DISCUSSION PAPER

March 2011 ■ RFF DP 11-09

Attributing Benefits to Voluntary Programs in EPA's Office of Resource Conservation and Recovery

Challenges and Options

James Boyd and Cynthia Manson

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org

RESOURCES FOR THE FUTURE

Attributing Benefits to Voluntary Programs in EPA's Office of Resource Conservation and Recovery: Challenges and Options

James Boyd and Cynthia Manson

Abstract

This paper reviews the economic justification for voluntary environmental programs to derive defensible measures of their positive social outcomes. We consider ideal experimental and statistical designs to detect and attribute benefits. We also explore a set of more practical approaches to benefit attribution that take into account the data gaps and statistical challenges that often make more rigorous approaches infeasible.

Key Words: voluntary programs, cost–benefit assessment, program evaluation

JEL Classification Numbers: Q53, Q58, L15, O33

Contents

Attributing Benefits to Voluntary Programs in EPA's Office of Resource Conservation and Recovery: Challenges and Options

James Boyd and Cynthia Manson*

1. Introduction

For more than a decade, the U.S. Environmental Protection Agency's (EPA's) Office of Resource Conservation and Recovery (ORCR) has implemented public voluntary or partnership programs (referred to collectively as PVPs) as a collaborative, nonregulatory approach to encouraging pollution prevention, recycling, and input and waste reductions. As these efforts have gained in popularity in both ORCR and other EPA offices, the need to measure their impacts and benefits has emerged as a key challenge for ORCR and EPA. In particular, the White House Office of Management and Budget (OMB) has emphasized the need for ORCR, and EPA more broadly, to demonstrate defensible approaches for documenting voluntary program outcomes.

A key component of gauging program success is identifying and *attributing* specific positive outcomes (or benefits) to the initiative under consideration (Ferraro, 2009). In the specific context of ORCR's and EPA's voluntary programs, the goal of attribution analysis is to identify socially desirable changes in behavior that would not have occurred, or would not have occurred as widely or as quickly, in the absence of PVPs. The nature of voluntary programs presents a variety of challenges that often call for variations on the rigorous statistical approaches that analysts might otherwise pursue to assess attribution.

To respond to questions from OMB and other stakeholders, ORCR has investigated existing research on the attribution of voluntary program benefits, including the theoretical basis for voluntary program benefits, and a range of empirical and theoretical methods for assessing benefits. This effort, though initially focused on ORCR programs, has broader relevance to EPA partnership programs in all offices and is designed to address agencywide challenges.

Our study explores PVP attribution in several ways. First, the analysis reviews the economic justification for voluntary programs, highlighting the categories of programs that EPA most frequently administers. Second, the analysis considers experimental design and full-scale

* Boyd, Resources for the Future, boyd@rff.org; Manson, Industrial Economics, Incorporated. This research was funded by EPA's Office of Policy and Office of Solid Waste and Emergency Response.

statistical approaches to benefit attribution as well as the challenges that frequently make these rigorous approaches infeasible. Finally, the analysis examines a series of practical approaches for attributing benefits to PVPs. The objective is to identify three distinct analytical levels that progress from qualitative assessment of program design to quasi-experimental designs that can provide robust quantitative estimates of program impact.

2. Voluntary Programs and Their Economic Justification

From an economic standpoint, the case for PVPs (or any environmental regulation or policy), is strongest if the case can be made that the government activity is addressing a *market failure*—a situation in which the market forces alone to provide efficient environmental outcomes. PVPs are one category of policy intervention designed to compensate for market failures.

The general market failure that creates the need for all environmental policy interventions is the lack of adequate incentives for private parties—whether organizations or individuals—to contribute to the quantity or quality of shared, common-pool resources, or *public goods*, such as clean air, clean water, and uncontaminated land (see, for example, Hardin 1968; Ostrom 1990). A related motivation for policy intervention is the desire to address externalities—a situation in which a polluter or injurer imposes costs on others, including private property owners. These market failures are the fundamental justification for all environmental regulation, public investments, and protections. In the case of PVPs, the goal is to encourage behavioral change that goes beyond regulatory and other legal requirements to provide the shared public goods associated with environmental improvement (e.g., improvements in human health and ecological resources).

All environmental PVPs are designed to address this general failure of the market to adequately price the protection of public goods, such as human health and environmental quality. PVP design, however, typically focuses on two distinct market failures more specific to private sector users of resources and generators of waste. Depending on the type of program design, PVPs may address inefficient investments in environmental process, management, and product innovation.

- PVPs designed to encourage organizational and technical learning (which we call *technical assistance* PVPs) address the failure to invest in research and development (R&D) and technological innovation (e.g., because these are, in part, public goods).
- PVPs that provide *signaling features* to the market (e.g., suppliers and customers) address information imbalances, or *asymmetries*, associated with complex products and

processes; specifically, incomplete information can result in decisions that create negative environmental impacts, or externalities (see Darnall and Carmin 2005; Borck and Coglianese 2009).

Recent research has considered another potential failure: the failure of organizations to individually behave rationally in their own self-interest—that is, they fail to identify and act on profit-making opportunities, such as pollution prevention or other money-saving efforts (Sarokin et al. 1985). The explanations for why firms might not act in their own self-interest are various and include information barriers, accounting-based distortions, or inappropriate managerial incentive schemes. At the same time, so-called behavioral economics increasingly emphasizes the importance of psychological realities that conflict with purely rational maximizing behavior (e.g., Thaler and Sunstein 2008).

This paper focuses on two market failures—the public good nature of intellectual property and information asymmetry—to judge the case for government intervention. Arguably, organizational failures and behavioral economics should also be considered, though the literature on these areas is still evolving, and it is not clear that the evidence supports a broad case for government intervention. This paper does not, therefore, incorporate organizational failure or behavioral economics into a broad framework for assessing voluntary programs.

Importantly, because PVPs are voluntary, economists also consider whether these programs merely identify behavior that would have occurred without the program (as a result of existing market conditions).[1] It is important for PVPs, therefore, to clearly identify how they encourage behavioral change beyond that which would occur in the market. The discussion below describes in more detail these two types of PVPs and the market failures they address as well as implications for attributing benefits to such PVPs.

2.1 Technical Assistance PVPs and the Promotion of Innovation

Technical assistance PVPs, as discussed in this paper, include programs that promote information sharing and knowledge transfer regarding green production processes, organizational improvements that foster environmental efficiency, and environmentally preferred product innovations, as well as the financial benefits of these innovations to program participants. Examples of the technical assistance that PVPs offer to members include:

[1] This phenomenon is referred to as “crowding out,” where public investments reduce private investment (Spencer and Yohe 1970).

- technical guides and best-practice documents;
- hotline assistance;
- online technical resources;
- environmental impact calculators (e.g., carbon calculators);
- peer mentoring;
- publication of partner practices and success stories; and
- conferences, working groups, and teleconferences that facilitate knowledge transfer.

EPA sponsors several technical assistance PVPs. For example, ORCR's WasteWise program facilitates knowledge transfer between program partners through dedicated conferences. In addition, the office also sponsors an annual Resource Conservation Challenge (RCC) conference that provides an information-sharing venue for several specific programs, including the Coal Combustion Products Partnership (C2P2), WasteWise, the National Partnership for Environmental Priorities (NPEP), and a range of municipal and electronic waste efforts. The RCC conference provides opportunities to share best practices and successes among program participants and across programs. In addition, EPA's Office of Air and Radiation (OAR) sponsors the AgSTAR Program, which encourages the use of methane recovery (biogas) technologies at certain confined animal feeding operations. AgSTAR assists agricultural producers in adopting anaerobic digester technologies through conferences, a hotline, "how-to" project development tools, industry listings, and by conducting performance evaluations for digesters and conventional waste management systems. OAR also sponsors Natural Gas Star, which partners with the oil and natural gas industry to identify and promote the implementation of cost-effective technologies and practices to reduce emissions of methane. As part of Natural Gas Star, EPA works with companies to quantify emissions from across the natural gas system, facilitate technology transfer, and demonstrate successful greenhouse gas–reduction technologies.

Technical assistance PVPs address a market failure created by the fact that certain types of information, such as intellectual property and R&D, are often (but not always) viewed as a public good that benefits others—including competitors—as well as the firm undertaking the investment. Public goods have two traits that limit their value as investments for a single company. First, public goods are *nonrivalrous:* one's use of the good (clean air) does not reduce

its availability for others. More importantly, public goods are *nonexcludable*; that is, it is difficult or impossible to exclude others from using the good.[2] The value of organizational and technical information can be difficult for private parties to fully capture because, once created, information tends to be readily copied or used by others.

Because private parties operating in a free market may not be the only ones to use the innovation developed through R&D, they tend to underinvest in creating such information. Therefore, it is common for governments to intervene in the market to address this market failure. Within economics, R&D is a classic public good example, where the justification for public assistance and investment is clear. Policies such as the patent system; R&D tax credits; and financial subsidies for research by government bodies, nongovernmental organizations, and academic institutions are explicitly designed to encourage R&D because, without such incentives, private parties would tend to underinvest in it.

Technical assistance PVPs can be considered investments in R&D in a broad sense because they facilitate knowledge transfer and technological innovation. R&D can promote both innovations that reduce production costs, and innovations that lead to more desirable and saleable products. Firms' motivation to participate in PVPs need not be thought of as altruistic. To the contrary, lower costs and higher revenues (i.e., higher profits) are likely to be the most material drivers for participation in PVPs. Indeed, effective PVPs provide benefits to both the participant and broader society by helping to leverage existing (complementary) incentives, such as cost savings. As a result, the existence of other motives for undertaking a project does not itself prove that a PVP has no impact, though it becomes much more difficult to isolate the "cause" of an operational improvement.

An example of the complexity of motives for participating in PVPs is the decision by Okonite to become an NPEP partner. Company representatives noted in a 2009 interview that a broad corporate dedication to environmental stewardship, combined with interest in market competitiveness and cost savings, were drivers for participation. Participation in the voluntary program helped the company decide to invest in upgrades and hastened its efforts to remove lead from its processes (James Groome, The Okonite Company, Inc., pers. comm., July and November, 2009).

[2] As discussed below, the appropriability of new information determines its nonexcludability. For reviews of the public policy issues related to R&D as a public good, see Griliches (1992) and Tassey (2005).

Although PVPs do not provide direct financial support to firms or their R&D activities, PVPs can lower the cost of innovation by facilitating the distribution of information that reduces production costs and/or leads to more desirable and saleable products. This information can be costly for an individual firm to acquire on its own. In this way, a PVP can be thought of as a subsidy to private R&D. And this government subsidy is justified to the extent that the programs provide a public good, namely, technical assistance that encourages socially desirable, efficient innovations that would not otherwise have occurred in the same time frame.

> **Focus for Attribution of PVP Benefits:** Attribution of technical assistance benefits should be focused on the ways in which PVPs subsidize and encourage R&D that has strong informational public good features.

2.2 Signaling PVPs and Improving Information

Signaling PVPs, as used here, are those that provide participants with an independent acknowledgement of the environmental quality of their products, processes, or supply chains. These programs reward, acknowledge, or certify environmental performance, and thus provide firms and other institutions with an independent, credible signal of environmental quality that they can market to consumers and clients. It can also be valuable for firms to signal environmental performance to financial markets to lower capital costs. Signaling PVPs may provide their members with benefits such as:

- awards and other public recognition;
- logos that signal participation and performance;
- certification assistance and verification; and
- marketing assistance.[3]

[3] Observers often highlight the similarities between PVPs and social marketing—that is, the use of marketing techniques (e.g., advertisement, messaging, and branding) to achieve a broad social health, safety, or welfare goal. Commonly cited examples are pro-seatbelt, antismoking, and antilitter public education campaigns. Signaling PVPs are designed to allow firms to market their achievements in a way that is good for social welfare, suggesting a clear link to social marketing. Several features of PVPs counter this general similarity, however. First, whereas social marketing is broadly defined as mass media messages targeting the general public (e.g., don't smoke), signaling PVPs target specific subgroups (e.g., consumers of a particular product). Second, signaling PVPs are driven by firms' desire for competitive advantage. To some, the fact that PVPs provide participating firms with tools to market their own products for their own competitive advantage might disqualify them as social marketing. Finally, social marketing emphasizes long-term, lasting behavioral change driven by a change in underlying social values. PVPs are not currently designed to achieve such goals; instead, they allow consumers to make purchasing decisions based on values they already hold. Nonetheless, the basic goal of social marketing—"behavioral change for the social good"—is one way to judge the performance of signaling PVPs.

EPA's voluntary programs typically provide one or more of these benefits, particularly awards or public recognition (e.g., though website recognition and the use of logos). For example, the SmartWay Transport program promotes the adoption of advanced clean energy technologies and strategies by encouraging retailers (like IKEA and Target) to choose trucking and/or rail companies that are environmental leaders in their sectors. The SmartWay label is also used to help buyers identify environmentally top-performing passenger vehicles and heavy-duty trucks. SmartWay has more than 2,600 partners, such as trucking companies, rail lines, shippers, logistics companies, and vehicle manufacturers. EPA's Indoor airPLUS program is a signaling program that provides a label for new homes that meet a comprehensive set of indoor air quality specifications. Indoor airPLUS recruits partners in the building industry, including builders and home energy raters, as well as home buyers, to promote healthier homes.

Signaling PVPs address information asymmetries, or market failures in which information is incomplete and consumers are unable to easily observe the environmental features, qualities, and footprints of the products they purchase. Increasingly, consumers and firms seek—and are willing to pay a premium for—products that are environmentally state-of-the-art. Motivations for this focus can be both altruistic (green consumerism) and financial (firms that want to avoid liability arising from their supply chains). Unfortunately, the environmental quality of products, processes, supply and distribution chains, and disposal is often difficult for consumers to observe. In economic terminology, this information breakdown creates a *lemons problem* (Akerlof 1970).[4] The classic example of a lemons problem involves used cars whose underlying quality is difficult for buyers to observe on their own. Lacking precise information, rational consumers will purchase cars based on their understanding of a used car's average quality. This means that sellers of above-average quality cars cannot get the price they seek for their high-quality cars, and will thus withhold them from the market. Buyers can anticipate this, which leads them to further reduce their expectations of average quality. In the end, and left to themselves, some markets may provide only the lowest-quality products because of this information asymmetry between buyers and sellers.

If consumers cannot easily distinguish truly "green" firms and products from other firms and products, then their willingness to pay more for green products will rationally decrease. As in the used car example, this weakens the incentive for producers to provide greener products and processes. The public typically does not have comparative information about the materials

[4] The lemons problem predicts that markets for certain products will not come into existence when consumers are uncertain about quality.

and waste management issues that ORCR voluntary efforts address (with the possible exception of recycled content in goods).

The economic justification for government intervention as a response to information asymmetry is well-established (see Mankiw 2008 485). Market failure associated with information asymmetry is a key motivation for truth-in-advertising laws enforced by the Federal Trade Commission, product certification by the Food and Drug Administration and Consumer Product Safety Commission, and compliance enforcement of technical standards throughout the government.

Signaling can improve the ability of firms to gain a private advantage when they innovate environmentally.But in this case, the profit motive yields socially desirable outcomes that occur through participation in the program. Government assistance is in the social interest because it leads firms to innovate and be able to sell greener products that consumers demand.

> **Focus for Attribution of PVP Benefits:** Attribution of the benefits of signaling features should focus on the ways in which PVPs encourage green product and process innovation in situations where green branding and marketing would benefit from external validation.

2.3 Intended Outcomes (Benefits) of PVPs

It is instructive to characterize the general types of outcomes that technical assistance and signaling PVPs seek to produce. Often these programs seek to encourage firms to switch to the use of processes that require fewer natural resource inputs or result in smaller amounts of pollution released into the environment. For example, PVPs including WasteWise, C2P2, NPEP, Natural Gas Star, and focused efforts such as Carpet America Recovery Effort (CARE) and AgStar all are designed to produce this type of process change outcome. PVPs may also encourage end-product design changes that enhance recyclability or reuse; alterations to the supply chain that either directly or indirectly yield environmental benefits; managerial practices (e.g., lifecycle assessment, environmental accounting rules, or health and safety protocols); and changes in a firm’s overall product mix.

Participation in a PVP by a firm or other organization is not prima facie evidence that the PVP is producing its intended outcomes. In some cases, PVPs could be encouraging innovations that firms would invest in regardless of the PVP. In many situations, firms conduct their own R&D and marketing without government involvement. Recent interviews conducted with NPEP participants illustrated this point: more than half of the interviewees indicated that their companies were already considering the activities undertaken as part of NPEP participation. In

some cases, NPEP commitment either sped up implementation or prioritized the project to continue through difficult economic circumstances. The challenge, then, is to characterize the degree to which PVPs result in outcomes above and beyond those that firms would pursue out of their own self-interest.

2.4 Conventional Regulatory Requirements as an Alternative (or Complement) to PVPs

This analysis argues that PVPs are most likely to be socially beneficial in the presence of an underlying market failure that nonregulatory government intervention can address. However, the same set of circumstances often justifies consideration of conventional regulatory interventions. Although a comprehensive comparison of regulatory interventions versus PVP interventions is beyond the scope of this research, we acknowledge the importance and value of such a study.

A thorough study would address several questions.

- What regulatory options are available to address market failures targeted by PVPs?
- What conditions favor regulatory versus “PVP-type” interventions?
- Are there opportunities for coordinated regulatory and PVP interventions?

Consider the market failure associated with the nature of innovation/R&D as a public good. Regulatory interventions to address this market failure (i.e., interventions with the same basic aim as technical assistance PVPs) might include mandatory pollution-prevention planning and industrywide training programs. Reporting mandates, such as the Toxics Release Inventory (TRI), represent another category of potentially applicable regulatory strategies. These types of interventions are intended (at least partly) to stimulate knowledge creation and innovation within the industrial sector.

As noted, signaling PVPs address the market failure that arises when consumers are unable to easily observe the environmental features, qualities, and footprints of the products they purchase. Regulatory alternatives to signaling PVPs include mandatory labeling requirements (e.g., mandatory energy or fuel efficiency ratings), tighter product design standards, manufacturing rules, and standards applied to suppliers (e.g., forest product certification).

A comprehensive environmental and economic comparison of PVPs to regulatory alternatives would help underscore the unique strengths of PVPs. It is also possible that, under some circumstances, regulation (or PVPs that include regulatory elements) could serve the public interest better than PVPs.

2.5 The Role of Organizational Failures

The framework proposed in this study rests on the assumption that firms are motivated by a desire to maximize profits and are able to identify and act on profit-making opportunities. A competing school of thought holds that firms are not always rational, profit-maximizing "machines." One argument is that pollution prevention can often save firms money but is not undertaken. For example, an analysis of waste reduction at chemical plants concluded that waste reduction occurred in some cases only in response to regulation, though plants found it to be cost-effective once in practice (Sarokin et al. 1985). The explanations for why firms might not act in their own self-interest are various and include information barriers, accounting-based distortions, or inappropriate managerial incentive schemes. At the same time, behavioral economics increasingly emphasizes the importance of psychological realities that conflict with purely rational maximizing behavior (e.g., Thaler and Sunstein 2008).

This paper focuses on two market failures—the public good nature of intellectual property and information asymmetry—to judge the case for government intervention. Arguably, organizational failures should also be considered because, clearly, firms are not perfectly rational, profit-maximizing machines. Indeed, organizational failures are reflected in the value of certain forms of technical assistance, such as improvement in environmental accounting tools and practices. EPA (1995) has documented problematic accounting practices with the potential to bias environmental decisionmaking. Frequent targets for criticism are the allocation of environmental costs to general overhead accounts, the failure to account for future contingent liabilities, and the failure to measure the impact of environmental decisions on corporate image and customer and supplier relationships. This perspective has, in turn, motivated a growing literature on financial and accounting methodologies to improve accounting practices (Boyd 1998a; Moilanen and Martin 1996; Epstein 1996).

It is difficult, however, to argue that organizational failures create a broad rationale for any government assistance or investment. First, firm irrationality is a matter of degree. For their part, corporate environmental managers tend to be more skeptical of pollution prevention's inherent profitability. They point to regulatory barriers that reduce the financial incentive to change production processes or to introduce new products with uncertain regulatory mandates. Evidence for the profitability of pollution-prevention opportunities is decidedly mixed. For instance, EPA (1991) evaluated a broad set of source-reduction options at a large-scale petroleum refinery. Most of the options were found to have negative rates of return and only one had a rate of return higher than the historical rate of return for projects at the refinery. Another study (Boyd 2000, 1998b) found evidence that contradicts the view that firms suffer from organizational weaknesses that make them unable to appreciate the financial benefits of

pollution-prevention investments. Instead, projects foundered because of significant unresolved technical difficulties, marketing challenges, and regulatory barriers.

These complexities make the detection of organizational failures difficult and often subjective. Consequently, organizational-failure arguments can complicate attribution claims and even weaken the case for PVPs. Theoretically, if firms do not predictably make environmental improvements that save them money, the case for mandatory command-and-control type regulation is strengthened. Voluntary efforts, even more than command-type regulation, depend on firm rationality in that they require firms to be rational enough to voluntarily take action.

As a result, the approaches laid out in this paper do not attempt to isolate the potential impacts of organizational failures, though these types of information and decision failures are in some cases reflected in responses to technical assistance efforts.

3. Analytical Approaches To Attributing Outcomes to PVPs

The primary objective of this paper is to identify practical and theoretically defensible approaches that EPA program managers can use to characterize the benefits of PVPs. Experimental design and rigorous statistical analysis represent useful starting points when examining analytical options. The discussion below describes the structure and goals of these rigorous approaches as well as the challenges they pose; these challenges motivate the more practical approaches examined in subsequent sections.

3.1 Experimental Design and the Two-Stage Method

In theory, the strongest approach for measuring a program's benefits involves experimental design (sometimes called randomized control trials, or RCTs). This approach begins with random assignment of potential program participants to a treatment group that participates in a program, and a control group that does not. (In other words, entities have no choice about whether they participate in the program.) This is the approach that scientists often use when testing a new drug: they randomly assign participants to be in the treatment group, which receives the test drug, or the control group, which receives no drug (or a placebo). Because researchers randomly assign potential participants, the treatment and control groups can be considered statistically equivalent. This allows researchers to control for other variables (e.g., age or, in the case of PVPs, firm size or sector) and directly attribute observed differences between the treatment group and control group to the intervention (e.g., the drug being tested or the program being studied).

A fundamental problem with applying an experimental design to assessing the impacts of PVPs is that an experimental design relies on random assignment to a treatment or control group. By the very nature of PVPs as voluntary programs, it is not possible to assign entities to participate in a PVP, nor is it generally considered acceptable for a government program to exclude eligible entities from a PVP. Therefore, experimental designs are not well-suited for evaluation of PVPs.

An alternative approach to assessing the outcomes of PVPs is to approximate an RCT by comparing existing program participants to those that do not participate (i.e., a comparison group).[5] Researchers would ideally choose a comparison group of nonparticipants that is as similar as possible to the participant group in known characteristics (e.g., company type, markets served, and firm size). It is also possible to use statistical approaches to control for characteristics in comparison groups that might affect outcomes. We use the term *two-stage method* to describe approaches that correct for nonrandomly selected samples (Heckman, 1979)..

The essential steps for conducting a two-stage method to assess a PVP include the following.

- First, program managers would segregate potential program participants into treatment and comparison groups; that is, one group that participates in the program and another that does not. The groups would not be randomly assigned because program participants would volunteer to participate in the PVP. However, program managers could try to ensure that the characteristics of the comparison group match those of program participants.
- Second, program managers would identify key outcomes of interest and collect the requisite data. For instance, this might involve tracking waste generation among firms in the treatment and comparison (nonparticipant) groups.
- Finally, program managers or evaluators would use quantitative methods to test for statistically significant differences between the treatment and comparison groups. The analysis would control for other influences on the outcome of interest to characterize the true net effect of the program intervention.

[5] We do not consider this group a true "control group" because entities are not randomly assigned between the treatment and control groups.

3.2 Challenges Associated with the Two-Stage Method

The specific features of voluntary programs make it difficult, and often impractical, to use a full-scale statistical analysis to assess the outcomes of PVPs.[6] The discussion below highlights several important obstacles: (a) sample bias, (b) breadth of confounding factors, (c) non-uniform interventions and outcomes, (d) ambiguous intervention events, and (e) spillover effects.

Sample Bias

Using full-scale statistical approaches to analyze comparison groups is subject to *sample bias*; in other words, PVP partners (i.e., program participants) and nonpartners may systematically differ in some way. This is a distinct possibility, given that participation in a PVP is voluntary. For example, firms most likely to benefit from a PVP are more likely to be participants. If so, a study could overstate program benefits because PVP benefits would not be as great if the PVP were also applied to nonparticipants.[7] Also, for a variety of reasons, large firms, greener firms, or firms with good management may be more likely to participate. If these factors are independently related to outcomes (as they probably would be), the partner sample will be biased. Accordingly, the analysis would have to control for these biases.

However, researchers may not know all of the relevant characteristics that could distinguish between the treatment and comparison groups. Even if researchers could identify all relevant characteristics, some key factors (e.g., a firm's motivation to innovate, or the presence of a dynamic leader to implement changes) may be impossible to measure. The effort to develop a complete set of control variables creates extensive data demands, particularly for PVPs that address more than one sector.

Breadth of Confounding Factors

Many factors may affect the performance of PVP participants, aside from their participation in a PVP. Product and process innovations are driven by sector- and firm-specific competitive pressure, regulations, liability concerns, financing options, and changing consumer tastes. Ideally, PVP participation enables firms to respond more effectively to these pressures and to increase innovation, but it is not likely that all innovation at a firm would result from PVP participation.

---

[6] For examples in the literature of attempts to examine ORCR programs using statistical approaches, see Industrial Economics, Incorporated (2008b).

[7] For an empirical study that highlights this concern, see Alberini (2007).

To correctly measure PVP benefits, the analysis must control for these external factors. A further complication is that firms are generally in a constant state of innovation. This means that baseline conditions are "moving," or intertemporal. Accordingly, benefits attributable to PVPs probably are due to the capacity for PVPs to accelerate innovation. Detecting innovation is easier than detecting earlier adoption of innovation.

Because many, if not most, of the other drivers of innovation are changing across time (e.g., regulatory requirements, consumer tastes, and competition), the two-stage method also requires the use of carefully designed time series and comparison groups. For example, it may be possible to measure conditions in a treatment and comparison group at a baseline time period, conduct a follow-up measurement of both groups some time later, and then compare the rate of change or innovation in each group (i.e., a differences-in-differences approach). Although this approach is possible,[8] it is somewhat complex, and any differences in the effects of external factors on the treatment and comparison groups can create significant uncertainties in the analysis.

Non-Uniform Interventions and Outcomes

To compare treatment and comparison groups, PVP interventions must be standardized. For example, technical assistance should be grouped into like forms of assistance and classified according to the following characteristics:

- subject matter (e.g., process reengineering, product reformulation, and recycling);
- form of assistance (e.g., in-person site or product evaluation, phone queries, written guidance documents, or conferences and workshops); and
- intensity of assistance (e.g., hours or duration of assistance).

Likewise, signaling interventions could be standardized in terms of the form of recognition, certification, and auditing conducted.

[8] For example, EPA's Office of Enforcement and Compliance Assurance is currently implementing a pilot project that uses experimental design and the two-stage method to measure the impacts of EPA Region 1's compliance assistance outreach for the auto body sector. However, the assistance effort is not a PVP because it focuses primarily on encouraging compliance with existing or forthcoming regulations, and participation is not voluntary (assistance is offered to all entities in the treatment group).

Second, the beneficial outcomes of the interventions must be standardized. In the case of both technical assistance and signaling features, these outcomes will differ across partners in the following ways:

- type of outcome (e.g., process innovation, product innovation, input reductions, change in product mix, changes to supply chain, and recycling);
- scale of outcome (e.g., pounds of pollutant reduced, tons of input recycled, or number of processes affected); and
- economic, health, and other social benefits of the outcome (e.g., the monetary value of innovations or public health improvements).

Numerous statistical issues remain, but even the basic standardization of interventions and outcomes may be difficult when assessing PVPs. This is particularly true of outcome measures. Because they are voluntary and idiosyncratic to the needs of particular partners, there is no simple way to standardize, integrate, or compare different kinds of outcomes. Innovations tend to be specific to firms or sectors and this can prevent the development of data sets with uniform outcome measures and observations numerous enough to achieve statistical power. Also, because program partners typically self-report outcomes, it is a challenge for the agency to require the use of consistent outcome measures.

In some cases, evaluators may be able to take advantage of standardized data sets in assessing the beneficial outcomes of PVPs. However, options are generally restricted to a limited set of existing databases (e.g., the TRI and Biennial Report data) that are not universally relevant or useful. The limitations of these data sets have curtailed the application of statistical approaches. In addition, these approaches may not be well calibrated to address the impacts of information diffusion.[9]

Ambiguous Intervention Events

To detect program effects it is important to clearly identify the time frame of the program's intervention. In one ORCR voluntary effort, the C2P2 program, a single industry adopted new policies all at once, and the changes in practice that followed are easily identified in data collected annually by the industry. In contrast, however, most PVP interventions are staggered across time, with rolling program admission and negotiated projects of differing

[9] For a recent, detailed assessment of the limitations of using TRI data to assess ORCR's NPEP program, see Industrial Economics, Incorporated 2008a.

lengths. For instance, with WasteWise and NPEP, entry to the program is rolling, and innovations or projects do not have consistent time frames or implementation deadlines. As a result, it may be difficult to establish a baseline and to identify clear temporal thresholds for before-and-after analyses.

Spillover Effects

Finally, the diffusion of innovations poses a crucial challenge to the use of experimental design or statistical approaches to measure PVP benefits. As noted, PVPs seek to create R&D and other information that is appropriable. By its nature, appropriable R&D tends to diffuse quickly and easily, and it may be shared outside the treatment group. If PVPs facilitate new knowledge and practices that quickly diffuse, the comparison group may be observed implementing new innovations. If so, the attribution analysis will show less difference between the treatment and comparison groups, leading to a smaller estimate of the program's effects. Of course, in such a case the opposite would actually be true—that is, if diffusion occurs, the program is creating particularly large benefits.[10]

In short, given that a key goal of many PVPs is the exchange of information regarding new technologies and practices, social benefits may arise because of behavioral change on the part of companies that do not participate in the programs at all. This *spillover* effect means that it may not be possible or desirable to prevent information from reaching the comparison or control group.[11]

4. Practical Approaches for Attributing Benefits

Having described the obstacles to experimental design and the two-stage method, this paper now focuses on alternative approaches for benefit attribution. The objective is to recommend practical methods that are theoretically defensible but which place fewer, more realistic, demands on data collection and analysis. Specifically, the discussion presents a three-level approach to attribution analysis in the context of ORCR initiatives.

- **Level 1—Initial Threshold Assessment:** Threshold assessment ensures and documents that the PVP design is appropriate for achieving targeted benefits. The assessment

[10] See Lyon and Maxwell (2008), who argue that PVPs are a knowledge-diffusion mechanism and that successful diffusion seriously complicates the measurement and detection (attribution) of program effects.
[11] See Maxwell and Lyon (2007), for a detailed discussion of the spillover benefits of voluntary programs.

explicitly considers the market failures discussed earlier and helps verify that the PVP under consideration has a defensible economic rationale.

- **Level 2—Intervention–Outcome Assessment:** Assessment of interventions and outcomes helps ORCR verify that the resources and activities it invests in a PVP are logically aligned with the desired outcomes.
- **Level 3—Quasi-Experimental Designs:** In some instances, ORCR may pursue quantitative analyses that effectively attribute benefits of a PVP, while avoiding the strict demands of experimental design or the two-stage method.

Throughout, the discussion recognizes the complexity of PVP interventions and data limitations associated with ORCR programs. The section concludes with a discussion of several special outcome categories and measurement considerations unique to PVPs.

4.1 Level 1, Threshold Evidence To Support Attribution

The case for benefit attribution will generally be strongest in situations where the underlying economic rationale for government-supported PVPs is clear. This section describes the types of information needed to link PVP features with market failures where PVPs are most likely to be socially beneficial and in need of government intervention. To the extent that a program directly recognizes and addresses key market failures, it can be thought of as satisfying a fundamental *threshold test* for potential success.

The discussion below first examines threshold assessment for technical assistance programs designed to address the public good nature of innovation. Next, the discussion considers threshold assessment for signaling programs designed to address information asymmetry.

Threshold Evidence To Support Technical Assistance Attribution

The case for attribution is strongest when technical assistance helps participants create and apply green R&D that is easily appropriable by others. Appropriable information can be thought of as discoveries that quickly and easily become common knowledge; it is easily copied, used, or duplicated by other institutions or competitors. In contrast, nonappropriable information is easier to keep private and deploy to one’s private advantage.

Specific discoveries, R&D, and information vary in the degree to which their value is appropriable by the discoverer. Appropriability depends on many factors, including the nature of the new information; the competitive application of the information by markets; the difficulty of applying new knowledge to practical problems; and laws designed to reward discovery.

At one extreme, basic scientific discoveries (e.g., that bacteria cause disease) are often highly appropriable. Once discovered, they easily become common knowledge, due in part to the nature of the information itself (explaining or defending the new knowledge provides the key to its duplication) and also cultural norms that value the sharing of basic scientific knowledge.

At the other extreme are proprietary or trade secrets that can be, and are, jealously guarded by discoverers. For example, many firms take great care to prevent knowledge of production processes, product formulas, and even management practices that give them a competitive advantage. “Noncompete” contracts that prevent employees from being hired by competitors are evidence of the incentive and ability of firms to keep some information private, or nonappropriable. Patentable discoveries are also nonappropriable because (by design) the patent system protects a discoverer’s exclusive ownership of new knowledge for a period of time.

The public policy case for interventions that subsidize or assist discovery—such as PVP technical assistance—is strong in the former case (appropriable information) and weak in the latter (nonappropriable information). In the former case, the discoverer is less likely to privately benefit from any new knowledge their R&D investments produce. This means that the private financial incentive to invest is less than it should be to induce socially justified investments in new knowledge.

Accordingly, attribution of technical assistance benefits to PVPs is justifiable (at least as a threshold test) when technical assistance encourages the discovery and application of R&D and other information that is appropriable. Attribution is more difficult to justify when it supports proprietary information. There is little or no market failure associated with proprietary discoveries because the private incentive to discover them is already strong. PVPs will benefit participants when they support proprietary R&D, but the rationale for government intervention is weak.

Focus for Attribution of PVP Benefits: Attribution of technical assistance benefits to PVPs should focus on assistance that clearly targets *appropriable* versus *proprietary* environmental information and practice.

Therefore, when evaluating a technical assistance program, a practical challenge is to distinguish between technical assistance that fosters appropriable versus proprietary innovation and discovery. For example, technical assistance that was used by a partner to develop a patented product or process, or a valuable trade secret, would not pass the threshold test and should be excluded from attribution on the grounds that the PVP in this case is subsidizing discovery that is in the partner's private interest (i.e., no market failure is being addressed). On the other hand, when technical assistance supports innovations that are not patentable, are easily copied by other firms and institutions, and can quickly and easily diffuse throughout a market or industry, the economic justification for attributing program benefits to the PVP is strong. Box 1 summarizes these principles.

BOX 1.

Attribution threshold for appropriability met:

- Nonpatentable innovation
- Innovation with broad applicability to other firms
- Innovation other firms can cheaply duplicate
- Innovation other firms can quickly duplicate
- Innovation applicable to small firms with numerous competitors

Attribution threshold not met:

- Patentable innovation
- Innovation that is only applicable to specific firms
- Innovation that is expensive to duplicate
- Innovation that can be adopted by others only after significant lag

Strong evidence in favor of PVP benefits is technical assistance that leads to the diffusion of innovation beyond a specific partner to other PVP partners or nonparticipants. Evidence of such knowledge spillovers points to an underlying market failure addressed by the PVP—namely, the "free" appropriation of new intellectual property by competitors and other users. In short, innovations and green practices facilitated by technical assistance that diffuse to other firms—that is, spill over—are particularly attributable to PVPs.

The attribution factors identified above are not "bright line" threshold tests. It is not necessary for a PVP to satisfy all of the threshold criteria to make the case that the program has attributable impacts. It is also possible that a PVP can demonstrate attributable benefits even if one or more proprietary features exist within a program. For example, in general, a PVP intervention that enables a partner to patent a new process or product argues against attribution. In specific cases, however, patents may be only weak

protection against appropriation, as in a case where competitors can cheaply and quickly "patent around" the partner's patent.

Although a "perfect score" is not critical, it is important to consider these attribution factors when designing and prioritizing program elements and interventions. The social and economic argument for both the desirability of intervention and the attribution of public benefits to the agency's actions is strongest when the largest number of threshold tests is passed.

Program managers should recognize the potential for furnishing participants with an undesirable degree of competitive advantage. Innovations that quickly diffuse and are nonproprietary tend to be less competitively advantageous—precisely because they are nonproprietary. Competitive advantage by itself does not disqualify the attribution of benefits, however; partners must expect some competitive advantage to justify their participation in the PVP. What is crucial is the degree of competitive advantage. PVPs that provide extreme competitive advantages are not socially desirable because they duplicate what is already in a firm's private interest. Instead, a balance must be struck: enough competitive advantage to induce participation, but not so much that the diffusion of R&D public goods is inhibited.[12] In the context of EPA, publication of achievements (e.g., project summaries) may be a key factor in ensuring that information is appropriable.

BOX 2. NPEP AND INFORMATION SHARING

ORCR's NPEP program provides several examples of the complexity of the technical processes and products that technical assistance can address. Among nine partners interviewed in 2009, one company (a manufacturer) specifically noted that innovations comprised trade secrets that were partially, and confidentially, shared with clients, but not with competitors; this company nevertheless feels that its NPEP participation has furthered innovation within the company, and therefore within the sector. In contrast, another partner (an airport) has publicized its innovations and has actively encouraged participation in similar projects by other departments in the local government. In addition, other manufacturing partners have noted that interaction with firms in different sectors (but with similar waste-related challenges) have helped further their participation in the program.

Threshold Evidence To Support Signaling Attribution

The case for attribution of signaling benefits to PVPs is strongest when signaling gives stakeholders knowledge of green production and product features that are difficult to observe. As reviewed, signaling seeks to overcome market failures associated with asymmetric information.

[12] This balance is a key feature of public policies geared toward innovation. For example, patents provide intellectual property rights, but only for a limited time. This feature of patents takes into account both the need to induce innovation and the social efficiency of its diffusion across the economy.

In many cases, it is difficult for consumers to make the technical assessments necessary to judge whether a product is truly green. A consequence of these information asymmetries is that consumers cannot condition their purchases on products' true environmental features. This weakens the incentive for high-quality green firms to invest in green products and processes because it is difficult for them to pass on those extra costs via premium prices.

Also, information asymmetries create market failures beyond end-product markets. For example, firms wishing to purchase green inputs from suppliers and other business partners may find it difficult to judge their environmental practices. Similarly, financial markets often look for signals of firms' green practices to judge the risks associated with investment. Individual consumers, employees, and communities may also look for independent verification of a firm's green credentials. In other words, many stakeholders may wish to base their purchase, investment, employment, and community development decisions on information that is hard for them to independently assess and verify.

Of course, when information is clear, obvious, and credible there is no information asymmetry and no presumption of market failure. Some products and processes may have obvious environmental consequences. For example, electric vehicles and garden tools obviously emit less local air pollution than their gas-powered alternatives. Consumers can also rely on a range of green signals provided by government certification programs (e.g., Energy Star, recycled content, and fuel efficiency ratings). In some cases, the private sector establishes its own certification and standards programs, primarily to deal with supply chain information issues. ISO 14000 certification and testing by Underwriters Laboratories are examples of such private sector–driven signaling programs. Markets also rely—to some extent—on government enforcement of truth-in-advertising laws.

Attribution of signaling benefits to EPA PVPs is most justifiable when recognition, awards, and auditing are directed at environmental quality information that is: (a) difficult to observe and (b) not already addressed by some other third-party certification or auditing scheme. For example, ORCR's CARE (which promotes carpet recycling) highlights carpets made from secondary materials; this market has not traditionally enjoyed the publicity of other recycled product markets. Likewise, EPA's Indoor airPLUS program is the only labeling specification in the marketplace that comprehensively deals with indoor air issues. Attribution is more difficult to justify when one or both of these conditions is not met. When both conditions are not met, PVPs may still be beneficial (because an additional signal is being provided), but the rationale for government intervention is weaker.

Focus for Attribution of PVP Benefits: Attribution of signaling benefits to PVPs should focus on awards, recognition, and auditing that target environmental features, costs, and footprints that are both difficult to observe and not already addressed by an existing third-party certification or auditing scheme.

When analyzing PVPs, the practical challenge is to identify awards and recognition that provide quality signals to stakeholders that would otherwise be absent from the market. Resource intensity, pollution releases, green management practices, and many product features are almost always difficult for key stakeholders (e.g., customers) to observe easily. The complexity of products, processes, and supply chains; the geographic distance over which production occurs; and the often hidden nature of environmental damages mean that environmental performance will almost always be difficult to observe readily. Awards, recognition, and other signals of environmental performance provided by PVPs are likely to be socially beneficial when they address stakeholder information constraints. Attribution of benefits is most justifiable when the PVP does not duplicate signals provided by other third-party performance assessment schemes.

4.2 Level 2, Intervention–Outcome Assessment

A second level of analysis for attributing benefits to PVPs focuses on developing a thorough inventory of interventions and outcomes. Interventions essentially represent the activities and inputs that EPA program managers invest in the PVPs under consideration (e.g., hotline assistance, guidance documents, and so on)..Outcomes are the results that the program seeks to achieve. They can take many forms, ranging from near-term behavioral changes (e.g., reduction in the quantity of a toxic manufacturing input) to ultimate, long-term environmental outcomes (e.g., reduced exposures and improved public health).

Overall, the objective of the intervention–outcome exercise is simply to provide a framework in which program managers can critically assess whether inputs align with the outcomes and environmental goals that EPA is seeking to achieve. In early planning stages, this type of accounting exercise can highlight key program design issues and can help allocate resources across a suite of related initiatives. In later stages of retrospective assessment, it can aid in the reallocation of resources between programs and can ensure that data needed for quantitative analyses are readily generated.

Program managers have many alternatives for organizing the information in an intervention–outcome assessment. This paper reviews two possible approaches: (a) intervention–outcome matrices and (b) logic models.

Intervention–Outcome Matrices

PVPs could collect, organize, and report their interventions and outcomes in a standardized matrix format that distinguishes between type of intervention and type of outcome. This "binning" of interventions and outcomes will organize and clarify evaluations by making transparent the full range of program interventions and effects.

The intervention side of the matrix reflects data collected by the program itself and could include information on the subject of the PVP (i.e., the process or activity EPA is seeking to influence), the form that intervention takes, and the intensity of the intervention, as shown below.

Subject of intervention	Form of intervention (for each subject)	Intensity of intervention (for each subject–form combination)
• Process reengineering • Supply chain alternations • Recycling • Disposal practices • Product reengineering	• On-site evaluations and support • Hotline assistance • Guidance documents • Online resources • Conferences and workshops • Recognition and awards	• EPA hours devoted to intervention • Type of EPA personnel involved • Duration of partnerships and assistance • Other EPA investments

A simple illustration might involve a program to encourage the use of a less-toxic production input (process reengineering). The intervention may take the form of an on-site evaluation in which EPA representatives identify chemical substitution opportunities. The intensity of this intervention could be assessed in terms of personnel or contractor hours devoted to the evaluation visits.

Classifying interventions in this manner serves several purposes. First, it highlights the range of interventions in which PVPs are currently or potentially applicable. Second, it allows evaluators to assess the relative frequency with which different interventions are used. Third, it makes it easier to associate program costs with particular kinds of interventions (i.e., what types of interventions impose the greatest resource costs on the agency?). Fourth, when combined with outcome measures, the matrix allows a clearer analysis of cause and effect (intervention and outcome).

Now consider the outcome side of such a matrix, reflecting the specific achievements of the PVP and its participants. Here, program managers can organize information according to the type of outcome, the scale of the outcome, and the ultimate benefit (environmental and/or economic) sought, as shown below.

Type of outcome	Scale of outcome (for each type)	Ultimate benefit (for each type and scale)
• Process type affected • Chemical input reduced or changed • Resource input reduced or changed • Waste stream recycled • Waste stream disposal method or change • Product or feature changed	• Volume of reduction (net of increases due to substitution) • Number of processes or products affected • Number of facilities affected • Diffusion of changes to other PVP partners (using same metrics as above) • Diffusion of changes to nonpartners (using same metrics as above) • Internal diffusion of product or process	• Public health benefits expected (Y/N) and type (e.g., air or water) • Ecological benefits expected (Y/N) and type (e.g., species or land cover effects) • Production cost savings and form (e.g., input costs, liability, or disposal) • Consumer benefits (e.g., more desirable product features or greater product variety)

Continuing the illustration introduced above, the type of outcome is a chemical input reduction. The scale of the change could be measured in terms of the volume reduction at key facilities that received on-site evaluations; the overall volume reduction across all facilities; or in terms of the number of facilities that participated. Ultimately, benefits might be characterized in terms of both production cost savings for the participants and reduced exposures to the subject chemical at landfills or other disposal facilities.

PVPs already collect some of the information needed for the development of intervention–outcome matrices. For example, NPEP's enrollment form (questions 1a, 1b, 2a, and 2b) provides some information on the type and scale of outcome associated with enrollment. The key to attribution, however, is to distinguish between outcome types and then relate them to the specific interventions that facilitated them.

Again, classification will help communicate the broad range of PVP outcomes and will also underscore the difficulties associated with statistically controlled experimentation. But even a simple classification scheme such as this can help direct attention to the role that specific interventions play in the generation of particular benefits and the scale of these benefits—the essence of performance evaluation.

Another dimension of the matrix could highlight key external factors—that is, other factors that could be particularly important drivers of behavioral change but that occur independently of the program intervention. External factors may include new regulatory requirements, significant changes in economic conditions, and firm size and profitability.

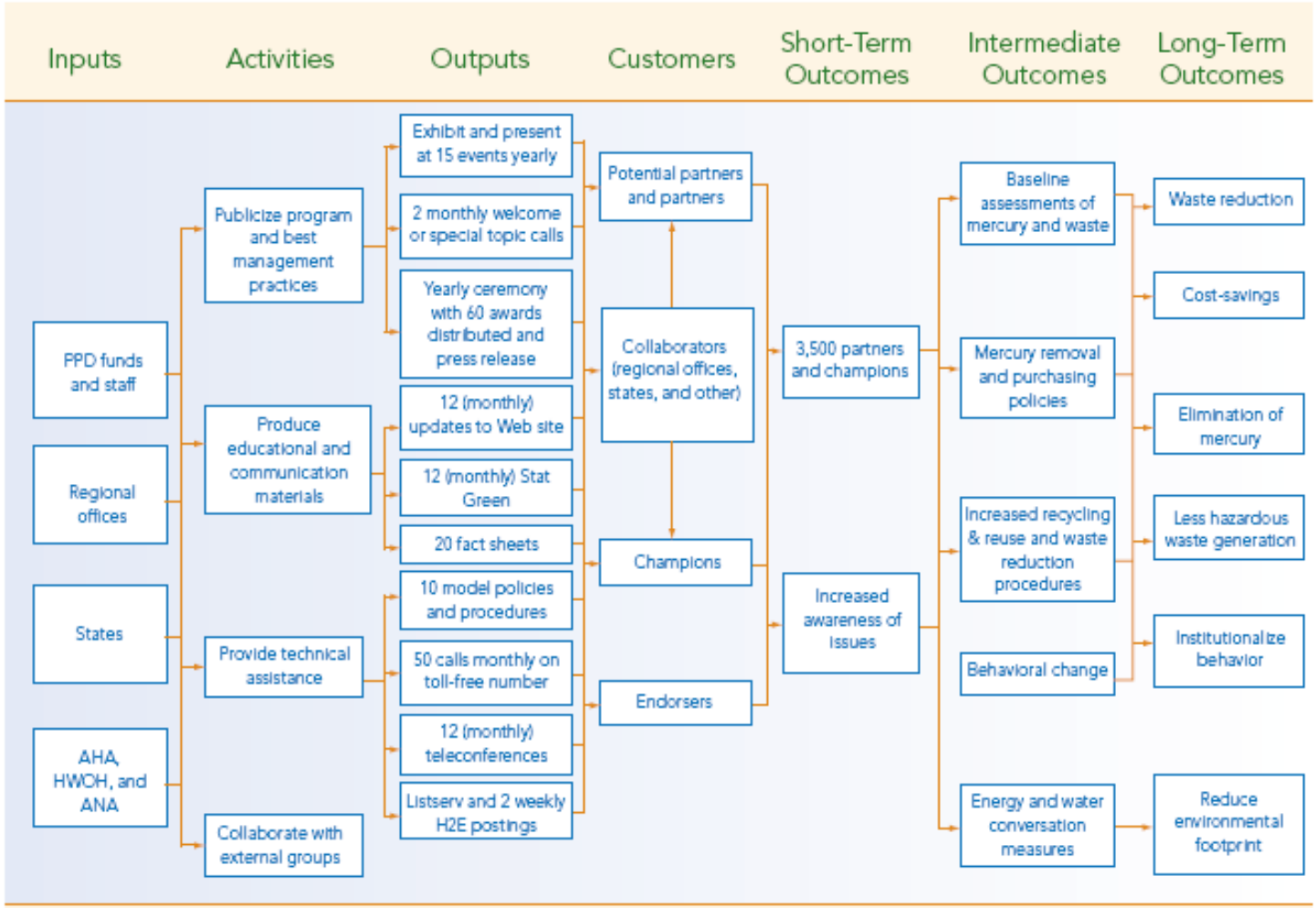
Logic Models

A closely related alternative to the matrix approach is the documentation of program activities using a logic model (described in detail in EPA [2009]). Commonly used in EPA program evaluations, logic models represent a different way of arraying much of the same intervention and outcome information discussed above (albeit with somewhat different terminology). Organized in flow chart fashion, logic models typically include several elements:

- resources and inputs for the program (e.g., staff and funding);
- activities that the program performs;
- outputs, defined as intermediate products or services that the program provides;
- target decisionmakers;
- short-term, intermediate, and long-term outcomes; and
- external influences.

Logic models may provide a useful organizing principle for arraying program information. In assessing the success of a PVP at addressing market failures, however, the key exercise, using either the intervention–outcome matrix or the logic model, is documenting a clear link between program outcomes and the elements of program design that focus on market failures. Figure 1 presents a logic model for Hospitals for a Healthy Environment (H2E), an EPA partnership program launched in 1998 with the goal of advancing waste-reduction and pollution-prevention efforts in the nation's hospitals. Specifically, H2E directed its efforts toward (a) virtually eliminating mercury-containing waste, (b) reducing the overall volume of regulated and nonregulated waste, and (c) identifying hazardous substances for pollution-prevention and waste-reduction opportunities by providing a variety of tools and resources to its partners.

Logic Model for Hospitals for a Healthy Environment (H2E) Program (August 23, 2005)



Abbreviations:

H2E: Hospitals for a Healthy Environment
PPD: Pollution Prevention Division
AHA: American Hospital Association
HWOH: Health Care Without Harm
ANA: American Nurses Association

4.3 Level 3, Quasi-Experimental Approaches to Benefit attribution

When feasible, program managers may choose to progress to a third level of benefit attribution in which they develop quantitative analyses for key PVP outcomes. Quasi-experimental designs do not involve random assignment and may not involve the two-stage method of comparison groups. However, quasi-experimental designs often involve comparison groups and/or measurements of the study group over time (i.e., time series designs). Although these approaches do not definitively prove a program's impact, they provide a rigorous assessment of a program's influence by ruling out other plausible explanations through rigorous measurement and control.

Quasi-experimental approaches can take a vast array of forms. Overall, the analysis generally involves tracking one or more quantitative outcome indicators and logically associating changes in the indicator(s) with the program intervention. A simple example might involve an outreach program to encourage small businesses to recycle waste oil at centralized recovery facilities. The analysis could simply compare recycled quantities at the local recovery facility before and after institution of the outreach program. If adequate data were available over time, the analysis could apply simple statistical techniques to verify that the change in recycled quantities is not due to normal economic variations.

This type of trend analysis involves a variety of compromises in statistical rigor. The brand of quasi-experimental design recommended here might typically involve the following simplifications relative to the two-stage method.

- The analysis may identify a suboptimal comparison group, comparing participants and nonparticipants without regard to aligning the secondary characteristics of the two groups.
- The analysis may simply consider longitudinal data (i.e., before-and-after data) without establishing a comparison group.
- The analysis may consider changes in the outcome variable without providing a quantitative assessment of external factors that may have contributed to the change.
- The analysis may track changes in the outcome of interest without performing formal statistical tests to demonstrate statistically significant changes in the "before" and "after" observations.

Overall, the key message for program managers is to formulate quantitative analyses that adhere as closely as possible to "ideal" statistical analysis while making practical compromises where necessary. An excellent example of this principle is provided by a trend analysis developed for C2P2 (Lange, 2008), which compared waste-reduction metrics across different samples of firms and over time. The study had several analytical advantages. First, the researchers had access to particularly good data on waste changes over time. Second, the interventions and outcomes of the program are reasonably uniform across a large number of facilities. Third, the study was able to normalize waste levels as a function of output and other variables. Fourth, the C2P2 program featured a clear point in time when purchasing guidelines took effect across the industry. The study found a significant increase in waste reuse when the program was introduced. However, this finding is only suggestive of program benefits. In particular, it does not rise to the level of the two-stage method, because the analysis was unable to include many of the controls necessary to isolate the C2P2 program's effect.

Notably, Lange (2008) evaluates the confounding role diffusion can play in attribution efforts. The study found no statistical difference in reuse rates between partners and nonpartners. The C2P2 evaluation explored one particular "diffusion variable"—the geographic proximity of partners—and found that positive nonpartner outcomes were related to the nonpartners' proximity to partners. This is not an iron-clad method for detecting spillover effects, however. First, physical proximity is one of many possible diffusion pathways; other pathways have little to do with geographic proximity (information flows via trade meetings, publications, and online resources are not geographically dependent). Second, geographic location is a proxy for many other intervening variables that have nothing to do with C2P2 interventions but that might explain adoption rates, such as state regulations, community pressures, employee socioeconomic characteristics, media attention, and so on. Overall, the C2P2 case study demonstrates that an imperfect quantitative analysis can still supply compelling evidence of program benefits, guiding ORCR's decisionmaking.

Another example of a quasi-experimental approach is the Environmental Results Program (ERP), an innovative approach to improving facilities' management practices within small business sectors (e.g., auto body repair, dry cleaners, and printers). ERP combines compliance assistance, state or EPA inspections, and statistically based performance measurement to drive improved compliance and reduced environmental impacts of the target sector, while fostering efficient and strategic use of government resources. In many cases, business participation in the compliance assistance aspects of ERP is voluntary (though compliance with applicable regulations is not). ERP relies on longitudinal data by measuring performance at representative

samples of regulated entities before and after those businesses are offered technical assistance. Changes in performance may be associated with ERP. In addition to considering measured changes in performance, the agencies implementing ERP also qualitatively consider what other factors may be affecting entities' performance (e.g., changes in economic conditions or newly promulgated regulations).

4.4 Specialized Outcome Categories and Related Measurement Considerations

This final section highlights a variety of analytical issues that program managers may encounter when applying the methodologies discussed above, particularly the quasi-experimental approaches associated with Level 3. The objective is to highlight outcome indicators that are relevant to PVPs and provide guidance on effective measurement metrics. The topics addressed include the following:

- partners' demand for program interventions;
- partners' leveraging of recognition;
- devotion of partner resources;
- spillover metrics; and
- firm size as an indicator of PVP benefits.

Partners' Demand for Program Interventions

Program partners may demand, apply, and invest in PVP participation. Measurement of this behavior represents practical evidence of program benefits, particularly when combined with the other types of evidence articulated in this paper. Qualitatively, ORCR collects and presents some data on partners' revealed preferences when it reports success stories and associates these successes with program actions. But it is possible to expand this set of data to provide more detailed information on demand for PVPs and the services and features they provide.

EPA offices should develop standardized demand measures that begin with the following:

- participation rates (number of partners);
- duration and consistency of participation (whether partners stay in the programs);
- PVP resource usage rates (hotline calls, conference participation, and guidance downloads and requests); and

- evidence of repeat demand.

Many programs (e.g., NPEP, WasteWise, SmartWay, and Indoor airPLUS) collect some of this information already.For example, see the analysis of hotline and web content usage in EPA's 2010 "Evaluation of the WasteWise Program.[13] But data collection could potentially be further standardized and analyzed across programs to be consistent with the matrix approach described above and to provide consistent, officewide information about program organization, size, and achievements.

Leveraging of Recognition

In the case of awards and recognition (and their role in environmental signaling), it may be feasible for EPA to track the ways in which partners advertise or otherwise communicate awards, recognition, or certification to stakeholders. For example, programs could track the use of awards, logos, and other forms of recognition in the following contexts:

- product labeling (i.e., whether partners feature the recognition on their products);
- marketing materials and other advertising;
- annual reports and other financial reports;
- corporate websites; and
- corporate presentations.

If partners perceive value in recognition and awards, there should be measurable evidence of their use in the partners' own interactions with stakeholders.

Devotion of Partner Resources

As a substitute for direct evidence of innovation, EPA program offices could track partner resources devoted to program participation. Relevant indicators might include financial investments associated with the partnership or the number of employees and hours devoted to participation in the PVP. Where data exist to document them, analyses of partner investments are a particularly credible signal that PVP benefits are occurring.

[13] Available at http://www.epa.gov/evaluate/pdf/wastewise_eval_rpt.pdf.In addition, EPA's Performance Track Program routinely collected and maintained detailed data on memberships, renewals, and withdrawals of its partners. Although this program has been completed, the data collected may provide insights into data collection protocols for other programs.

Spillover Metrics

Because diffusion and spillovers are closely tied to PVP benefits—and because of the difficulty of attributing benefits to PVPs—we recommend that programs develop spillover metrics to detect knowledge diffusion.

Spillovers are likely to occur along particular knowledge and information-sharing pathways that may depend on the type of innovation involved. For example, diffusion can occur within a firm via the partner's management procedures, engineering standards, and employees. Knowledge is also likely to be disseminated via supply chain relationships. External pathways include trade publications, academic journals, migration of employees between firms, and conferences.

Ideally, PVPs should identify specific diffusion pathways and should look for evidence of PVP information along these pathways in terms of:

- internal diffusion of knowledge and practices by partners;
- citation of PVP publications and guidance in various literatures;
- discussion of PVP-related innovations in various literatures;
- PVP-related procurement, disposal, recycling standards, and goals communicated to the supply chain; or
- reference to PVP interventions in trade shows and other conferences.

The WasteWise partnership program is particularly ripe for such data collection because it explicitly encourages endorsers to diffuse knowledge via advertising, recognition programs, publications, and the endorsers' own technical assistance to other institutions. Evidence of diffusion is a particularly strong signal of program benefits because diffused innovations are most likely to suffer from underinvestment because of their public good–like properties.

Firm Size as an Indicator of PVP Benefits

Evaluators sometimes assume that the involvement of smaller firms is an indicator of program success, with the underlying logic being that smaller firms are more dependent on intervention to encourage innovation. However, considered closely, firm size does not have clear implications for benefit attribution.

It is true that small firms will have greater difficulty conducting their own R&D. They also tend to operate in markets composed of large numbers of competitors. Large firms can internalize more benefits of R&D because they can apply discoveries to their own (presumably

wider) range of processes and products. Alone, these factors suggest that small firms will be particularly susceptible to R&D market failures because it could be more difficult for small firms to create, protect, and enjoy the benefits of new intellectual property.

However, small firms often aggressively innovate, with no need for intervention. Moreover, large firms may find it easier to diffuse the fruits of technical assistance within their organizations (e.g., to multiple manufacturing facilities). This implies that a given technical assistance intervention may yield larger social benefits when applied to larger firms.

Likewise, it is not clear that signaling benefits are more likely when provided to small firms. Small firms may have more difficulty generating recognition, certification, and awards because of their smaller environmental health and safety staffs. On the other hand, large firms may reach larger numbers of consumers with correspondingly large benefits to performance signaling.

5. Summary

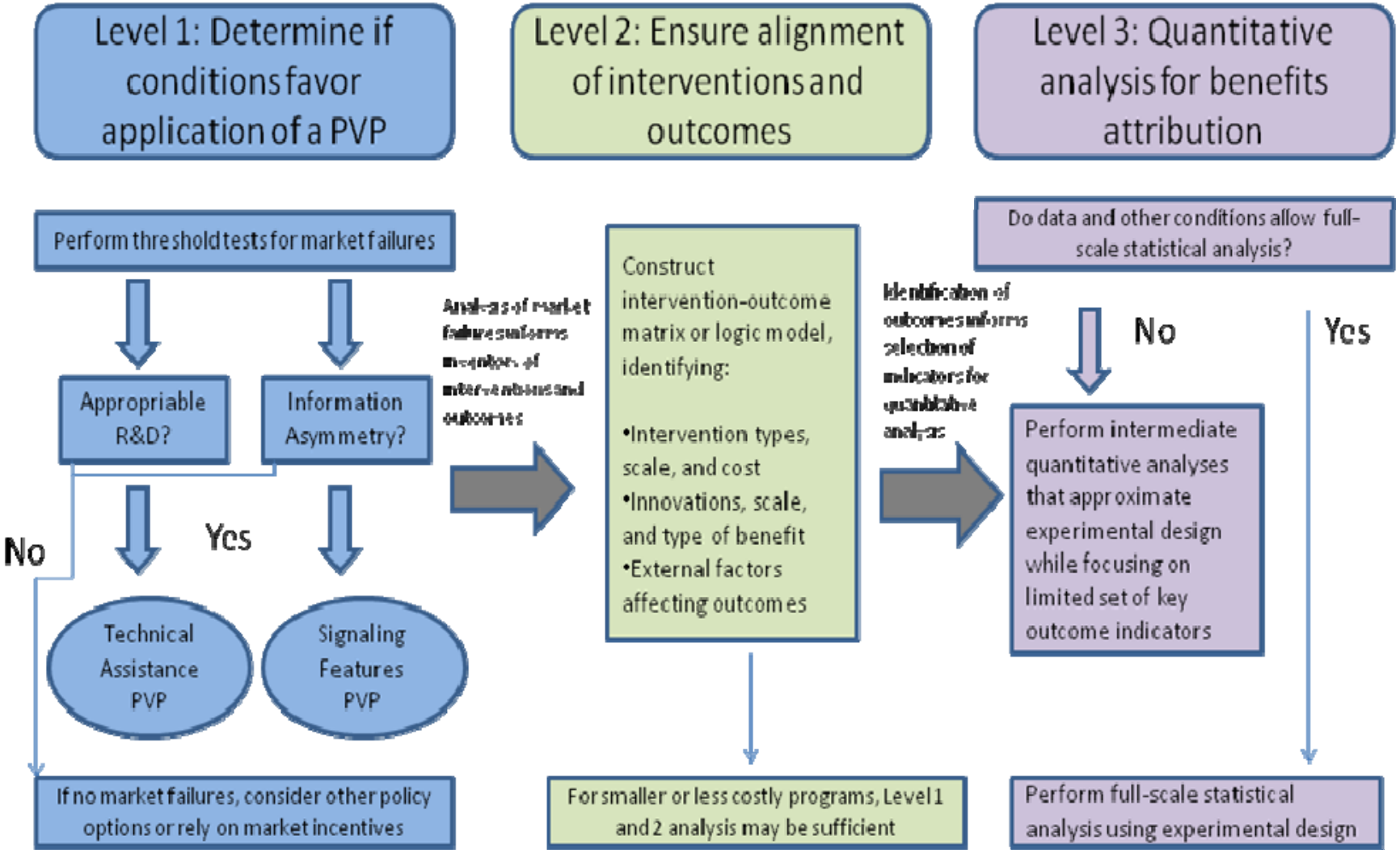
This document outlines the challenges associated with benefit attribution in the context of EPA voluntary programs. The analysis suggests a variety of arguments and evidence that programs can use to demonstrate the benefits of ongoing or planned initiatives. The proposed methods are intended as a practical alternative to the two-stage method or experimental design. Although strict statistical approaches represent an ideal method for rigorously demonstrating program benefits, a variety of obstacles can impede their application. In response, this paper recommends a more practical approach to attribution that relies on core principles of economic public policy analysis and data that are more readily available.

The proposed methodological approach is summarized in diagram below. As shown, the three major levels are as follows.

- **An initial threshold assessment** to determine whether economic conditions (i.e., market failures) favor the application of a voluntary program, such as one providing technical assistance or signaling features.
- **An assessment of interventions and outcomes** to help verify that the resources and activities invested in a PVP are logically aligned with the desired outcomes.
- **A quasi-experimental design and quantitative analyses** that effectively attribute the benefits of a PVP and adhere to basic program evaluation principles while avoiding the strict demands of the two-stage method.

EPA program offices can pursue these three assessment levels to the degree consistent with the nature of the programs under consideration and the resources available. For instance, a small voluntary program may require only the application of an initial threshold assessment, whereas larger, more costly programs may justify the application of all three levels of analysis.

Three-Level Benefits Attribution Approach for ORCR Voluntary Programs



References

Akerlof, George. 1970. The Market for Lemons: Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics* 84: 488–500.

Alberini, Anna. 2007. Determinants and Effects on Property Values of Participation in Voluntary Cleanup Programs: The Case of Colorado. *Contemporary Economic Policy* 25(3): 415–432.

Borck, Jonathan, and Cary Coglianese. 2009. Voluntary Environmental Programs: Assessing Their Effectiveness. *Annual Review of Environment and Resources* 34: 305–324.

Boyd, James. 1998a. *Searching for the Profit in Pollution Prevention: Case Studies in the Corporate Evaluation of Environmental Opportunities*. EPA 742-R-98-005. Washington, DC: U.S. Environmental Protection Agency.

Boyd, James. 1998b. The Benefits of Improved Environmental Accounting: An Economic Framework To Identify Priorities. Discussion paper 98-49. Washington, DC: Resources for the Future.

Boyd, James. 2000. Corporate Evaluation of Pollution Prevention Opportunities. In *Empirical Studies of Environmental Policies in Europe*, edited by Jurgen von Hagen and John Maxwell. Amsterdam, the Netherlands: Kluwer Academic Publishers, 193-210.

Chu, Edward, and Ian Lange. 2008. Evaluating Market-Making Voluntary Programs: The Case of Coal Combustion Products Partnership, powerpoint presentation, on file with authors.

Darnall, Nicole, and Joann Carmin. 2005. Greener and Cleaner? The Signaling Accuracy of U.S. Voluntary Environmental Programs. *Policy Sciences* 38: 71–90.

Epstein, Marc. 1996. *Measuring Corporate Environmental Performance*. Montvale, NJ: Institute of Management Accountants.

Ferraro, Paul. 2009. Counterfactual Thinking and Impact Evaluation in Environmental Policy. In M. Birnbaum & P. Mickwitz (Eds.), *Environmental Program and Policy Evaluation.New Directions for Evaluation, 122*, 75–84.

Greer, Linda, and Christopher van Loben Sels. 1997. When Pollution Prevention Meets the Bottom Line. *Environmental Science and Technology* 31: 418–422.

Griliches, Zvi. 1992. The Search for R&D Spillovers. *The Scandinavian Journal of Economics* 94(Supplement): 29–47.

Hardin, Garrett. 1968. The Tragedy of the Commons. *Science* 162(3859): 1243–1248.

Heckman, James J. 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47(1): 153–161. http://links.jstor.org/sici?sici=0012-9682%28197901%2947%3A1%3C153%3ASSBAAS%3E2.0.CO%3B2-J (accessed August 15, 2010).

IEc (Industrial Economics, Incorporated). 2008a. Analysis of the Feasibility of Statistical Approaches for Performance Evaluation for the NPEP Program. Memorandum submitted to EPA Office of Resource Conservation and Recovery, August 21.

———. 2008b. Draft Literature Review of Approaches to Estimating Attribution of Voluntary Program Benefits. Memorandum submitted to EPA Office of Resource Conservation and Recovery, February 25.

Lange, Ian. 2008. Evaluating Voluntary Programs with Spillovers: The Case of Coal Combustion Products Partnership. U.S. EPA, National Center for Environmental Economics, Working Paper # 08-12.

Lyon, Thomas, and John Maxwell. 2008. Corporate Social Responsibility and the Environment: A Theoretical Perspective. *Review of Environmental Economics and Policy* 2(2): 240–260.

Mankiw, N. Gregory. 2008. *Principles of Microeconomics*. Mason, OH: Cengage Learning.

Moilanen, Tuula, and Christopher Martin. 1996. *Financial Evaluation of Environmental Investments*. Houston, TX: Gulf Publishing.

Ostrom, Elinor. 1990. Governing the Commons: *The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press.

Porter, Michael E. 1995. Green and Competitive: Ending the Stalemate. *Harvard Business Review* Sept.–Oct.: 120–134.

Ross & Associates Environmental Consulting, Ltd. 2009. NPEP Member Interviews: Synthesis of Findings. Prepared for Industrial Economics, Incorporated, under contract to EPA Office of Resource Conservation and Recovery, December.

Sarokin, David, Warren R. Muir, Catherine G. Miller, and Sebastian Sperber. 1985. *Cutting Chemical Wastes: What 29 Organic Chemical Plants Are Doing To Reduce Hazardous Wastes*. An Inform Report. New York: Inform.

Spencer, Roger, and William Yohe. 1970. The "Crowding Out" of Private Expenditures by Fiscal Policy Actions. *Federal Reserve Bank of St. Louis Review* October: 12–24.

Stiglitz, Joseph E. 2000. The Contributions of the Economics of Information to Twentieth Century Economics. *Quarterly Journal of Economics* 115(4): 1441–1478.

Tassey, Gregory. 2005. Underinvestment in Public Good Technologies. *The Journal of Technology Transfer* 30(1–2): 89–113.

Thaler, Richard H., and Cass Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

U.S. EPA (Environmental Protection Agency). 1991. *Amoco–U.S. EPA Pollution Prevention Project: Yorktown, Virginia, Executive Summary*. December. Revised May 1992. Washington, DC: U.S. EPA.

———. 1995. *Environmental Cost Accounting for Capital Budgeting: A Benchmark Survey of Management Accountants*. EPA 742-R-95-005. September. Office of Pollution Prevention and Toxics. Washington, DC: U.S. EPA.

———. 2009. *Guidelines for Evaluating an EPA Partnership Program (Interim)*. March. National Center for Environmental Innovation. Washington, DC: U.S. EPA.