

INDEPENDENT VARIABLE	DEPENDENT VARIABLE (PERFORMANCE MEASURE)[C]							
	FILTER GOOD	MIXROOM ENCLOSED	MIXROOM VENT	AVOID MECL	CLEANING COMPLIANT	TRAIN RECORDS	RAGS CLOSED	NO SPILLS
state_MA	0.78 (1.96)	**-2.48**** (-2.14)	0.06 (0.11)	**1.13**** (2.08)	**1.36*** (3.52)	-0.18 (-0.49)	-0.37 (-0.66)	-0.29 (-0.52)
year_2011	**1.06*** (2.83)	-1.83 (-1.63)	**-0.99**** (-2.10)	**0.93**** (2.00)	-0.05 (-0.16)	**0.55*** (1.73)	0.35 (0.66)	0.90 (1.24)
MA_X_2011	-0.57 (-1.14)	**3.07**** (2.15)	0.88 (1.24)	-1.04 (-1.40)	-0.05 (-0.09)	0.23 (0.50)	**0.95*** (1.36)	-0.38 (-0.44)
corp_chain	-0.08 (-0.17)	-0.28 (-0.29)	1.01 (1.26)	0.63 (0.87)	0.59 (1.36)	-0.19 (-0.49)	-0.64 (-1.03)	-0.10 (-0.18)
num_jobs	**0.05*** (2.61)	-0.01 (-0.41)	**0.04**** (1.99)	-0.01 (-0.74)	0.01 (0.96)	0.02 (1.46)	0.03 (1.36)	0.02 (0.67)
SQG	0.10 (0.32)	**1.65*** (1.83)	0.39 (0.90)	0.44 (0.88)	**0.84**** (2.52)	**0.57*** (1.95)	0.38 (0.94)	-0.53 (-1.19)
aware_b2009	**1.08*** (1.71)	--[d] --	1.42 (1.09)	0.63 (0.55)	-0.46 (-0.65)	-0.40 (-0.76)	-0.42 (-0.66)	-0.35 (-0.45)
nonEPA_visit	-0.24 (-0.86)	-0.16 (-0.23)	0.52 (1.31)	-0.43 (-1.20)	**0.65**** (2.27)	-0.03 (-0.12)	0.05 (0.12)	0.27 (0.55)
pvt_assist	0.21 (0.62)	1.05 (1.53)	0.05 (0.10)	-0.41 (-0.80)	0.19 (0.61)	**1.26*** (3.46)	0.75 (1.08)	0.08 (0.15)
constant	**-0.98**** (-2.40)	**3.43**** (2.55)	0.71 (1.15)	**1.49*** (2.89)	**-0.77**** (-2.07)	**-1.53*** (-3.53)	**-1.68**** (-2.08)	**2.19*** (3.50)
n	317	231	224	336	356	360	185	357

Notes:

[A] ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (one-sided test for "MA_X_2011" and two-sided test for all other variables).

[B] Design-based sampling weights were used in estimation.

[C] See Exhibit 2-5 for definitions of performance measures.

[D] The aware_b2009 variable was omitted from this regression due to quasi-complete separation: all ten shops that were aware of the spray coating regulations before 2009 also had an enclosed mixing room. Unique maximum likelihood estimates do not exist for this variable.

COMPARISON OF SHOPS THAT RECEIVED INTERACTIVE COMPLIANCE ASSISTANCE VERSUS SHOPS THAT DID NOT

An important part of the compliance assistance offered to shops in Massachusetts was the opportunity to attend EPA-led workshops/webinars and the opportunity to have an EPA employee or contractor provide customized compliance assistance during a site visit. Only about 18 percent of the Massachusetts 2011 sample received this interactive EPA compliance assistance (18 of the 101 shops selected).

Exhibit 4-5 compares the long-term environmental performance of shops that received interactive compliance assistance with shops that did not (as measured in 2011). Four of the 20 performance measures evaluated had differences between the two groups that were statistically significant (one-sided test). Specifically, the difference for booth_enclosed was significant at the 10 percent level (performance for shops with intensive assistance was 2.6 percentage points higher), the difference for filter_exists was significant at the 5 percent level (performance for shops with intensive assistance was 5.1 percentage points higher), the difference for mixroom_enclosed was significant at the 10 percent level (performance for shops with intensive assistance was 3.9 percentage points higher), and the difference for drums_labeled was significant at the 1 percent level (performance for shops with intensive assistance was 34.5 percentage points higher). Of these four differences, the 34.5 percentage-point difference for drums_labeled is probably the only one that is practically significant, as the differences are five percentage points or lower. As the interactive form other three of EPA's compliance assistance was voluntary, impacts associated with compliance assistance efforts may be conflated with self-selection bias. That is, the shops that chose to participate in the workshops/webinars or receive a site visit may have been systematically more inclined to improve their performance than other shops.[40]

[40] While we do not have evidence to show that self-selection bias was occurring, we also do not have evidence to disprove it.

EXHIBIT 4-5. AVERAGE PERFORMANCE FOR MASSACHUSETTS SHOPS THAT RECEIVED INTERACTIVE ASSISTANCE VERSUS THOSE THAT DID NOT (2011)

CATEGORY	PERFORMANCE MEASURE [A]	TYPE OF MEASURE	RECEIVED INTENSIVE ASSISTANCE		DID NOT RECEIVE INTENSIVE ASSISTANCE		DIFFERENCE [B]	MARGIN OF ERROR FOR DIFFERENCE
			PERCENT	N	PERCENT	N		
SPRAY BOOTH	**booth_exists**	Air – Spray Booth	94.4%	18	95.2%	83	-0.7%	9.7%
	not_outside	Air – Spray booth	88.2%	17	97.5%	80	-9.3%	13.2%
	booth_enclosed	**Air - Spray booth**	**100.0%**	**17**	**97.4%**	**78**	**2.6%***	**2.9%**
	booth_ventilated	Air – Spray booth	100.0%	16	98.7%	77	1.3%	2.1%
	filter_exists	**Air - Spray booth**	**100.0%**	**16**	**94.9%**	**78**	**5.1%****	**4.1%**
	filter_good	Air – Spray booth	78.6%	14	65.8%	76	12.8%	20.1%
	capture98	Air – Spray booth	100.0%	8	100.0%	29	0.0%	0.0%
PREP STATION	**prep_enclosed**	Air – Prep Station	50.0%	4	92.9%	14	-42.9%	42.7%
	prep_vent	Air – Prep Station	100.0%	4	100.0%	14	0.0%	0.0%
MIXING ROOM	**mixroom_enclosed**	**Air - Mixing Room**	**100.0%**	**9**	**96.1%**	**51**	**3.9%***	**4.5%**
	mixroom_vent	Air – Mixing Room	88.9%	9	77.6%	49	11.3%	19.8%
PAINT STRIP	**avoid_mecl**	Air – Paint Stripping	88.2%	17	90.9%	77	-2.7%	13.9%
SPRAY GUNS	**guns_compliant**	Air – Spray Guns	100.0%	18	100.0%	82	0.0%	0.0%
	cleaning_compliant	Air – Spray Guns	66.7%	18	78.1%	82	-11.4%	19.8%
	train_records	Air – Spray Guns	66.7%	18	55.4%	83	11.2%	20.4%
WASTE MGMT	**drums_labeled**	**Waste Management**	**61.1%**	**18**	**26.6%**	**79**	**34.5%*****	**20.6%**
	drums_closed	Waste Management	72.2%	18	59.8%	82	12.5%	19.5%
	rags_closed	Waste Management	50.0%	10	53.7%	41	-3.7%	29.0%
	no_spills	Waste Management	94.4%	18	91.5%	82	3.0%	10.2%
	waste_doc	Waste Management	61.1%	18	72.0%	82	-10.8%	20.6%

Notes:

[A] See Exhibit 2-5 for definitions of performance measures.

[B] ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (one-sided test). These measures are in bold text.

CHAPTER 5 | TELEPHONE SURVEY VALIDITY

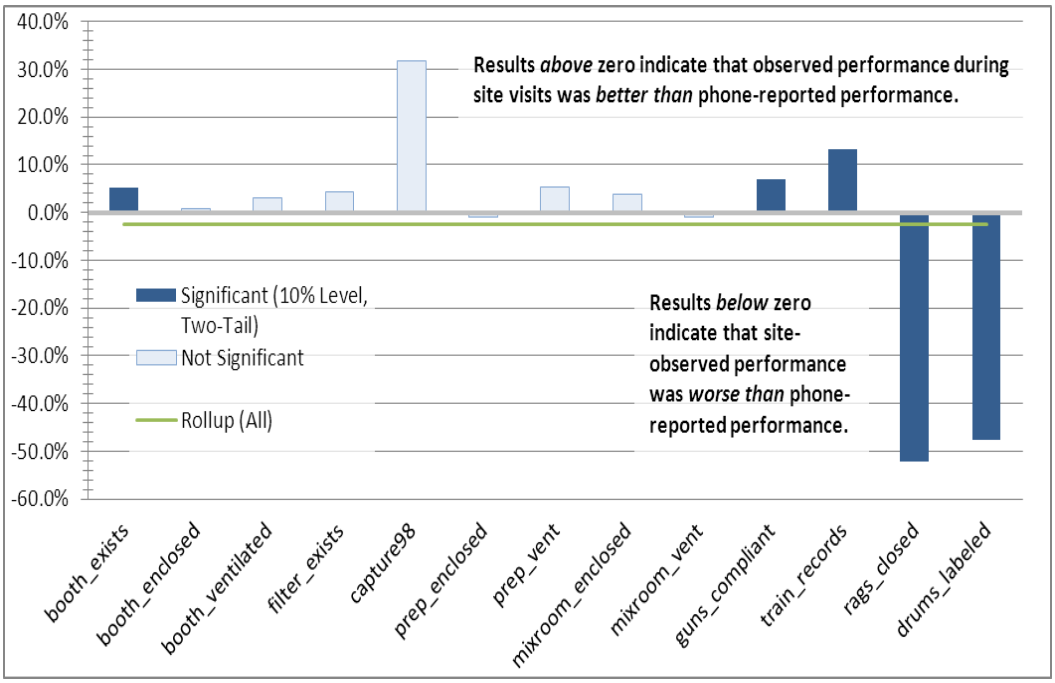
The phone survey accuracy component of the study was designed to test whether phone surveys are a reliable source of information about auto body shop performance. EPA frequently relies on telephone surveys to gather information about environmental performance, so these findings may inform the Agency's future data collection efforts. The analysis compares performance data collected through telephone surveys and site visits to assess the validity of phone survey data, accounting for both self-reporting bias and non-response bias. (See Chapter 2 for a detailed description of the measurement and analytic approach.)

The findings from the analysis of the phone survey validity are as follows:

- This study finds that, while the phone survey results were similar to the site visit results for the majority of the performance measures examined, very large differences were observed for several performance measures. The differences in performance as measured on-site and through telephone surveys are statistically significant for five of 13 measures (shown in dark blue on Exhibit 5-1). For three of these measures, observed performance during site visits is better than expected based on phone surveys; for two of these measures observed performance during site visits is worse than expected based on phone surveys. The largest differences were observed for performance measures related to storing used paint applicators and labeling drums. For these two measures, the performance reported in the phone survey was approximately 50 percentage points higher than the performance observed during the site visits.
- A "rollup" measure was developed to provide an overall indication of performance across all 13 measures. The difference between the site visit data and the phone survey data for the rollup measure was not statistically significant. It is important to keep in mind, however, that the rollup measure behaves like an average. As a result, to the extent that for some measures observed performance during site visits was *better than* phone reported performance, and for other measures it was *worse*, these effects cancel each other out, and the roll-up measure does not show a significant difference between site visit and phone survey data.
- The study finds that a greater number of measures showed self-reporting bias than showed non-response bias. These findings are somewhat different than reported in the literature, and may merit further exploration.

Exhibit 5-1summarizes the difference in performance levels on measures reported from both the phone survey and the site visits (which included phone respondents and non-respondents). The remainder of this chapter presents a review of literature on telephone survey validity and a detailed description of findings from this study.

EXHIBIT 5-1. DIFFERENCE BETWEEN ON-SITE AND PHONE-REPORTED RESULTS



FINDINGS FROM RELEVANT LITERATURE

The literature review focused on the following question: *What studies have been conducted to assess the reliability and/or validity of phone surveys as a way to understand behaviors, in the context of requirements?* The sections below summarize the literature search methods and findings. Appendix G presents a more thorough description of both, along with an annotated bibliography.

Literature Search Methods

The review involved a thorough search of databases and journals, as well as relevant academic, professional and government institutions**.** The review identified a mix of sources covering both theoretical discussions and experimental comparisons of different survey modes, most material either not recent or not directly germane to the question. Most of the theoretical articles were written the late 1970s and 1980s, a time when telephone surveying techniques were becoming a much more popular alternative to face-to-face interviews. Among experimental studies that measured differences between the modes, only one focused on compliance or compulsory behaviors. This study, conducted by the U.S. EPA's Office of Compliance, compared compliance and facility behaviors

from a mailed survey to an on-site survey. Most of the experimental studies examined opinion surveys and surveys that attempt to measure behaviors of individuals. However, these studies do provide important lessons about the validity of different survey approaches by trying to verify the reported information from different survey modes.

Summary of Findings

Overall, the literature indicates that face-to-face surveys gather better quality data than telephone surveys, but these differences are often small. This finding supports the conclusion that telephone surveys can provide accurate survey data. However, the literature is relatively sparse with regard to compliance-specific issues and focuses primarily on individuals, rather than facilities.

Studies have found that differences between the survey modes can often be corrected by thoughtful and creative survey design. For example, if a face-to-face interview includes a visual aid to help respondents answer a question about types of equipment used in a facility, a telephone interview obviously cannot include this same visual. Therefore, survey developers will have to come up with another way to communicate this question, such as careful description of the equipment (e.g., think of easily recognizable names for the types of equipment).

Both non-response bias and self-reporting bias play a role in the observed differences between the two survey modes:

- **Telephone surveys tend to have a lower response rate than face-to face interviews, which may result in relatively higher non-response bias.** Face-to-face interviews have a higher response rate (potentially reducing non-response bias, relative to phone surveys), while telephone respondents are more likely to not respond to individual questions, or to give more socially desirable responses.[41,42,43]
- **Self-reporting bias from telephone surveys appears to be small, except with regard to sensitive questions.** One study found study found that any differences in data accuracy between telephone surveys and face-to-face interviews were extremely small and statistically insignificant.[44] However, the similarity of responses between survey modes depends on the type of question

[41] Van der Zouwen, Johannes and de Leeuw, Edith D. "The Relationship Between Mode of Administration and Quality of Data in Survey Research." Bulletin of Sociological Methodology, Vol. 29, No. 3, 1990.

[42] De Leeuw, Edith Desiree. "Data Quality in Mail, Telephone and Face to Face Surveys." Netherlands Organization for Scientific Research, 1992.

[43] Bonnel, Patrick, and Le Nir, Michael. "The Quality of Survey Data: Telephone versus Face-to-Face Interviews." Transportation, Vol. 25, No. 2, May, 1998.

[44] Bonnel, Patrick, and Le Nir, Michael. "The Quality of Survey Data: Telephone versus Face-to-Face Interviews." Transportation, Vol. 25, No. 2, May, 1998.

(e.g., whether the question was "sensitive" or not).[45] For example, one study found that telephone respondents may be more likely to misreport their behaviors in telephone interviews (e.g., young adults were found to be more likely to underreport their smoking behaviors in a telephone interview than they were in a face-to-face interview).[46]

RESULTS OF PHONE SURVEY ACCURACY ANALYSIS

The sections below describe the findings from three different analyses conducted to explore the accuracy of phone surveys: 1) an overall comparison of performance estimates received from phone surveys and on-site surveys; 2) a comparison of observed on-site performance levels of phone respondents and non-respondents, to assess the contribution of non-response bias; and 3) a comparison of phone and site-visit results for shops that responded to both surveys, to explore the contribution of self-reporting bias.

Each of these comparisons relied upon the use of two-sided hypothesis tests to determine statistical significance. Two-sided tests are appropriate when the researcher does not have strong a priori expectations about the direction of the difference. In the case of phone survey accuracy, the two-sided test is justified because reasonable arguments could be made for expecting performance rates from the phone surveys to be higher or lower than performance rates based on site visits observations. For instance, one might expect phone respondents to over-report good performance, to appear to meet EPA expectations. On the other hand, it's conceivable that shops might improve their performance after the phone survey has made them aware of issues they need to resolve; in such cases, the phone survey performance would be lower than the performance observed during site visits.

Overall Accuracy of Phone Survey Data

This analysis compares the overall population performance estimates derived from each mode. The results speak to the primary interest of the study – the degree to which performance estimates based upon phone surveys are reliable. This approach takes into account that both self-reporting and non-response bias may influence accuracy of phone surveys. This analysis finds that:

- **On a summary measure of performance, there is no detectable difference between survey modes.** On a rollup measure summarizing the performance level of facilities across all 13 measures, observed performance at site visits was 77.8 percent, meaning that the average shop was achieving 77.8 percent of relevant performance measures. The performance level reported by shops over the phone was 80.1 percent, meaning that the phone survey, on the whole, over-reported

[45] Van der Zouwen, Johannes and de Leeuw, Edith D. "The Relationship Between Mode of Administration and Quality of Data in Survey Research." Bulletin of Sociological Methodology, Vol. 29, No. 3, 1990.

[46] Luepker, Russell V. et al. "Validity of Telephone Surveys in Assessing Cigarette Smoking in Young Adults." American Journal of Public Health, Vol. 79, No. 2, February 1989, pp. 202-204.

performance by an observed 2.3 percentage points. The difference is not statistically significant at the 10 percent level.

- **Phone surveys modestly under-reported performance levels on most measures.** Observed on-site performance was better than was estimated by phone surveys on nine of 13 measures, all air-related.[47] Differences on three of these measures – related to the existence of spray booths and spray gun compliance/training – were statistically significant at the 10 percent and 5 percent levels, respectively. The observed differences were not typically very large, with a median difference of 5.1 percentage points. The largest observed instance of phone survey under-reporting, 31.9 percentage points, was not statistically significant; it was for a measure of filter efficiency for spray booths, which had a small sample size in both telephone and on-site surveys.
- **Phone surveys substantially over-reported performance levels on two measures.** Phone-reported performance levels were better than measured on-site for four measures. Differences on the two waste-related measures – related to container closure and labeling – were substantial and statistically significant at the 1 percent level. It is important to note that the respective questions on each survey instrument related to container labeling may not have been sufficiently comparable. Whereas the phone survey only asked respondents if containers were labeled, the on-site survey asked interviewers to assess whether containers were *properly* labeled, and gave guidance on the appropriate determinants for that decision.

Exhibit 5-2 provides data on the overall comparison between performance as measured by phone surveys and site visits.

[47] This modest under-reporting of performance on most measures was more than balanced by substantial over-reporting of performance on two measures (discussed below).

EXHIBIT 5-2. OVERALL COMPARISON OF DATA FROM PHONE SURVEYS AND SITE VISITS

CATEGORY		PERFORMANCE			SAMPLE SIZE	
	PERFORMANCE MEASURE	PHONE	SITE VISIT	DIFFERENCE (SITE - PHONE)	PHONE	SITE VISIT
SPRAY BOOTH	**booth_exists**	**92.7%**	**97.8%**	**5.1%***	**80**	**169**
	booth_enclosed	98.6%	99.4%	0.8%	70	163
	booth_ventilated	94.4%	97.4%	3.0%	70	161
	filter_exists	94.3%	98.6%	4.3%	53	165
	capture98	51.6%	83.5%	31.9%	16	33
PREP STATION	prep_enclosed	87.3%	86.3%	-1.0%	22	32
	prep_vent	91.2%	96.6%	5.4%	24	27
MIXING ROOM	mixroom_enclosed	89.3%	93.1%	3.8%	48	112
	mixroom_vent	79.1%	78.0%	-1.1%	48	104
SPRAY GUNS	**guns_compliant**	**92.9%**	**100.0%**	**7.1%****	**80**	**168**
	train_records	**36.8%**	**50.0%**	**13.2%****	**74**	**169**
WASTE MGMT	**rags_closed**	**79.2%**	**27.0%**	**-52.2%*****	**48**	**128**
	drums_labeled	**79.8%**	**32.3%**	**-47.5%*****	**63**	**167**
ALL MEASURES	rollup	80.1%	77.8%	-2.3%	80	169

***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (two-sided test). These measures are in bold text.

Non-Response Bias

This analysis compares site visit data collected from phone survey respondents to that collected from phone survey non-respondents. This comparison relates to non-response bias, and assesses whether phone survey respondents are systematically different from non-respondents. Note that since this analysis is drawing on site visit data, there are more data available to make comparisons (since 20 performance measures were included on the site visit survey, compared to 13 on the telephone survey). Therefore, the rollup measure for this analysis is not directly comparable to the rollup measure for the analysis of overall accuracy or self-reporting bias.

This analysis of non-response bias finds that:

- **The rollup measure indicates that the performance of respondents was slightly lower than the performance of non-respondents.** The overall performance of respondents was 3.6 percentage points lower than the overall performance of non-respondents. This difference is statistically significant at the 10 percent level, although it is unlikely to be practically significant.

- **There was a significant difference between respondents and non-respondents for only one measure.** Non-respondent performance was a 21 percentage points higher with regard to proper ventilation of mixing rooms. This result is statistically significant at the 5 percent level.[48] Differences were not statistically significant on any other measure.

Exhibit 5-3 provides data on difference in performance for phone respondents and non-respondents, as measured by site visit observations.

EXHIBIT 5-3. COMPARISON OF ON-SITE DATA FOR PHONE SURVEY RESPONDENTS AND NON-RESPONDENTS

		PERFORMANCE			ON-SITE SAMPLE SIZE	
CATEGORY	PERFORMANCE MEASURE	PHONE RESPONDENT	PHONE NON-RESPONDENT	DIFFERENCE (RESP.-NON)	PHONE RESP.	PHONE NON-RESP.
SPRAY BOOTH	booth_exists	96.5%	98.3%	-1.9%	52	117
	not_outside	98.2%	99.2%	-1.0%	51	116
	booth_enclosed	97.9%	100.0%	-2.1%	49	114
	booth_ventilated	95.1%	98.2%	-3.0%	48	113
	filter_exists	97.2%	99.1%	-1.9%	50	115
	filter_good	54.0%	62.0%	-8.0%	48	113
	capture98	82.1%	80.0%	2.0%	13	20
PREP STATION	prep_enclosed	86.7%	87.2%	-0.6%	8	24
	prep_vent	84.9%	100.0%	-15.1%	7	20
MIXING ROOM	mixroom_enclosed	86.1%	96.3%	-10.2%	31	81
	mixroom_vent	**63.1%**	**84.1%**	**-21.0%****	**28**	**76**
SPRAY GUNS	guns_compliant	100.0%	100.0%	0.0%	52	116
	cleaning_compliant	74.9%	78.4%	-3.5%	52	116
	train_records	42.5%	53.2%	-10.7%	52	117
PAINT STRIP	avoid_mecl	92.8%	93.1%	-0.3%	51	116
WASTE MGMT	rags_closed	23.1%	28.3%	-5.2%	40	88
	no_spills	81.9%	84.6%	-2.7%	52	117
	drums_labeled	30.9%	33.0%	-2.1%	52	115
	drums_closed	57.5%	58.4%	-0.9%	52	116
	waste_doc	80.7%	71.6%	9.1%	52	116
ALL MEASURES	**rollup**	**75.2%**	**78.8%**	**-3.6%***	**52**	**117**

[48] In the overall comparison of results, there was no significant difference observed with regard to this measure; results of the analysis of self-reporting bias (presented later in this chapter) suggest that respondents may have been over-reporting performance levels on this measure, balancing out the non-response bias.

***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (two-sided test). These measures are in bold text.

Self-Reporting Bias

This analysis compares phone and site-visit results for shops that responded to both surveys. By including only respondents, the analysis can explore self-reporting bias by examining differences in performance estimates provided by the different modes. Self-reporting bias in phone surveys may occur for a variety of reasons, such as greater difficulty in understanding or answering survey questions, or greater reluctance to share accurate information. There may also be differences between results from the two different modes simply because data collection occurred at different points in time. This analysis finds that:

- **The rollup measure indicates that phone-reported performance was somewhat higher than performance observed during site visits.** Phone-reported overall performance levels were 6.6 percentage points higher than overall performance levels observed on site. This difference is statistically significant at the 5 percent level.
- **There are significant differences for three measures.** Phone surveys under-reported compliance with spray gun requirements by 6 percentage points, which is statistically significant at the 10 percent level. On the other hand, phone surveys substantially over-reported compliance with container closure and labeling requirements, by 53 and 46 percentage points, respectively. Both of those differences are statistically significant at the 1 percent level. Notably, levels of over-reporting on those two measures are nearly the same as in the comparison of overall results, suggesting that self-reporting bias is the overwhelming driver of the difference on these two measures.

Table 5-4 presents detailed findings of this comparison between phone survey data and site visit data for the set of shops that completed both surveys.

EXHIBIT 5-4. COMPARISON OF DATA FROM SHOPS COMPLETING TELEPHONE AND SITE VISIT SURVEYS

CATEGORY	PERFORMANCE MEASURE	PERFORMANCE			SAMPLE SIZE
		PHONE	SITE VISIT	DIFFERENCE (SITE - PHONE)	SHOPS WITH BOTH SURVEYS
SPRAY BOOTH	booth_exists	95.9%	96.4%	0.5%	51
	booth_enclosed	100.0%	97.7%	-2.3%	45
	booth_ventilated	92.9%	97.9%	4.9%	43
	filter_exists	94.3%	100.0%	5.7%	35
	capture98	61.0%	72.0%	11.0%	8
PREP STATION	prep_enclosed	84.9%	84.9%	0.0%	7
	prep_vent	82.6%	82.6%	0.0%	6
MIXING ROOM	mixroom_enclosed	87.4%	87.4%	0.0%	26
	mixroom_vent	82.5%	65.4%	-17.1%	24
SPRAY GUNS	**guns_compliant**	**94.1%**	**100.0%**	**5.9%***	**51**
	train_records	33.7%	44.2%	10.5%	50
WASTE MGMT	**rags_closed**	**83.6%**	**30.6%**	**-53.0%*****	**24**
	drums_labeled	**79.2%**	**33.6%**	**-45.6%*****	**42**
ALL MEASURES	**rollup**	**80.1%**	**73.6%**	**-6.5%****	**51**

***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (two-sided test). These measures are in bold text.

CHAPTER 6 | CONCLUSIONS

This study represented an ambitious effort by EPA to understand the effects of three sources of bias that may have affected the Agency's prior performance measurements. In particular, this study sought to account for the effects of *self-selection bias, non-response bias*, and *self-reporting bias* in measuring the effects of EPA compliance assistance on regulated industries. While the specific findings of the study are limited to the particular sector and context where performance measurement occurred, the broader lessons learned about what types of measurement are feasible are broadly instructive for EPA going forward. This chapter summarizes what the study finds for each of the evaluation questions, and implications for future work.

1. Did EPA Region 1's compliance assistance activities contribute to behavior change in the auto body sector?

This study does not provide evidence that EPA assistance to auto body shops affected sector-wide performance in the short-term. While it appears that EPA assistance may have had a positive effect on sector-wide performance in the long-term for a few measures (3 out of 17 measures), the statistical evidence for an impact is not entirely compelling. There are a number of potential explanations for the absence of evidence of an impact:

- The direct assistance provided by EPA simply may not have been effective in influencing the targeted population. It is possible that other approaches to providing information to auto body shops would be more effective, though the study does not suggest what, if any, changes to direct assistance should be made.
- In addition to providing assistance directly to auto body shops, EPA also provides information to vendors and suppliers, who in turn educate shops. This study did not measure the indirect effects of EPA assistance. It is possible that the indirect approach of influencing auto body shops is more effective than direct assistance from EPA because it channels information on compliance requirements and best practices through vendors and suppliers with whom shops already have a trusted relationship.
- Despite considerable outreach efforts by EPA Region 1, fewer than 20 percent of the shops in Massachusetts received interactive assistance during the study (i.e., workshops, webinars, or site visits). Thus, even if the interactive assistance was extremely effective for the shops that received it, the impact may be difficult to

detect when this small group of shops is pooled with the remainder of the auto body population.

- For many of the performance measures evaluated, baseline performance was high, leaving little room for performance improvement. The auto body sector in Massachusetts had been exposed to considerable government assistance efforts over the last few decades, which may have limited the impact of additional assistance.

All of these potential explanations are speculative. The study itself does not demonstrate why EPA's assistance did not have a substantial impact on the sector as a whole. Further research, particularly interviews with regulated entities, might offer insights as to what factors are most important in influencing their behavior.

2. Are the measurement methods employed in the pilot transferable to other assistance activities?

EPA realized at the start of the pilot study that the methodology would require considerable time and resources, and that it would not be possible to replicate the methodology in its entirety on a regular basis. Nevertheless, the Agency sought to identify what components of the methodology might be transferable. The study findings suggest that several measurement methods might be broadly useful and could be applied in future projects. In particular:

- It appears that obtaining representative data on baseline performance would be helpful in targeting assistance. For example, implementing on-site surveys at a subset of randomly selected shops would provide an initial gauge of performance, and compliance assistance could then be optimized. This approach is particularly relevant where the universe is not well characterized (e.g., information about performance is anecdotal). Other agencies have also found value in measuring baseline performance through statistical samples. For example, numerous states have conducted Environmental Results Programs, which begin with establishing a statistical baseline for sector performance, and then use information from the baseline to target assistance. Other federal agencies (e.g., the Department of Labor) are also using statistical baselines as a tool to understand compliance problems, design interventions, and test the impact of those interventions over time. While establishing a statistical baseline does require an investment of resources, it can save time and effort in the long term by pointing out where Agency attention is most needed. Moreover, sample sizes do not need to be large to approximate performance. For example, several ERP states have developed statistical baselines by sampling as few as 40 – 50 shops. While such a small sample may not offer a very precise picture of compliance, it can be sufficient to give a general picture of whether sector compliance is relatively high or low.

- Phone surveys might also be used to assess baseline performance at a reasonable cost, but this would require further study to better understand the factors impacting phone survey validity (see question 4 below).
- Delay of treatment to establish a control group could be broadly applied to test compliance assistance impacts. This approach (sometimes called a pipeline control group) has advantages in that it can allow agencies to randomly assign entities to receive assistance, while still ensuring that all facilities are ultimately offered assistance. This approach may be particularly relevant where demand for assistance outstrips EPA's capacity to offer assistance, and therefore some entities would have to wait to receive assistance regardless of any efforts to measure performance. In this situation, randomly assigning the facilities that wait to receive assistance can allow EPA to compare performance of those that did and did not receive assistance and use this comparison to understand the effectiveness of assistance. However, this approach is necessarily limited by the length of delay that is considered tolerable. If EPA could reasonably offer assistance to all entities at one time (e.g., if the planned assistance is a webinar all entities could access simultaneously), then it may not be reasonable to prevent some entities from accessing the assistance until after measurement has occurred.

3. What specific characteristics of the auto body sector influence the transferability of the measurement approach in this evaluation?

The auto body sector is characterized by many small businesses, businesses often opening or going out of business (business turnover), and businesses that are not registered with state agencies (informal businesses). These characteristics present particular challenges for measuring sector performance, and these challenges may also hold true for other similar sectors. For example:

- It is more difficult to draw statistically-based samples in sectors with high business turnover. For example, this study encountered considerable "list problems" (i.e., shops that were listed on in business databases, such as InfoUSA or Dunn and Bradstreet, turned out to be not auto body shops, to be out of business, or to have moved.) These list problems may plague other studies trying to gauge performance in sectors dominated by small businesses, rapid growth, or changes in business entities.
- In sectors with a considerable number of informal businesses, alternative approaches are needed to identify entities that are not licensed. For example, a cluster sampling approach could be used within urban areas: neighborhoods could be randomly selected, then all businesses operating within the selected neighborhoods would be visited.

4. Is the telephone survey a valid and reliable technique for performance measurement and program evaluation?

Overall, this study found the phone survey to be fairly accurate for most measures. However, the study found high levels of inaccuracy on a small number of questions. The study does not provide enough information to predict what types of questions are more likely to be inaccurate than others. Further research would be needed to test a variety of types of questions in a variety of sectors to better understand factors that influence the accuracy of phone surveys.

The study assessed two potential sources of bias in phone survey results: non-response bias and self-reporting bias. The study found very little non-response bias in the phone survey results. This was a surprising finding, and it differs somewhat from results reported in the literature. It would be helpful to track if future studies of phone survey accuracy verify this finding. This study found that self-reporting bias was a more substantial source of inaccuracy. This could be due to facilities not understanding the survey question over the phone, intentionally providing inaccurate information, or changing their behavior between the time of the phone survey and the follow-up site visit.

This study suggests a few implications for future measurement and compliance assistance efforts. In particular, in future measurement efforts, EPA may wish to consider developing a streamlined list of performance measures that can be independently verified, and using those for measurements at the baseline and over time. EPA could also combine on-site assistance and baseline measurement for future pilots. In future compliance assistance efforts, EPA could consider 1) conducting baseline assessments to assess performance before launching an assistance effort, 2) conducting phone surveys to understand extent of reliance on private parties for assistance, before investing in direct EPA assistance, and 3) focusing on outreach to suppliers to disseminate EPA's accurate compliance information.