Chapter 9: Concluding Observations

Section I: Introduction

This book began with the question of whether voluntary environmental programs work – that is, whether they actually improve environmental outcomes. Breaking the question down further, we have asked: Quantitatively, how large are the gains? Are there significant differences between energy and non-energy related programs? How do participation incentives, as well as the process for setting goals, affect outcomes? How convincing are alternate approaches to baselines and evaluation? We now want to re-examine these questions looking across the seven case studies in this volume.

In looking across the seven case studies, a natural starting point for our examination is to summarize what the authors conclude about the performance of the individual voluntary programs. Do the authors find that participating firms (or households in the case of the California Demand Side Management program) achieved better environmental results than the determined baseline? Is the baseline credible? If so, and particularly if the results are shown to persist over time, was participation in the voluntary program the principal motivating factor, or might other planned or coincidental factors play important roles?

Our value as editors, however, comes from added insights that arise from looking across this collection of studies viewed as a group. Specifically, we can compare the

institutional and regulatory contexts of different voluntary programs, including the extent to which specific 'carrots' and 'sticks' may have influenced program participation and effectiveness; the importance of using voluntary approaches to address energy as opposed to toxics-related issues; and the key methodological concerns relative to program design and evaluation that cut across the studies. Finally, we make a bold attempt to compare the effectiveness of the seven different programs on a quantitative basis. That is, based on the range of case study results, we attempt to draw some general conclusions about the likely range of effective impacts from voluntary programs, with particular attention to transient versus long-lived effects.

Section II: Conclusions of the Individual Authors

In this subsection we briefly describe the different programs and review the key findings and observations drawn by the individual authors.

The 33/50 Program

The U.S. EPA's first voluntary program, 33/50, was established in 1991 amid rising interest in finding a quick, cost-effective, relatively non-controversial approach to address concerns about toxic releases. Focusing on seventeen high priority chemicals reported to the Toxic Release Inventory (TRI), the program emphasized pollution prevention as an environmental management technique. The 33/50 name derives from the program's goals of 33 percent reductions by 1992, and 50 percent reductions by 1995, below a 1988

baseline. According to case study author Khanna, the specific goals were first proposed by EPA administrator Reilly in 1990.

At the outset, industry leaders expressed concern about possible mandatory requirements and praised 33/50 as a preferred alternative to regulation. In part, EPA chose the 1988 baseline to permit companies to take credit for activities already underway and thus not to penalize them for cutting emissions prior to 1991. In fact, more than a quarter of the recorded reductions took place before the official start of the program.

In early 1991, EPA invited the top 600 companies, which accounted for two-thirds of total 33/50 releases in 1988, to participate in the program. Meetings were held with top corporate executives, trade associations and others to motivate action. Subsequently, second and third rounds of invitations were sent to smaller firms and outreach was also conducted. Sixty-four percent of the top 600 firms agreed to participate, far more than the 10-18 percent of the invitees from later rounds. The relatively flexible goals could be met by reductions to air, water or land releases of any of the listed chemicals. Timetables for individual firms were also flexible and there were no requirements for the use of specific abatement methods.

Although some of the reductions were clearly driven by mandatory provisions of the Montreal Protocol and the 1990 Clean Air Act Amendments, covered releases declined considerably between 1988-1995, well in excess of the established goals. The author cites several sophisticated studies, including her own analysis of the chemical industry,

which find that the program had a statistically significant, negative impact on releases of the 17 chemicals targeted by 33/50. A major innovation of these studies was to model explicitly the relationship between a firm's decision to participate and its environmental performance once in the program. These statistical analyses were complemented by case studies documenting that once signed up for 33/50, some participants made significant efforts – often at great cost – to reduce their releases.

In contrast to these affirmative results, Khanna reports on a recent study that found the program's overall impact on toxic releases to be ambiguous and, in some cases, adverse – joining 33/50 increased emissions relative to non-participating facilities in chemicals and a number of other key industries. This study specifically excluded the two ozone depleting chemicals that were being phased out by the Montreal Protocol. While Khanna draws no firm conclusions on these new results, at a minimum, they raise questions about the performance of this program.

Analyses of firms' motivation for cutting their toxic releases suggest that the desire to differentiate themselves from rivals, to garner positive publicity, and to respond to perceived regulatory threats were important in stimulating participation in the program. Some companies simply welcomed formal recognition for the efforts already underway. Not surprisingly, a number of studies found that firms with greater consumer visibility were more likely to participate in the program, e.g., those producing final goods, or in direct contact with consumers, or in industries with higher advertising expenditures per unit of sale, or those operating in states with high environmental group membership.

Other studies found that participation was motivated by the desire to offset adverse publicity or, perhaps, to reduce the frequency of EPA inspections. Interestingly, there were no significant differences in the extent of early reductions carried out by firms signing up in the first round of invitations to join 33/50 versus those invited to join in later rounds. Research also indicates that an environmentally aware public and a credible threat of mandatory regulation if voluntary approaches failed were also important elements, as was the presence of clearly defined numerical goals and mandatory reporting requirements via the TRI.

Japan's Keidanren Voluntary Action Plan on Environment

Japan's Keidanren Voluntary Action Plan on the Environment was initiated by industry in 1997, just prior to the negotiation of the Kyoto Protocol. It encompasses large enterprises drawn from 58 different business associations including the industrial, electricity, construction, commercial, and transport sectors. Neither small/medium enterprises nor households are part of the Keidanren. Based on 1990 data, Keidanren members represented more than four-fifths of the total greenhouse gas emissions associated with industrial activity and electricity generation, and almost one-half of Japan's total emissions.

Under the Plan, non-binding targets were established at the sector level through the individual industry associations and did not apply to individual enterprises. As case study authors Wakabayashi and Sugiyama explain, the Plan was initially embraced by industry

as a means of demonstrating cooperation with government on greenhouse gas emissions while, at the same time, it sought to avoid mandatory requirements. At present, they argue, the non-binding targets are widely recognized as commitments with which industries are to comply.

Although the government has provided tax and other incentives to help address various other environmental problems in Japan, no specific economic incentives are available to Keidanren members to help achieve the Kyoto targets. At the same time, the existence of relatively stable, long-term institutional relationships among the government, individual industries and the Keidanren are critical to the success of the Plan. Because the Keidanren has considerable influence on the operation of Japan's industrial policy, it can help to ensure the performance of its members, since noncompliance would undermine the multi-faceted benefits of the long-term relationships.

Overall, Wakabayashi and Sugiyama note three factors that seem to be motivating industry to comply with the Plan: i) the cooperative relationship between Keidanren and companies, ii) threats of mandatory policies such as tax or cap and trade schemes, and iii) awareness of private companies' social responsibility. As the authors suggest, some of the motivating factors may be unique to the special relationship that exists in Japan between government and business and may not be readily applicable to other nations.

In terms of absolute emissions, Keidanren members are committed to stabilize their collective greenhouse gas emissions at 1990 levels by 2010 – a goal for which they are

now on track. The key question -- confounded by the slowdown in GDP growth during the early years of the Plan -- is whether this goal is significantly different from business-as-usual. Evidence presented in the case study shows that the emissions of the participating industries have fallen by an average of 0.62 percent per year over the period 1997-2004. This contrasts with annual emission *increases* averaging 0.55 percent for the seven preceding years. A 2005 follow-up reported by industry found that for 35 of the individual associations that participate in the Plan – mostly in the industrial and electricity sectors – emissions were slightly below base year levels. Were it not for certain unplanned nuclear shutdowns, the industry estimates that greenhouse gas emissions would have been 2.4 percent below base year levels.

There are a number of reasons why it is difficult to evaluate whether or to what extent the goals of the Plan exceed business-as-usual. Because of its wide coverage it is not possible to identify a comparable set of non-member industries (or firms) against which to compare the environmental performance of the Keidanren members. Time series analysis is hindered by the absence of historical data on emissions by Keidanren members. However, some evidence of the seriousness of the commitment is seen in the actions taken by specific industries. For example, power companies have announced their intention to acquire Kyoto credits from other nations, in case they do not meet the targets. The steel industry has expanded efforts to increase efficiency and reduce raw material waste, in some cases with governmental support. In 2002 Keidanren established its own evaluation committee to oversee the performance of its members. To date it has issued a number of recommendations to avoid double counting, adopt more standardized

assumptions, and others. At the same time, the NGO community has raised various concerns about the stringency of the targets, and the lack of consistency and transparency of the process.

In sum, the authors present somewhat of a mixed picture on the environmental success of the Plan. Even though it is labeled as voluntary, there are clearly some subtle, semi-binding aspects to it. While at this point emissions of the covered industries are on track with the stated goals, there is uncertainty about the future. Moreover, despite the difficulty in documenting the situation, there is some concern that the stated goals may not differ greatly from business-as-usual.

UK Climate Change Agreements

The UK has been an early and strong supporter of the Kyoto Protocol, and has adopted a proactive position on climate change, both domestically and internationally. In various white papers the government set out a far-reaching strategy for achieving and then moving beyond the Kyoto targets and timetables, aiming to reduce CO_2 emissions by 20 percent below 1990 levels as an interim step in reaching a long-term goal of 60 percent reductions by about 2050. In laying out the context for the case study, authors Glachant and de Muizon note the large-scale substitution of gas for coal in both the electricity and industrial sectors that occurred in the early and mid-1990s, driven largely by air pollution issues and the privatisation and restructuring of the electricity sector.

In 2001 the UK government established voluntary, quantified, climate change agreements

(CCAs) with 48 sectoral associations in the industrial, commercial, and public sectors as part of a complex policy mix involving an energy tax, a climate change levy, and an emissions trading system. One of the stated goals of the 5-10 £ ($9-18) tax per ton of CO_2 emissions was to transfer the burden of taxation from employees to CO_2 emissions (the double dividend argument) and, simultaneously, to shift the burden away from energy intensive industries. The mechanism used to accomplish these goals was to reduce various employment-based taxes alongside implementation of the CO_2 levy. This, in turn, was combined with the use of the CCAs, which then exempted participating firms from 80 percent of the levy.

When first taking on a CCA, firms could choose either intensity-based or fixed targets (most chose the former), expressed in terms of either energy use or carbon emissions. Overall, the CCAs cover about 12,000 individual sites – virtually all those eligible -- representing almost 44 percent of total UK industry emissions.

Compliance with the CCA could occur via reductions in energy use or by purchase of emission rights in the recently established pilot emissions trading program. Because most companies were reluctant to enter into a scheme based on collective compliance, several options were offered. The most popular one involved two phases wherein verification is first made for the sector as a whole. If the sector's target is met all firms are deemed to be in compliance. If the sector target is not met, the performance of individual firms is then examined.

Citing a government-sponsored study, the authors argue that the selected targets were considerably below business-as usual. However, as they also note, the same study indicates that the targets were actually less stringent than the expected levels with the full CO_2 tax in place, i.e., without the CCA tax exemption. Further, the authors reference a study by the Association for Energy Conservation which challenges the initial finding that the targets were significantly below business-as-usual levels. Yet, there is no clear basis on which to judge the merits of the competing assessments.

Regardless of how the targets were set, aggregate emissions during the first two years of implementation were well below them. The case study authors then develop an evaluation methodology to assess the contribution of the CCAs to the implied emission reductions based on the observed prices for emission credits and on the behavior of individual firms. Although a government-sponsored study finds widespread compliance with the CCAs, given the low observed credit prices and the relatively small number of transactions, the authors conclude that the CCAs were only modestly effective in encouraging reductions beyond business-as-usual. At the same time, they observe that the use of permit prices as a measure of the effort required to meet the targets may understate the true effectiveness of the CCA if, in fact, the voluntary agreements were effective in reducing organizational and informational barriers that may have hindered the operation of profitable abatement actions.[1] Unfortunately, no specific evidence is available to test this hypothesis.

Denmark's Voluntary Agreements on Energy Efficiency

Beginning in 1996, the Danish Energy Agency established voluntary agreements on energy efficiency as part of a set of revenue-neutral CO_2 and other 'green' taxes imposed on the industrial, trade and service sectors. After an initial phase-in period, the full tax rate was 13.3 Euros ($18) per ton of CO_2. Lower rates were applied to energy-intensive firms and those most vulnerable to foreign competition. Virtually 100 percent rebates were given to energy intensive-firms if they entered into a voluntary agreement on energy efficiency with the Energy Agency. The voluntary agreements were thus considered complements to the tax scheme. If companies failed to follow through on their agreement, there was an explicit sanction: they had to repay the rebate in full. As case study authors Krarup and Millock note, the rebate effectively lowered taxes on industry to the point where they were only about one-third as much as those on households.

Although the voluntary agreements did not involve quantitative targets, rebates were initially conditioned on the completion of verified energy audits and the implementation within three years of all measures estimated to have a payback that exceeded given criteria. If no energy savings projects were identified in the audit, the company was considered to be energy efficient and need not carry out new projects in order to qualify for the rebates. Participating firms were also eligible for a subsidy of up to 50 percent of the cost of the audits. The investment criteria for the efficiency measures (generally a 4-6 year payback) were somewhat more relaxed than those typically applied to private investments. Measures to be undertaken include energy savings projects, so-called special investigations, and energy management systems such as those specified in ISO

14001. Environmental and consumer groups had only limited involvement in the voluntary agreements.

From the outset, issues arose concerning the high administrative costs of the program. As a result, the voluntary agreements were revised in 2000 and again in 2003, including expanding coverage to additional industries, dropping the requirement for the verified energy audit, and emphasizing adoption of energy management systems as opposed to audit-based requirements.

Government-sponsored evaluations of the impact of the agreements suggests a 2.6% reduction in energy use over the period 1996-99 and 1.9 percent over the period 2000-2003. Other studies estimate somewhat higher reductions, on the order of 3-8 percent, based on econometric analysis of a panel of both participants and non-participants. However, given the small number of participants in the sample (~2%), the authors interpret these quantitative results with considerable caution. Qualitatively, in a number of interviews, company managers generally thought the program had favorable impacts on energy use.

Overall, the authors argue that the true savings are likely quite modest. Importantly, the evaluations also found that the most profitable investments were realized in the agreements implemented in the early years of the program (prior to 2000). In the future, as the EU ETS is implemented, the authors see only a very limited role for this type of voluntary agreement in Danish greenhouse gas mitigation policies.

German Cement Industry

In 1995, the Federation of German Industries, a group of 16 industrial associations representing major sectors of German industry, voluntarily issued a "Declaration of German Industry on Global Warming Prevention (GGWP)" which called for voluntary reductions in specific fuel consumption (e.g., energy per unit of output) of up to 20 percent below 1987 levels by the year 2005. The initial commitments undertaken by the German industries did not involve any government-provided incentives, nor were they accompanied by threats of future regulation. The initial GGWP was simply a unilateral commitment by the nation's principal industries to reduce their emissions. By the year 2000, five years in advance of the target date, most of the commitments of the individual industrial associations were already fulfilled, an indication, in the view of case study authors Bohringer and Frondel, that the targets were not very ambitious. Subsequently, as result of pressure by the government and the desire of industry to avoid mandatory requirements, the GGWP goals were made more stringent – a 28 percent reduction in energy-related specific CO_2 emissions (e.g., emissions per unit of output) below the new base year 1990 by 2012. Moreover, the declaration was joined by other industry associations.

For purposes of evaluating the effectiveness of the GGWP, the case study authors emphasize the importance of establishing a credible baseline. For that reason they single out the cement industry for detailed analysis because it is the only one among the 19

industries now in the German Federation for which sufficient historical data are available to compare the CO_2 emissions of the industry following development of the GGWP to emissions in prior years.

Using data for the period 1974-1995, which begins after the 1973 oil shocks and ends prior to the start of the GGWP, the authors estimate annual fuel efficiency improvements of 0.63 percent in the cement industry. By extrapolating this trend forward and comparing the observed to the predicted levels of fuel efficiency, Böhringer and Frondel conclude that the energy and emission-reducing activities 'have not gone much beyond good intention.' Indeed, they find that the carefully monitored performance of the cement industry during the period of the GGWP does not deviate significantly from business-as-usual, noting that the margin of error on the forecast business-as-usual baseline is +/- 5 percent. More generally, the authors argue that because of asymmetric information between the industry and government, it is quite difficult for outsiders to decide whether or not the commitments made by the cement or other industries push beyond the levels that would have occurred without the voluntary agreements in place. That is, the impossibility of observing the counterfactual, along with the absence of pre-declaration data that could be used to develop a credible assessment, make it relatively easy for industry to declare the effort a success. Performance monitoring, by itself, they argue is insufficient to gauge the effectiveness of the program.

As regards policy recommendations, the authors call for the use of firm-specific targets as opposed to aggregate, industry-wide agreements. Further, they argue that such

agreements, if negotiated between firms and regulators rather than being unilaterally set by industry, have a far better prospect of encouraging emission reductions beyond business-as-usual.

Climate Wise

Climate Wise is a voluntary program with the non-utility industrial sector developed by the U.S. EPA to encourage the reduction of CO_2 and other greenhouse gases. Originally established in 1993, Climate Wise remained in operation until 1999-2000 when it was renamed and placed under the Agency's Energy Star umbrella. Unlike EPA's well known technology-based programs, e.g., Green Lights, which require the adoption of specific technologies, Climate Wise members have the flexibility to adopt whatever technologies or strategies they choose to reduce their emissions. The requirements are simply that a participating firm develop baseline emission estimates of its greenhouse gases (GHGs) for any year since 1990, self-designate forward looking emission reduction actions, and make periodic progress reports. To assure that the proposed reductions were substantial in nature, EPA provided a checklist of major actions such as specific boiler modifications, waste heat recovery systems and others. Firms were strongly encouraged to select at least some of their proposed actions from this list. Participants were also encouraged to report their progress to the U.S. Department of Energy through the 1605(b) registry program, and to provide a copy of the completed form directly to EPA.

EPA offered several kinds of technical assistance to participating firms, including a guide to industrial energy efficiency, various government publications on energy efficiency, and free phone consultation with government and private sector energy experts retained as consultants by the Agency. EPA also set up an annual event open to the public to recognize the performance of outstanding Climate Wise Participants. As part of these events, a series of workshops were held which allowed participating firms to exchange experiences about their efforts to improve industrial efficiency and reduce GHG emissions. Informal reactions from Agency staff and industry representatives suggested that these workshops were seen as quite valuable by the participating firms. At its peak, Climate Wise had enrolled more than 600 industrial firms covering several thousand facilities nationwide.

While EPA has published estimates of emission reductions associated with Climate Wise ranging from 3 to 20 million metric tons of greenhouse gases, these estimates have been criticized by other researchers. In particular, they expressed concern that the extremely wide range of activities covered by Climate Wise makes the program's role in the decision to undertake these activities somewhat questionable. The principal goal articulated by case study authors Morgenstern, Pizer and Shih is to try to tease out the unique contribution made by Climate Wise, apart from other factors that may have contributed to the reported emission reduction by the program participants.

To conduct their evaluation, the case study authors obtained access to confidential plant-level data files for the manufacturing sector collected by the U.S. Census Bureau. These

files contain technical and economic information on individual plants in multiple industries, e.g., data concerning location, output, energy expenditures, and other relevant variables. The authors first use this data to create a matched set of non-participants, where each Climate Wise participant is paired with another facility with similar observable characteristics. The authors then compare the performance of Climate Wise participants and non-participants.

Although a number of important caveats apply, the principal result reported by the case study authors is that Climate Wise appears to have had little to no effect on fuel use while slightly increasing demand for electricity. Comparing the change in fuel and electricity expenditures before and after a facility joins the program, and then across participants and non-participants over the same horizon – a "difference-in-difference approach" – their best estimate is a temporary (1-2 year), 3 percent decline in fuel use and 6 percent increase in electricity use associated with joining the program. Given a roughly +/- 5 percent confidence interval, the effect on fuel use is not consistently significant.

Given the counter-intuitive result for electricity, the authors discuss a number of reasons why electricity use may have risen. On an econometric level, the variable used to match participants and non-participants on growth – changes in the value of shipments – may be inadequate and we might see faster growth (and therefore increased electricity use) among participant. On a substantive level, it may be that firms are making choices to increase electricity use in order to reduce direct CO_2 emissions. Finally, the fuel expenditure variable may not be a good proxy for greenhouse gas emissions. Although

the focus of the Climate Wise program was on energy efficiency and the reduction of CO_2 emissions, a number of firms in the chemical industry proposed reductions of nitrous oxide (N_2O) and some breweries proposed reductions of methane (CH_4). Other firms may have pursued fuel-switching projects that would not show up as a decrease in expenditures.

The one result that is clear, regardless of the effect on electricity, is that all of the significant effects vanish after 2 years, suggesting any program consequence is temporary.

Residential Demand Side Management (DSM) Programs in California

Beginning in the 1970's, at the instigation of the regulatory authorities, California electric and gas utilities sponsored programs to promote the adoption of energy-efficient technologies and energy-conserving behavioral practices. These programs included general and site-specific information programs, and financial assistance. Case study author Sanstad focuses on information-only programs for single-family dwellings conducted by investor-owned utilities in the 1990s, wherein the utility would assess customers' energy consumption patterns, equipment holdings, and other energy-relevant characteristics and, at the same time, offer advice on conservation practices and potential efficiency investments. Programs where explicit financial assistance was offered are not considered. Unlike the other programs considered in this volume, the focus in this case study is on the residential rather than the industrial sector.

The literature on programs of this type has emphasized the need for both undertaking ex post performance measurement in addition to ex ante calculations and for addressing the issues of counterfactual baseline and self-selection. The author notes that California has in recent decades established regulatory requirements for DSM program evaluation to address these issues. He reviews the results of evaluations reflecting these requirements, conducted in the mid-1990s on programs undertaken by three of the state's four large, publicly-regulated investor-owned utilities: Pacific Gas and Electric, Southern California Gas, and Southern California Edison.

For each of the three programs, evaluations compared energy use before and after the program was implemented and applied one or more methods to address selection bias. Nonetheless, the author finds that discerning the true impact of the programs is technically and practically challenging even with state-of-the-art methodological controls. He also notes substantial differences in technical sophistication across the three evaluations. After carefully reviewing the results, he concludes that at least two of the three programs yielded energy savings on the order of several percent that would not have occurred in their absence. This magnitude is consistent with previous findings of the savings accruing from DSM programs, such as these, that do not include financial incentives.

In the more detailed Southern California Edison study, savings are reported to be driven principally by changes in the behavior of households rather than by the installation of

new equipment, e.g., improving maintenance of appliances, or discontinuing use of secondary refrigerators. At the same time, as the author notes, 'while vaguely aware of the energy benefits of the recommended actions, [customers]…do not always act on this knowledge until it is suggested by an expert.' This implies that a key barrier to action by homeowners may not be information *per se*, but information from an authoritative source.

Section III: Observations across studies

[Insert Pizer Morgenstern Chap 9 Table 1.doc here]

In Table 9-1, we attempt to quantitatively summarize and compare the effect of these different programs based on the work by the authors of the individual chapters. At first glance, the estimated effects range from zero in the case of German cement industry to 28% in the case of the 33/50 program. Most of the estimates, however, are in the 5-10% range. How can we understand these results?

Media and activity

As noted early on, a key difference between the 33/50 program and the others in this volume is its focus on toxics rather than energy or energy-related greenhouse gas emissions. Toxic emissions differ from energy and energy-related greenhouse gas emissions in three important ways. First, toxics are typically a local or regional pollutant while greenhouse gases are global. Second, toxics can have a direct, acute effect on

human health. They can also have long-term, chronic impacts, e.g., cancer and heart disease. Meanwhile, greenhouse gas emissions accumulate for many years, affect the climate, and thereby impact ecosystems and overall human well-being over the longer term. Finally, with no practical opportunity for end-of-pipe abatement, reductions in energy-related greenhouse gas emissions often amount to reductions in energy use itself. Given the underlying positive price on energy, there is *always* an incentive to reduce energy use. Toxic emissions, meanwhile, are often an un-priced industrial byproduct whose existence was widely ignored until the 1980s and 1990s.

What does this mean for program effectiveness? It should not be surprising that the potential effect of a voluntary program addressing toxic releases is higher – in terms of percentage reductions – than that of a voluntary, energy-related program. Each of the three reasons cited above implies a greater incentive and potentially lower costs for reducing toxics. As noted by Khanna, in Chapter 2, public recognition and goodwill is often a key motivation for corporate action in a voluntary program. Toxics arguably offer a greater reward in this area because of their more localized consequences. Reductions in mercury releases, for example, provide far greater public relations benefits for firms than, say, lower natural gas usage. This arises in part because households are more familiar with energy issues, making energy use and savings less dramatic, and in part because climate change affects society collectively and globally as opposed to the more localized impacts of most toxics.

In addition to larger benefits, firms may also find lower costs associated with efforts to reduce toxics versus additional energy efficiency. Why? Firms always have an incentive to seek out reductions in energy use because such reductions save money, especially when energy prices are high (as in recent years). The added incentive of a voluntary program to reduce energy use on top of existing energy prices seems less likely to turn up many new opportunities. In contrast, there is no obvious reason firms would have previously sought reductions in toxic emissions until they were alerted to the problem. In this way, reduction in toxic releases might represent a new arena of inquiry – with the consequent possibility that such inquiry might yield significant, low cost reduction opportunities.[2]

More generally, it is important to consider how the kind of environmental hazard and mitigation activity addressed by a voluntary program will affect both the incentives for participation and action, as well as the likelihood that achievable, affordable opportunities will be uncovered. Programs that address local, popular concerns will likely resonate more with companies and their constituents. Similarly, programs that focus attention on a relatively new issue may uncover more opportunities than programs that draw attention to areas of existing concern.

Incentives for participation and action

The preceding discussion points out that the nature of the environmental problem addressed by a voluntary program creates different incentives for both participation and

action. More generally, voluntary programs are structured with different incentives, ranging from information benefits and subsidies, to the threat of mandatory regulation, to the coercive effect of strong trade associations, to – in the case of the UK and Danish programs – specific relief from otherwise burdensome taxes. It is, arguably, the nature and magnitude of the different incentives that creates the strongest distinction among voluntary programs.

Certainly in terms of participation, we see differences. The Climate Wise program, with relatively weak incentives, had relatively low participation rates. Meanwhile, the UK agreements and Keidanren program in Japan, with strong incentives, had almost universal participation. More interesting is the unilateral, incentiveless German program, which covered entire industries, and the Danish program, which offered significant tax incentives, yet saw only partial participation.

To understand these differences we also have to consider the requirements of/for participation. In the Danish program, there were real costs to participating firms associated with the required audits (estimated at 17,000-33,000 Euros per facility). In the German program, there were no costs to participants. Taking such costs into consideration, the pattern of participation makes more sense.

The effect of incentives is perhaps even more interesting when we turn to the actual results. All of the energy-related programs have effects in the 0-10% range, and all but the UK program suggest an effect closer to 5% or less. This is true regardless of how

large or small the incentives for participation or action might be. This 5% estimate was also highlighted in Chapter 8 as being consistently observed across a variety of residential energy programs.

Methodology and Baseline

As we noted in Chapter 1, measuring the effect of these programs requires a baseline. The ideal baseline would come from a randomized experiment, with participation in the voluntary program assigned by pure chance to one group of firms or households, and the remaining non-participants serving as controls. Absent such an experiment, analysts have used one of two general approaches: either a baseline *forecast* based on historical data and/or other information, or a group of non-participants serving as a *control* group.

Among the programs surveyed in this volume, three use a forecast baseline to estimate program effects and four use a control group. Within each of the two general approaches, we also see a variety of implementations. Both the UK and Japanese programs use baselines collaboratively developed by business and government. The UK program, however, was criticized in part because so many industries over-achieved their targets and did so with minimal price incentives. While the Japanese program was subject to less criticism, the basis of the forecast remains unclear.

Böhringer and Frondel take an entirely different approach to providing a forecast baseline. While the commitment itself did not provide a baseline estimate to evaluate the

target (which was achieved in half the time originally planned), the authors simply extrapolated past trends. They demonstrate that the commitment itself represented nothing more than a continuation of historical trends (though this trend itself had a +/-5% margin of error). They conclude that the target did not represent an effort beyond business-as-usual, something they attribute to the fact that the industry came up with the target unilaterally instead of negotiating it with government or other stakeholders.

It is interesting that a similar argument was leveled in 2002 when the Bush administration announced their economy-wide target of an 18% improvement in greenhouse gas emissions per dollar of GDP over ten years. While forecasts at the time suggested a baseline of 14%, critics noted that intensity had improved by 18% over the preceding 10 years.[3] Based on the Böhringer and Frondel approach of simply extrapolating past trends, such a target would be viewed as business-as-usual, even though it represented – at the time – a 4% improvement over the more elaborate forecasts prepared by the Energy Information Administration. Now, as the U.S. finds itself currently on track to meet the 18% commitment, these issues have arisen all over again – as well as new ones. Namely, even if 14% was a reasonable BAU forecast and 18% a reasonable target in 2002, with higher energy prices in 2006, we would expect larger improvements in greenhouse gas intensity.

The underlying problem is that BAU forecasts can be quickly outdated by events completely unrelated to a particular environmental program.[4] Where participation is widespread (e.g., industry-wide programs such as the Keidanren and German industry

targets), there may be no alternative to such an approach. However, when only a portion of the population participates in a voluntary program, it is natural to look at the behavior of non-participants to provide a baseline.

Unfortunately, participation in voluntary programs is not random and, as noted, a typical participant often looks very different than a non-participant. In the Morgenstern, Pizer, and Shih study of the Climate Wise program, the authors find that participants are larger and faster-growing than their non-participating counterparts. In one of the studies surveyed by Sanstad, participating households had higher incomes and dramatically higher frequency of air conditioning (central and room) than non-participants.

Program effects can still be estimated when participant and non-participant demographics differ, so long as those demographics are included in the model and are correctly specified. This is the approach taken in the evaluation of the Danish climate change agreements, the 33/50 program, as well as some of the demand-side management programs just mentioned. An alternative is to more carefully match participants and non-participants, as was done in other demand-side management programs surveyed by Sanstad and in the Climate Wise evaluation.[5] Careful matching can eliminate the demographic differences between program participants and controls. It also puts less pressure on the model specification (although, as the Climate Wise study shows, the matching effort can then be as important as the outcome specification).

A final concern is that even if a non-participant control group *looks* similar to the participants we want to evaluate, there may be unobserved differences that explain why participants joined and non-participants did not. For example, participants may simply be more environmentally concerned – leading them to join *and* to have better environmental performance, but not reflecting a real effect of the program. In this case, we need an explicit (and correct) model to explain how unobserved features – errors – might be related across the participation decision *and* the outcome. The studies reviewed by Khanna and by Sanstad both consider this issue.[6]

Synthesis

Tying these observations together, an interesting result emerges. The context of the program, particularly the additional use of "carrots and sticks" to encourage and strengthen the effectiveness of voluntary programs, appears to have only a limited effect on the measured quantitative results among participants. While we have only one example, much larger differences seem to be associated with the *kind* of problem addressed by the voluntary program – e.g., toxics versus energy-related activities. Meanwhile, incentives do have a major influence on the degree of participation. Given that overall impact of a voluntary program is the product of effectiveness per participants and the number of participants, incentives clearly affect the overall impact of a voluntary program.

This observation about effectiveness emerges despite the wide variety of approaches to constructing baselines, an issue that has pervaded this volume. Concern about the UK baseline, for example, makes us wary of the estimated 9 percent effect – but it is still in the ballpark of the other estimates. Similarly, the estimated zero effect in the German program contains a +/- 5 percent error window. Summarizing, it would be hard to reject the hypothesis that *all* of the energy-related programs had a 5 percent impact among participants.

Should this be surprising? On the one hand, we began this exercise with the idea that context would be important; a key element of our case study design was that author teams were supposed to describe the incentives surrounding the voluntary programs. Some involved tax exemptions and others were not only voluntary, but unilateral on the part of business. On the other hand, energy is already a real expense for firms and households. If we imagine an energy demand elasticity of 0.5, a voluntary effort that increased the cost associated with energy use by 10 percent would only have a 5 percent effect on demand.[7] Is it therefore sensible to imagine a larger effect stemming from a voluntary program?

Even if we see this as a reasonable result, a number of caveats are in order, not the least of which is the limited scope of our analysis – we examined seven voluntary programs among thousands in operation in the U.S., Europe and Japan. Further, there are a variety of softer effects that may be equally if not more important – such as changes in attitude and practices that may pay off in the long run. There are non-trivial margins of error in

these evaluations; in some cases, they are not even specified. In other cases, larger questions lurk in the background (for example, inclusion of regulated toxic releases in the 33/50 evaluation). It is also worth noting that a 5 percent improvement in energy use is not a trivial accomplishment, especially for a voluntary program and especially if participation is fairly broad. Nonetheless, it does suggest certain limitations on such endeavors.

Conclusions

We noted at the outset of this chapter that our discussion of voluntary programs is by no means comprehensive. Yet, we have examined programs in three different regions of the world, with a variety of designs and incentives, and with a range of efforts at evaluation. All but one of the analyses suggested the voluntary program affected behavior; however, focusing on energy-related activities, the effect was less than 10 percent, and more typically closer to 5 percent. One evaluation suggested a zero effect (the German cement industry study) but offered an error bound of +/- 5 percent. We note that even in the case of the 33/50 program covering toxics, where a 28 percent improvement was observed, other studies that excluded substances regulated under the Montreal Protocol and/or domestic regulations found negative effects. A tentative conclusion, therefore, is that voluntary programs have a real but limited quantitative effect, particularly among energy-related activities.

Of course, a 5 percent reduction in energy use or carbon dioxide emissions is not trivial. Some nations' initial efforts under the Kyoto Protocol amount to roughly that order of magnitude. It also represents potentially billions of dollars in savings. Nonetheless, it represents what appears to be an outer limit on what these kinds of programs can achieve.

At the same time, many of the authors have noted important but hard to quantify soft effects: changes in attitudes, management practices, etc., that are viewed by participants and stakeholders as significant steps in improving long-term stewardship. While such effects may not show up as immediately quantifiable reductions in targeted emissions or energy use, they may lead to longer-term, broader improvements in environmental performance.

A key theme in the chapters and our discussion has been the methodology used to evaluate programs. An important lesson, then, is the importance in establishing a reasonable baseline. This involves a careful negotiation of a reasonable BAU forecast or, ideally, identification of a suitable control group for evaluation.

There is some evidence that incentives can affect the magnitude of emission reduction or efficiency improvement. The energy-related programs with the weakest incentives – Climate Wise and the German GWP declaration – had the weakest effects. Those with the strongest incentives, UK and Danish agreements and the Japanese Keidanren program, had the strongest effects. However, the difference is somewhat suspect,

especially because the UK and Japanese programs were evaluated against a forecast baseline. The difference is also small, with all the programs in the 5 percent range.

We also noted that, in contrast to the limited effect on the magnitude of impacts among participants, incentives played a significant role in the level of participation, with some programs with larger incentives and lower barriers to participation having near universal participation. Therefore, despite the lack of a large impact on estimated effects among participants, the fact that the pool of participants is larger means that the overall *impact* (e.g., effect × number affected) is also larger.

After estimating the quantifiable impacts of a voluntary program, a necessary follow up question should be whether the effect persists over time and whether, as the program continues, future participants will continue to find the same gains. Evidence from both the Climate Wise and the California DSM programs suggests that some initial gains may not persist. Analysis of the Danish agreements indicates that, as the most profitable gains were realized and larger firms joined earlier, later gains were smaller.

At the outset, we imagined this volume might offer advice to policymakers and stakeholders hoping to better design voluntary programs. Tentatively, that advice might look like this: create incentives to attract firms into a program, put some effort into negotiating reasonable action and providing a baseline for evaluation, but do not focus on holding participants feet to the fire in order to spur heightened effort. Among energy-related programs, the studies summarized in this book suggest only a slightly stronger

effect where taxes, further regulation, or industry coercion were threatened. That is, there seems to be only limited evidence that stronger incentives lead to stronger effects. Rather, the possibility that a larger group of participants can be inspired to act implies that whatever effect there is will occur over a larger population. The Böhringer and Frondel study offers some guidance on a lower bound, however; the threshold for action should not be so low as to be meaningless.

An important initial consideration in designing a program must be the environmental media and activities being addressed: If it is a novel and previously unstudied area, there may be opportunities for more significant improvements at low cost – for example, toxics. At the same time, if it is an arena that has already been carefully scrutinized, as we might believe to be the case for energy efficiency, such opportunities are less likely.

Summarizing, we find that voluntary programs can affect behavior and offer environmental gains, but in a limited way. By considering the media and activity, as well as the potential incentives that can be brought to bear, stakeholders can make crude assessments of the potential for a voluntary program. A critical step is having a realistic, agreed-upon baseline. In many cases, such programs make sense: when the arguments for mandatory programs are unclear or lacking legal or political support, or where such programs will take considerable time to implement, voluntary efforts can play an important role. However, none of the case study authors found truly convincing evidence of dramatic environmental improvements. Therefore, we find it hard to argue for voluntary programs where there is a clear desire for major changes in behavior.

References

Dahl, Carol. 1993. A survey of energy demand elasticities in support of the development of the NEMS. Washington, D.C.: U.S. Department of Energy (contract De-AP01-93EI23499).

Energy Information Administration (EIA). 2005. Impacts of modeled recommendations of the National Commission on Energy Policy. Washington. D.C.: Energy Information Administration.

White House. 2002. Global climate change policy book, addendum. http://www.whitehouse.gov/news/releases/2002/02/addendum.pdf. (accessed July 31, 2006).

[1] We return to this idea of "soft" qualitative results later in this chapter.

[2] This same line of reasoning partly explains why analyses of multi-gas policies, such as the emissions trading program proposed by the National Commission on Energy Policy, find that nearly half the reductions come from non-energy-related emission, even though such emissions are a much smaller fraction of the total inventory (EIA 2005).

[3] See White House (2002).

[4] The initial National Allocation Plans in the EU ETS have been criticized based on over-compliance, raising similar questions about whether the problem is with the baseline forecast or with events that might have occurred after the forecast was made.

[5] Sanstad provides an interesting history of the selection issue in the context of DSM program evaluation.

[6] The key part of this approach involves finding an instrument, or excluded variable, that only affects the participation decision and not the outcome directly. This allows a identification of an outcome

effect arising solely from participation. In the end, the credibility of this approach rests on the arguments for such a variable.

[7] See Dahl (1993) for a review of elasticities estimates, suggesting 0.5 is fairly typical for the intermediate to long term.

Table 9-1: Quantitative comparison of the effect of voluntary programs on behavior

	Quantity measured	Estimated Effect	Scope	baseline	comment
33/50 Program	Aggregate toxic releases	28%	Participating chemicals facilities	Non-participants with self-selection model	Effect reversed when ODS excluded.
UK Climate Agreements	GHG emissions	9%	Participating industries	Negotiated forecast	Baseline criticized; considerable over-achievement.
Danish Energy Efficiency Agreements	Energy Use	4-8%	Participating facilities	Non-participants	Estimate based on 60 participants.
German Cement Industry GWP Declaration	Energy per unit of cement	0	German cement industry	Econometric forecast using historic performance	Baseline error band is +/- 5%. 2005 target achieved by 2000.
Japanese Keidanren	CO_2 emissions	5%	Participating industries	Keidanren forecast of 2010 BAU	Basis of BAU estimate unclear.
Climate Wise	Fossil energy expenditures	3%	Participating facilities	Matched non-participants	Electricity expenditures estimated to rise 6%. Margin of error is +/- 5% and both effects vanish after 1-2 years.
California Demand Side Managment	Natural gas & electricity demand	2-4%	Participating households	Non-participants	Covers three programs; some evaluations more carefully matched non-participants / controlled for self-selection