NCEE
NATIONAL CENTER FOR
ENVIRONMENTAL ECONOMICS

Evaluating the Effectiveness of Voluntary Programs: Did Ohio's Tox-Minus Initiative Affect Participants' TRI Emissions?

Charles Griffiths, Will Wheeler, and Ann Wolverton

Working Paper Series

Working Paper # 16-05
December, 2016

U.S. Environmental Protection Agency
National Center for Environmental Economics
1200 Pennsylvania Avenue, NW (MC 1809)
Washington, DC 20460
http://www.epa.gov/economics

Evaluating the Effectiveness of Voluntary Programs:

Did Ohio's Tox-Minus Initiative Affect Participants' TRI Emissions?

Charles Griffiths, Will Wheeler, and Ann Wolverton[1] [2]

November 14, 2016

[1] All authors are employed at the National Center for Environmental Economics at the U.S. EPA. Questions or comments on the draft paper can be emailed to the authors at: griffiths.charles@epa.gov, wheeler.william@epa.gov, and wolverton.ann@epa.gov. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Environmental Protection Agency.

[2] We appreciate the very able assistance of Hannah Hennighausen and Nicole Condon on this project. This research benefits from comments made at the AERE Summer Conference, the AERE sessions at the ASSA meetings, and the NAREA annual meeting as well as from seminars at the University of Maryland and George Mason University. We also thank Mike Kelley, Ohio EPA, for helpful discussions.

Abstract: In this paper we evaluate whether Ohio's Tox-Minus Initiative had a discernible effect on participants' emission reductions relative to non-participants. We expect this to be the case if there are private benefits of program participation that outweigh its costs. To investigate whether the Tox-Minus Initiative resulted in greater reductions in TRI-reported air emissions from the top 100 emitters, we use a triple difference approach to compare emissions before and after the program. This is done using both the simple difference in emissions between 2003 and 2012 and a fixed-effects, panel regression. To form an appropriate comparison for participants, we use propensity score matching estimation techniques based on pre-participation attributes. Our results suggest that being invited to the program, regardless of whether a facility joined the Tox-Minus Initiative, produced a significant decline in the absolute level of air emissions. Degree of regulatory attention also appears important, though we find that participants reduced emissions subject to the Clean Air Act by significantly more than non-participants in the post policy period.

Key words: voluntary programs; toxic releases; air emissions; program effectiveness

JEL codes: Q53; Q58

1. Introduction and Related Literature

Ohio has regularly ranked as one of the most polluted states in the United States in terms of reported Toxic Releases Inventory (TRI) emissions. Two possible reasons for this are Ohio's concentration of heavy industry relative to other states and high emissions from electric utilities. The resulting negative media attention that Ohio received as a top emitter led it to initiate a voluntary program, the Tox-Minus Initiative, with the specific goal of relinquishing its high ranking status and "enhance[ing] its image as an environmentally proactive, yet economically competitive state" (Ohio EPA, 2007). In this paper, we examine whether the Tox-Minus Initiative had a measurable impact on participants' TRI emissions relative to non-participants in Ohio.

Ohio's decision to target its polluting facilities through a voluntary program fits with a more general trend towards using voluntary approaches to reduce emissions in the United States. Voluntary approaches have been used to complement existing regulations and as a substitute when environmental regulations are not in place. At last count, the U.S. EPA had almost 40 voluntary initiatives targeting issues ranging from air quality to pollution prevention to energy and climate change (US EPA, 2013). In addition, there are many state and local voluntary programs designed to address environmental issues, although, to our knowledge, no comprehensive list of these programs exists.

A question that is frequently raised in the literature is why a facility would voluntarily reduce its emissions beyond what it would do without the program. Further emission reductions are expected only when the private benefits of participation outweigh the costs. Researchers have hypothesized that voluntary programs may reduce a facility's costs (e.g., by encouraging the adoption of more efficient technologies (Blackman and Boyd, 2002)), increase its market share with green consumers, and/or enhance its reputation with green investors (Arora and Gangopadhyay, 1995; Hamilton, 1995; Arora and Cason, 1996; Khanna, Quimio, and Bojilova, 1998; Konar and Cohen, 2001).[3] Alternatively, facilities may join voluntary programs to influence or delay regulation associated with even larger potential costs (Henriques and Sadorsky, 1996; Segerson and Miceli, 1998; Lutz, Lyon, and Maxwell, 2000; Maxwell, Lyon, and Hackett, 2000; Brouhle, Griffiths, and Wolverton, 2009) or to cover up behavior that has received negative attention, a form of "green-washing" (Harrison, 1999; Kim and Lyon, 2011).

[3] A separate strand of literature looks at why consumers might demand green goods (Delgado, Harriger, and Khanna, 2015; Delgado and Khanna, 2015; Sexton and Sexton, 2014).

In the case of the Tox-Minus Initiative, participating companies may be able to convince the public that they are taking positive action towards reducing their environmental footprint, enhancing their reputation and perhaps even their market share. It is also possible that information gleaned from more regular monitoring of emissions could lead to unforeseen production improvements for some companies. However, given Ohio's goal to reduce its high TRI ranking, participation may be primarily viewed as a way to enhance a company's reputation with state regulators or as a way to forestall a possible regulatory threat. While nothing is stated in the recruitment materials to suggest that this is the case, it is possible that firms reasoned that if Ohio could not reduce TRI emissions voluntarily, it may pursue mandatory reductions. In this case, we might expect some initial differences in emission reductions between participants and non-participants, but because a regulatory threat would target all emitters, such differences would dissipate over time.

While a relatively large literature exists on the effectiveness of Federal voluntary approaches, it empirically evaluates only a relatively small number of U.S. programs (e.g., 33/50, Green Lights, EnergyStar, Climate Wise, and the Strategic Goals Programs).[4] Evidence from the available studies on the effectiveness of national-level voluntary approaches is mixed. For instance, Khanna and Damon (1999) find that participants in the EPA's 33/50 program reduced toxic emissions by more than non-participants, though they fell short of meeting the program's overall reduction goals. (GAO (1994) and Davies et al. (1996) confirm the more modest gains of the program.) Later studies by Vidovic and Khanna (2007, 2011) find little evidence that 33/50 participants reduced emissions by more than non-participants. They point to the ability of participants to count reductions that occurred prior to the start of the program toward their goals as the primary driver of this result. Similar results were found by Gamper-Rabindran (2006), who found that in most industries, participation in the 33/50 program did not lower releases relative to non-participation. However, another set of studies find that participants that were inspected prior to joining 33/50 were both more likely to participate and to reduce emissions as part of the program (e.g., Innes and Sam, 2008; Bi and Khanna, 2012). Finally, Carrión-Flores, Innes, and Sam (2013) find that industries with higher rates of participation in 33/50 experienced less technological innovation (as measured by patent applications) in the future.

GAO (1997) and Horowitz (2004) both examine the Green Lights program and find that it improved energy efficiency. However, evidence indicates that some lighting upgrades by participants are not

[4] See de Vries, Nentjes, and Odam (2012) for a thorough overview of the theoretical and empirical literature on voluntary environmental programs.

attributable to the program. Kim and Lyon (2011) find that participants in the Department of Energy's Voluntary Greenhouse Gas Registry increase emissions, while non-participants decrease emissions. Brouhle et al. (2009) find that while participants in the Strategic Goals Program do not initially reduce emissions by more than non-participants, they make relatively greater strides in the later years of the program. The threat of regulation is a significant factor in explaining emission reductions for both participants and non-participants.

Other studies find less evidence that national-level voluntary programs encouraged additional emission reductions. For instance, King and Lenox (2000) and Vidovic, Khanna, and Delgado (2015) find that Responsible Care did not encourage larger improvements in environmental performance for participants relative to non-participants. Instead, participants may actually have performed worse. Welch, Mazur, and Bretschneider (2000) find the same result for the Department of Energy's Climate Challenge. Morgenstern, Pizer, and Shih (2007) observe that participants in Climate Wise saw only temporary improvements in environmental performance that disappeared after 1 to 2 years. Rivera, de Leon, and Koerber (2006) find no difference in the environmental performance of participating and non-participating ski resorts in the first five years of the Sustainable Slopes Program. Delmas and Keller (2005) are not able to evaluate whether the Waste Wise program led to a decline in the amount of municipal solid waste disposed due to lack of data on non-participants. They examine the propensity of participants to submit annual reports and show that firms with a committed CEO, and those that joined to learn about waste reduction methods or with the objective of improving relations with EPA were more likely to submit reports, while firms that joined later in the program or because it was free were less likely to submit reports.

Even fewer state and local programs in the United States have been examined. This is primarily due to lack of data on the environmental performance of participants before and after the program is in place as well as a lack of information on non-participant behavior. Blackman et al. (2010a) examine two voluntary programs in Oregon designed to encourage the remediation of contaminated sites. They find evidence that highly contaminated sites join the program (not just sites with low levels of contamination) and that regulatory pressure plays a major role in inducing participation in the program. Bui and Kapon (2012) find that state pollution prevention voluntary programs reduced annual TRI releases by 10 to 15 percent for the average facility compared to facilities in states without comparable programs. Kotchen (2012) finds that households in cities that participated in Connecticut's Clean Energy Communities purchased more green electricity than those living in cities that did not join the program.

Mosier and Fisk (2013) examine the city of Fort Collins' Climate Wise program, noting that participants have reduced greenhouse gas emissions in line with program goals. However, no comparison is made to non-participant or pre-program behavior, so one cannot discern whether participants would have made these reductions absent the program.

Since the literature indicates that the effectiveness of voluntary programs varies, it is important to continue conducting research in this area. This paper contributes to the existing literature by expanding beyond the programs that have been examined historically, carefully evaluating the effect of participation as well as merely being invited to the program, and whether regulations that may apply to a subset of emissions act in concert with the voluntary program to reduce participant emissions. While the Tox-Minus Initiative is broad-based – applying to many industries – it differs from many other previously studied voluntary initiative as a state-run program. It also qualifies as a good candidate for study because of the availability of pre- and post-program data on TRI emissions, the focus of the Ohio Tox-Minus Initiative. We have emissions information for facilities that were invited to join but did not, as well as facilities that were not invited to join Tox-Minus, allowing us to compare participants with eligible but non-participating facilities in Ohio.

The paper is organized as follows. Section 2 describes the Ohio Tox-Minus Initiative. Section 3 presents the empirical approach taken in this paper. Section 4 describes the data and variables used in the regressions. Summary statistics are provided in section 5. Results are presented in sections 6, 7, and 8. Section 9 discusses why we did not pursue a regression discontinuity approach. Section 10 concludes.

2. Ohio's Tox-Minus Initiative

In September 2007, the Ohio Environmental Protection Agency invited 100 of the top emitters (as of 2005) to its new Tox-Minus Initiative, giving a month's time to send a commitment letter to join the program.[5] While formal invitations were limited to the top emitters, other polluting facilities could still participate (including those not reporting to the TRI). A total of 53 TRI-reporting facilities ultimately agreed to participate in the Tox-Minus Initiative, including 44 invited facilities and 9 additional facilities

[5] The Ohio EPA identified the top 100 emitters by adding land, air, and water emissions reported to TRI minus releases to publically owned treatment works (POTWs), on and off-site energy recovery, recycling, and treatment to destruction. One-time releases and closed facilities were not considered. Certain waste management facilities were excluded because of their limited ability to affect how much waste they receive from other facilities (and similar reasons) based on the expertise of Ohio EPA (Personal communication, Mike Kelley, Ohio EPA, 12/15/2011).

(Ohio EPA, 2008). Facilities are asked to "identify, evaluate and implement feasible and effective pollution reduction or prevention strategies to reduce waste, air and water-related TRI emissions."

Participants were required to specify their own five-year (voluntary) reduction goal starting from a 2007 baseline, although facilities were allowed more time to meet their goal if necessary. Emissions reduction goals were expressed in different ways by participating facilities. For instance, some facilities expressed their goal as a percentage reduction or as pounds of total TRI releases reduced; others set a goal to reduce releases of a specific chemical or group of chemicals, decrease off-site disposal, or identified a particular process change with or without quantifying the implied change in releases. The Ohio EPA has compiled and made publically available each facility's pollution reduction goal on its website. Facilities were also required to submit a plan to meet these goals by mid-2008, and to provide annual written reports describing progress toward reducing their releases each year beginning in 2009. All 53 participants submitted information to be included in the 2009 progress report on their 2008 activities (Ohio EPA, 2009), but only 42 participants submitted information for the subsequent progress report on emission reductions in 2009 (Ohio EPA, 2010).[6] Facilities are allowed to revise their emission reduction goals, but any changes must be reported to the Ohio EPA.

In exchange for participation in the Tox-Minus Initiative, facilities received public recognition of their participation. The Ohio EPA actively promoted facility success stories in the Tox-Minus Initiative through its annual program report, media reports, and its website. The Ohio EPA also offered facilities technical assistance. This could include a site visit by a non-regulatory arm of the Ohio EPA to help identify opportunities to reduce or prevent pollution. The program explicitly promised that any information gathered during the site visit would not be shared with inspection or enforcement programs.

3. Empirical Approach

We begin by examining whether participants in the Tox-Minus Initiative reduce TRI emissions more than a similar set of non-participants. We use a two stage evaluation process. First, we use propensity score matching techniques based on pre-participation attributes to select a defensible comparison group from non-participating facilities. Second, we use difference-in-differences estimation to investigate whether the Tox-Minus Initiative affected participants' emissions relative to both what occurred prior to the program and the performance of large emitting non-participating facilities in the state of Ohio. It is also

[6] Unfortunately, reports for later years are not available online.

possible that the Tox-Minus Initiative resulted in greater TRI emission reductions from the 100 emitters invited to the program, regardless of participation. We explore this possibility through the same difference-in-differences approach. We then explore the interaction between these two effects using a triple difference estimation that accounts for both invited status and participation.

In these cases, the regression technique selected attempts to compensate for the lack of a true counterfactual: we do not have data to determine what emissions would have been for Tox-Minus participants (i.e., the treated group) if they had not been invited or joined (i.e., been left untreated). The two stage propensity score plus difference-in-differences estimation technique matches a participant with its closest non-participant neighbor and then compares emissions across the two sets of facilities. In each case, non-participants are standing in for a counterfactual that is not directly observable.

For this reason, a difference-in-differences approach requires that the treated group is not too different from the non-treated group, so that any observed differences between them can be defensibly attributed to the policy being evaluated by the model. If the treated and non-treated groups are widely different in their key attributes, we may be over-extending the empirical technique's usefulness. For instance, if facilities participating in the Tox-Minus Initiative have a very different age, industry, or size profile then a difference-in-differences approach on its own may not yield convincing estimates of the Tox-Minus Initiative's effect on emissions. Introducing a first stage to refine the comparison group can help mitigate this concern.[7]

Propensity score matching refines the sample of comparable facilities: A treated facility is matched to a non-treated facility based on pre-treatment characteristics aside from the outcome variable, its TRI emissions. It uses a probit regression where the dependent variable is equal to 1 if the facility joined Tox-Minus and 0 otherwise, and the independent variables are pre-treatment characteristics that may affect a facility's propensity to participate in the program. The predicted probability of joining the program from this regression is the facility's propensity score. When the propensity score is within a defined distance, treated and untreated observations are considered a match – this means that the

[7] Another possible alternative to combining difference-in-differences and propensity score matching is to use an instrumental variable approach to estimate the local average treatment effect. In both cases, the goal is to reduce potential selection bias between the treated and control groups that render comparisons between them invalid. In the context of voluntary programs, obtaining a valid instrument is fairly difficult. In cases where this approach has been taken, researchers have gone to great lengths to build highly detailed facility-level databases. Because the voluntary program we study, Tox Minus, is cross-industry, constructing such a data set is even more challenging. While acknowledging the potential advantages of such an approach, we have not pursued it here.

observed covariate distributions are only randomly different from each other, thus replicating a natural experiment. In this way we are able to assemble a dataset that consists of the treatment group and its nearest neighbors. In other words, propensity matching attempts to separate out the effect of pre-existing differences between the treated and untreated groups. Morgenstern et al. (2007), Blackman et al. (2010b), and Kim and Lyon (2011) use similar approaches in their examination of the U.S. Environmental Protection Agency's ClimateWise, Mexico's Clean Industry programs, and the U.S. Department of Energy's Voluntary Greenhouse Gas Registry, respectively.[8] We examine the robustness of our results to several possible matched samples by matching with and without replacement (i.e., a non-treated observation can be selected more than once if it is the best match for multiple treated facilities vs. only being selected once), as well as varying the distance, or "caliper" of the match.

The difference-in-difference technique then estimates the average treatment effect after the Tox-Minus Initiative is introduced. Emissions in year *t* by facility *i* are denoted $emissions_{it}$. *Tox-Minus* (or TM) is a dummy variable that is set to 1 when a facility is in the treatment group. It captures any remaining pre-policy differences between facilities in Tox-Minus and those in the control-group. *Post Policy* (or PP) is a time dummy variable that is set to 1 in the post-policy time period (2008-2012). It captures any general factors that result in changes in facility emissions behavior over time in both the treated and untreated groups apart from Tox-Minus. When we interact these two variables, (*Tox-Minus*Post Policy)*, we have a dummy variable that is equal to 1 when a facility is in the treatment group in the second period. Finally, we include other covariates, *Z*, and a residual error term, e_{it}.

The basic model is

$$emissions_{it} = \alpha + \beta_1\, \text{Tox-Minus} + \beta_2\, \text{Post-Policy} + \beta_3\, (\text{Tox-Minus*Post-Policy}) + \beta_Z Z + e_{it} \quad (1)$$

The difference-in-difference estimator of interest is the parameter β_3. This is the estimate of the difference between the change in emissions for Tox-Minus participants (TM=1) between the post-policy (PP=1) and pre-policy (PP=0) time periods, and the change in emissions for facilities not in the Tox-Minus Initiative (TM=0) over the same time period. That is,

$$B_3 = (\underline{emissions}_{TM=1,PP=1} - \underline{emissions}_{TM=1,PP=0}) - (\underline{emissions}_{TM=0,PP=1} - \underline{emissions}_{TM=0,PP=0}) \quad (2)$$

[8] For examples of this method applied in different environmental contexts, see also Ferris, Shadbegian, and Wolverton (2014) for an examination of employment effects from Phase I of the SO2 trading program, Fowlie, Holland, and Mansur (2012) for an evaluation of Southern California's RECLAIM NOx trading program, and Greenstone (2004) for an investigation on nonattainment status on SO_2 reductions.

We refer to the time between the pre-policy and the post-policy emissions estimates as the "long difference." *Emissions* is underlined in the equation above to denote that the parameter measures the expected value (or average) difference-in-differences across the two groups.

We can take advantage of the panel nature of our data set by adding facility-specific fixed effects, a_i, and a time trend, *Time*. The inclusion of the fixed effect and time trend means that we can no longer independently identify the coefficient on *Tox-Minus*, so it drops out of the specification (see Bennear and Olmstead, 2008). The covariates, Z, are interacted with the post-policy dummy in the panel regression to avoid having them drop out as well. The basic panel fixed-effects model is

$$emissions_{it} = \alpha + \beta_2\,Post\text{-}Policy + \beta_3\,(Tox\text{-}Minus*Post\ Policy) + \beta_Z\,(Z*Post\text{-}Policy) + \beta_t\,Time + a_i + e_{it} \quad (3)$$

The same approach is used to estimate the average treatment effect of simply being invited to the program, regardless of participation. In this case, the *Tox-Minus* variable in equations (2) and (3) are replaced with an *Invited* variable. We define the dummy variable, *Invited,* to capture pre-policy differences between invited and non-invited facilities and set it equal to 1 if the facility was invited to join the program, regardless of whether the facility actually joined the program or not. As with the *Tox-Minus* variable, we also interact this variable with the dummy variable, *Post-Policy*, which is equal to 1 for an invited facility in the second period. The long difference and panel models are, respectively

$$emissions_{it} = \alpha + \beta_2\,Post\text{-}Policy + \beta_4\,Invited + \beta_5\,(Invited*Post\text{-}Policy) + \beta_Z\,Z + e_{it} \quad (4)$$

$$emissions_{it} = \alpha + \beta_2\,Post\text{-}Policy + \beta_5\,(Invited*Post\ Policy) + \beta_Z\,(Z*Post\text{-}Policy) + \beta_t\,Time + a_i + e_{it} \quad (5)$$

The β_5 parameter is now the differential effect of being invited

To test both effects, being invited to the program and actual participation, we introduce a difference-in-difference-in-difference, or "triple difference," model. In this specification, both the *Tox-Minus* and the *Invited* variables and their interaction with *Post Policy* are included, as well as the interaction of all three of these variables. The long difference model is now:

$$\begin{aligned} emissions_{it} = {} & \alpha + \beta_1\,Tox\text{-}Minus + \beta_2\,Post\text{-}Policy + \beta_3\,(Tox\text{-}Minus*Post\text{-}Policy) + \beta_4\,Invited \\ & + \beta_5\,(Invited*Post\text{-}Policy) + \beta_6\,(Invited*Tox\text{-}Minus*Post\text{-}Policy) + \beta_Z\,Z + e_{it} \quad (6) \end{aligned}$$

In this panel specification (with fixed effects and a time trend) the β_1 coefficient for both *Tox-Minus* and β_4 for *Invited* drop out of the regression since they do not vary over time. The model is then:

$$emissions_{it} = \alpha + \beta_2\, Post\text{-}Policy + \beta_3\,(Tox\text{-}Minus{*}Post\text{-}Policy) + \beta_5\,(Invited{*}Post\text{-}Policy)$$

$$+ \beta_6\,(Invited{*}Tox\text{-}Minus{*}Post\text{-}Policy) + \beta_Z\,(Z{*}Post\text{-}Policy) + \beta_t\, Time + a_i + e_{it} \qquad (7)$$

Table 1 summarizes the various effects of potential interest in order to facilitate interpretation of the coefficients from Equation 7. The coefficients in each cell represent the difference between *Post-Policy* =1 and *Post-Policy*=1 with different combinations of *Tox Minus* and *Invited* activated. For instance, the total post-policy effect (PP=1 vs. PP=0) for those who were both invited (I=1) and participated in Tox-Minus (TM=1) is found in the upper left quadrant of the table:

$$\beta_2 + \beta_3 + \beta_5 + \beta_6 = (\underline{emissions}_{TM=1,I=1,PP=1} - \underline{emissions}_{TM=1,I=1,PP=0}) \qquad (8)$$

When *Tox Minus* and *Invited* are instead set equal to zero, the average post-policy change in emissions for firms that were neither invited nor participated in Tox Minus is captured by β_2 (the unshaded lower right quadrant).

Table 1: Triple Difference - Post-Policy Effects of Being Invited and Participating in Tox Minus

	Invited =1	Invited = 0	Dif-in-dif
Tox Minus = 1	$\beta_2 + \beta_3 + \beta_5 + \beta_6$	$\beta_2 + \beta_3$	$\beta_5 + \beta_6$
Tox Minus = 0	$\beta_2 + \beta_5$	β_2	β_5
Dif-in-dif	$\beta_3 + \beta_6$	β_3	β_6

We can also derive the difference-in-difference effects from equations (3) and (5) using the information in Table 1. The effect of Tox-Minus participation without accounting for invited status, β_3, can be calculated by subtracting the participation effect from the non-participation effect (TM=1 -TM=0) in the column *Invited*=0. Likewise, the effect of being invited to the Tox-Minus Initiative independent of participation, β_5, can be calculated by subtracting the invited effect from the non-invited effect in the row *Tox-Minus*=0. The triple difference coefficient, β_6, is shown in the lower most, darkly shaded right quadrant of the table. This is the" difference" in either set of difference-in-differences in the lightly shaded row or column. Formally, it is defined as:

$$B_6 = [(\underline{emissions}_{TM=1,I=1,PP=1} - \underline{emissions}_{TM=1,I=1,PP=0}) - (\underline{emissions}_{TM=0,I=1,PP=1} - \underline{emissions}_{TM=0,I=1.PP=0})] -$$

$$[(\underline{emissions}_{TM=1,I=0,PP=1} - \underline{emissions}_{TM=1,I=0,PP=0}) - (\underline{emissions}_{TM=0,I=0,PP=1} - \underline{emissions}_{TM=0,I=0,PP=0})] \quad (9)$$

For our other covariates, Z, we include a dummy variable equal to one if the facility is located in a county that is in non-attainment for particulate matter and a dummy for manufacturing to control for some basic pre-treatment differences. For comparison purposes, we first run the difference-in-differences specifications without first limiting the sample based on propensity score matching.

4. Data and Variables

Basic information on invited and participating facilities in the Tox-Minus Initiative is available through online program materials (Ohio EPA, 2007; 2008). Using name and address, these facilities are matched to Toxics Release Inventory. We collect TRI total air emissions data for 2022- 2012 for invited and non-invited Ohio facilities.[9] While facilities are free to establish Tox-Minus pollution reduction goals associated with any media, air emission are the most consistently and widely reported information available in the TRI. We also calculate the proportion of total air emissions stemming from Clean Air Act regulated pollutants since these chemicals, which presumably pose a higher risk to human health, may receive increased scrutiny from EPA and, potentially, result in greater reductions by polluting facilities.

We also utilize several pre-program and facility characteristics collected in the TRI that we expect would affect a facility's propensity to join Tox-Minus: a facility's primary standard industry classification code, and pre-program average air emissions in 2002-2004.[10] We also considered including whether a facility is listed by RCRA as a large quantity generator or serves as a transfer, storage, or disposal facility (TSDF) as a way to proxy for size, but these were not included in the final set of regressions.

In addition to the variables from the TRI, we collect a number of independent variables from other EPA and state databases for consideration for both the propensity score matching and the difference-in-differences regressions. We collect the number of inspections and enforcement actions that occurred at a facility between 2000 and 2004 from EPA's Enforcement and Compliance History Online (ECHO)

[9] The TRI basic data files are available at https://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-data-files-calendar-years-1987-2015. Data used in this paper were downloaded on April 28, 2014.
[10] Historic emissions are considered for two reasons: First, this is another rough way to control for facility size. Second, a large, dirty facility may find joining Tox-Minus implies a lower cost to reduce emissions than a much cleaner facility that has already addressed any low-hanging fruit.

database. Our hypothesis is that a facility already subject to regulatory oversight may be more likely to join the Tox-Minus Initiative as a way to demonstrate improved environmental performance. We collect information on whether the county in which a facility is located was out of attainment for particulate matter between 2000 and 2004 from EPA's Green Book of Nonattainment Areas for Criteria Pollutants.[11] Facilities in non-attainment counties may be subject to greater emission control requirements or face greater regulatory scrutiny than facilities in attainment counties. While we collected county and zip code level data from the 2000 U.S. Census to capture the possibility that demographic characteristics play a role in a facility's decision to join Tox-Minus, we ultimately did not rely on these data in our regressions.

We restrict our sample to facilities that report to the TRI in both the pre-program time frame (2002-2005) and in the final year of study, 2012. This second requirement is particularly important for identifying a comparable set of non-participants. Seventy-eight of 358 non-participants (both invited and non-invited) for which we collected TRI data do not report air emissions in 2012. An investigation of corporate websites and newspaper coverage of these facilities confirms that 60 percent of them went out of business in 2012 or a prior year due to reasons unrelated to emissions (common reasons included the Great Recession, and moving production to another state or country). Only three participants are missing 2012 TRI data (two of which went out of business). In the end, we produced a final data set that consists of 48 (of a possible 53) participants and 264 non-participants, excluding those that exited the industry. Note that we had enough information to include 89 of the 100 invited facilities (a mix of 41 participants and 48 non-participants).

5. Summary Statistics

In Table 2 we present summary statistics for three samples. In the second column are statistics for the 48 Tox-Minus participants in our sample. The third column contains data for the full sample of 264 non-participants. In the fourth column are the summary statistics for a sample of 48 non-participants that were matched (without replacement) to our participants.

We begin with a comparison of Tox-Minus participants to the full sample of non-participants. Not surprisingly, a much higher percent of participants was invited to join the program than non-participants (85 percent vs. 18 percent). Likewise, participants' total air emissions are almost 4 times higher than

[11] The ECHO database is available at: http://echo.epa.gov/ , while the Green Book is available at https://www.epa.gov/green-book/green-book-data-download. Data from ECHO were downloaded on December 23, 2013.The file "PHISTORY" was downloaded from the Green Book on December 14, 2012.

non-participants, on average, in 2003 and 2008. This is expected as the program targeted the highest emitters. While both sets of facilities reduce average air emissions, participant emissions appear to fall much faster between 2008 and 2012; participant emissions are only about 3 times higher than non-participants, on average, in 2012. The key question that we are addressing is whether it is possible to attribute some portion of this greater reduction in emissions to the Tox-Minus Initiative.

Table 2: Summary Statistics for Tox-Minus Participants and Non-Participants
Mean and (Standard Error)

Variables	Participants (48 observations)	Non-Participants (264 observations)	Matched Non-Participants (48 observations)
Invited to Tox-Minus	0.85 (0.36)	0.18 (0.39)	0.71 (0.46)
Total Air Emissions – lbs (2003)	1,006,044 (2,560,128)	272,164 (1,126,518)	801,059 (1,654,005)
Total Air Emissions – lbs (2008)	702,945 (1,601,646)	189,367 (772,110)	624,531 (1,439,092)
Total Air Emissions – lbs (2012)	311,980 (662,110)	100,440 (343,875)	351,697 (653,850)
Manufacturing facility	0.77 (0.42)	0.90 (0.30)	0.71 (0.46)
Inspections – partial or full (2003)	2.29 (2.79)	1.92 (3.19)	3.54 (3.46)
PM non-attainment county (2005)	0.15 (0.36)	0.14 (0.34)	0.10 (0.31)
County population (2000)	261,355 (334,636)	353,260 (425,807)	337,131 (426,943)
County household income (2000)	39,464 (5,677)	41,279 (6,234)	40,398 (6,048)

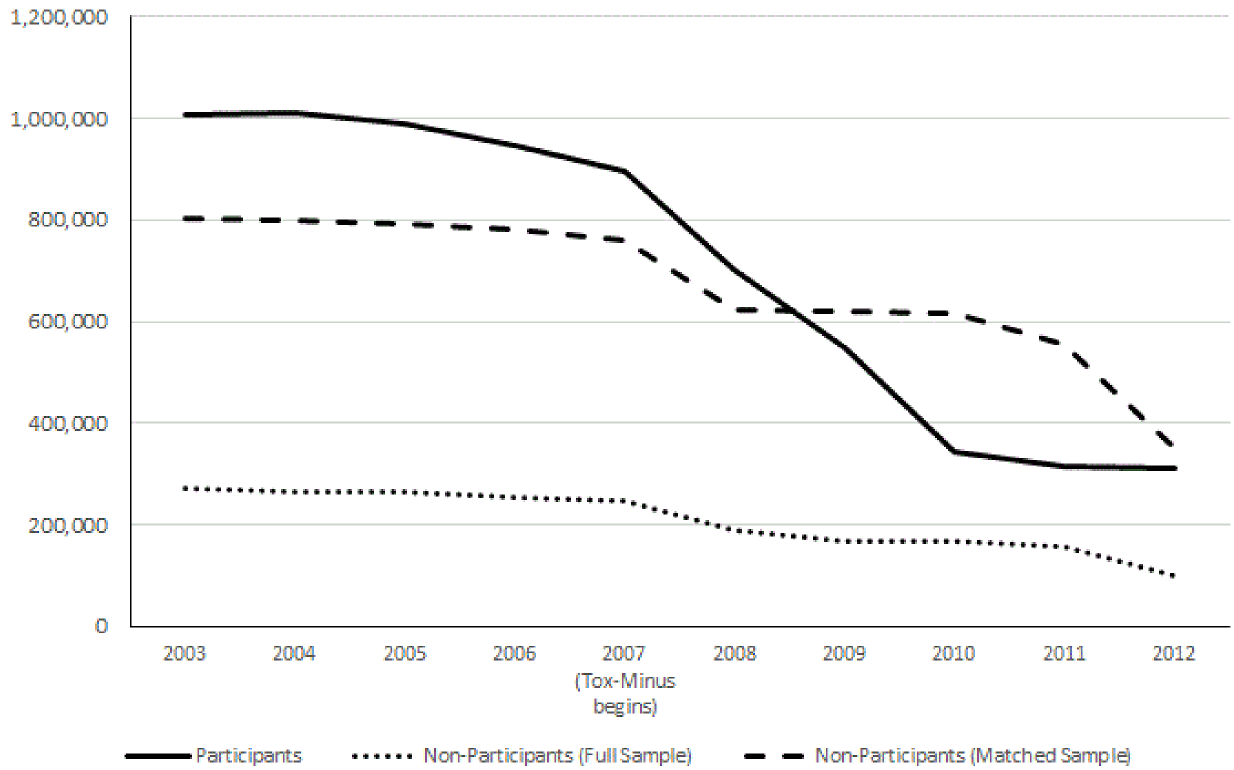
Fewer of the facilities that participate in the Tox-Minus Initiative are classified as manufacturing facilities than in the full sample of non-participants (the main alternative for participants is the electric utility sector). Participating facilities are also more likely to have received a full or partial inspection by EPA in 2003. Facilities look fairly similar across the two samples with regard to national air quality attainment status in 2005. Finally, while household incomes in the surrounding community also appear similar, on average, participating facilities tend to be located in counties with fewer people.

As mentioned above, to more carefully investigate whether the difference in the change in air emissions between participants and non-participants is due to the Tox-Minus Initiative, we also compare participants to a sample of non-participants with similar pre-program characteristics. (This matching estimation use to select this sample is discussed in detail below.) The summary statistics for the matched sample of 48 non-participating facilities is listed in column 4 of Table 2. By design, the matched sample looks more similar to the Tox-Minus participants with regard to pre-program characteristics than does the full sample of non-participants, though there are still distinguishable differences.

Even after matching, the percent invited to join Tox-Minus is lower among the matched sample of non-participants but has increased to 71 percent, compared to 18 percent in the full sample of non-participants. The percent of facilities classified as manufacturing is now 71 percent, which is slightly lower than among participants but closer than the unmatched sample. Non-participant air emissions in 2003 and 2008 are now only slightly lower, on average, than participant air emissions (i.e., between 10 and 20 percent lower). Also note that county attainment status for particulate matter is somewhat lower (indicating more counties in attainment) for non-participating facilities than it was previously, while non-participants in the matched sample now have had more inspections on average in 2003 than participants (the opposite of what occurred in the unmatched sample). Both samples look quite similar with regard to income but continue to differ with regard to population.

Figure 1 reiterates our interpretation of these summary statistics: (1) facilities in all categories – participant, the full sample of non-participant, and the matched sample of non-participants – experience a decrease in average total air emissions reported to the TRI after 2007; (2) Tox-Minus participants and the full sample of non-participants look quite dissimilar both before and after the Tox-Minus Initiative is introduced in 2007; and (3) participants and the matched sample of non-participants exhibit similar trends in average emissions prior to 2007, with Tox-Minus participants decreasing emissions by slightly more on average after 2007.

Figure 1: Total TRI Air Emissions (in pounds) 2003-2012 by Tox-Minus Participation



6. Naïve Results

We begin by running a set of "naïve" difference-in-differences regressions using Tox-Minus participants and the full sample of non-participants. Table 3 presents cross-sectional results using the long difference from 2003 to 2012 with the dependent variable defined in terms of either the level of total air emissions or logged total air emissions.[12] We present three different specifications: a difference-in-differences regression that interacts a participation dummy with the post-policy variable (equation 2); a difference-in-differences regression in which the invited dummy is interacted with the post-policy variable (equation 4); and a triple difference regression in which the participation and invited dummies are individually and in combination interacted with the post-policy variable (equation 6).

Cross-sectional results reported in Table 3 indicate that total air emissions for Tox-Minus participants are not statistically different from the mean over 2003 to 2012 (i.e., *Participant* is not significant). While facilities – both participating and non-participating – exhibit a statistically significant decrease in the

[12] We also examined long differences for levels and logged total air emissions for 2004-2012 and 2005-2012. The sign and significance of the independent variables did not change from what is presented in Table 3 for 2003-2012.

level of total air emissions in only one of the three specifications (the first three columns) after the Tox-Minus Initiative is in place, the coefficient on *Post-Policy* is always significant for logged air emissions (the fourth, fifth, and sixth columns). As expected due to targeting of highly polluting facilities by the Ohio EPA, facilities invited to join the Tox-Minus Initiative have emissions that are statistically different and higher than the mean across all specifications. Manufacturing facilities also have emissions that are statistically significant and lower than the mean, and areas out of attainment status have higher, but not statistically higher, emissions across all specifications.

Table 3: Naïve Difference-in- Difference Estimation Results for Cross-Sectional Long Difference of the Level and Logarithm of Total Air Emissions (p-values in parentheses)

	Total Air Emissions (lbs)			ln[Total Air Emissions (lbs)]		
	(1)	(2)	(3)	(4)	(5)	(6)
Participant (β_1)	206,866 (0.21)	-55,365 (0.66)	-79,339 (0.65)	0.457 (0.35)	-0.152 (0.68)	0.353 (0.50)
Post-Policy (β_2)	-169,364*** (0.04)	-11,899 (0.89)	-17,348 (0.85)	-0.782*** (0.00)	-0.756*** (0.01)	-0.747*** (0.01)
Participant*Post-Policy (β_3)	-524,700*** (0.01)		173,486 (0.67)	-1.223* (0.05)		-0.269 (0.82)
Invited (β_4)	591,845*** (0.00)	1,008,434*** (0.00)	1,018,524*** (0.00)	2.459*** (0.00)	2.832*** (0.00)	2.614*** (0.00)
Invited*Post-Policy (β_5)		-832,322*** (0.00)	-831,143*** (0.00)		-0.744 (0.14)	-0.183 (0.77)
Participant*Invited *Post-Policy (β_6)			-164,215 (0.69)			-0.979 (0.43)
PM attainment	-64,044 (0.57)	-63,552 (0.57)	-60,711 (0.58)	-0.510 (0.13)	-0.510 (0.13)	-0.491 (0.15)
Manufacturing	-1,040,664*** (0.00)	-1,040,974*** (0.00)	-1,042,206*** (0.00)	-1.532*** (0.00)	-1.531*** (0.00)	-1.540*** (0.00)
Constant	1,105,162*** (0.00)	1,026,111*** (0.00)	1,027,615*** (0.00)	10.522*** (0.00)	10.509*** (0.00)	10.498*** (0.00)
Adj. R-Squared	0.2158	0.2387	0.2365	0.2057	0.2036	0.2042

*** Significant at the 1 percent level, ** Significant at the 5 percent level, * Significant at the 10 percent level.

Looking at the first difference-in-difference estimator that interacts *Participant* and *Post-Policy* (first and fourth columns), we find that participants decreased total air emissions significantly more than non-participants in the post-policy period. When we instead interact *Invited* and *Post-Policy* (second and fifth columns), ignoring whether a facility ultimately participated in the program in the post policy period, we find that invited facilities reduced total air emissions by more than non-invited facilities once the Tox-Minus Initiative is in place but this result is only statistically significant when total air emissions are expressed in levels, not in logarithms.

To explicitly examine the relative importance of being invited versus actual participation with regard to emission reductions, we include a triple difference specification (third and sixth columns) that accounts for both participation and invited status in the post-policy period. We find that there is now no statistical difference between participant and non-participant emissions in the post-policy period. Invited facilities reduce total air emissions by more than non-invited facilities in the post-policy period, but this result is still only statistically significant when total air emissions are expressed in levels. The triple difference (the coefficient on *Participant*Invited*Post-Policy*) is not significant in either specification, indicating that there is no additional effect on emissions from the combined effects of being both invited and participating in the Tox-Minus Initiative. These results suggest that simply being targeted by the regulatory agency, rather than actual participation in the program, may have driven the decline in air emissions.

Table 4 includes the same specifications as Table2 but using panel data methods instead of the cross-sectional long difference. All facilities now exhibit a statistically significant decrease in the level of total air emissions (the first three columns) after the Tox-Minus Initiative is in place. However, unlike the cross-sectional results, the coefficient on *Post-Policy* is never significant for logged air emissions (the fourth, fifth, and sixth columns). We continue to find that participants decreased total air emissions more than non-participants in the post-policy period when we interact *Participant* and *Post-Policy* ignoring invited status (first and fourth columns), but it is now only statistically significant when total air emissions are expressed in logs. When we interact *Invited* and *Post-Policy* but do not account for whether a facility ultimately participated in the program (second and fifth columns), we find a result consistent with the cross sectional results: invited facilities reduced total air emissions by more than non-invited facilities in the post-policy period. However, this result is still only statistically significant for levels of total air emissions.

Table 4: Naïve Difference-in- Difference Estimation Results for Panel Regressions (p-values in parentheses)

	Total Air Emissions (lbs) (3080 observations)			ln[Total Air Emissions (lbs)] (3157 observations)		
	(1)	(2)	(3)	(4)	(5)	(6)
Post-Policy (β_2)	-929,128***	-768,378***	-767,158***	0.196	0.216	0.231
	(0.00)	(0.01)	(0.01)	(0.39)	(0.44)	(0.41)
Participant*Post-Policy (β_3)	-286,981		-73,735***	-0.890**		-0.656
	(0.12)		(0.01)	(0.03)		(0.28)
Invited*Post-Policy (β_5)		-375,294***	-346,061**		-0.436	-0.045
		(0.00)	(0.02)		(0.12)	(0.81)
Participant*Invited*Post-Policy (β_6)			5,615			-0.244
			(0.99)			(0.75)
PM attainment*Post-Policy	40,398	35,547	37,026	0.085	0.061	0.094
	(0.23)	(0.32)	(0.35)	(0.82)	(0.88)	(0.80)
Manufacturing*Post-Policy	1,018,381***	910,191***	911,090***	-0.305	-0.338	-0.334
	(0.00)	(0.01)	(0.01)	(0.18)	(0.19)	(0.20)
Time Trend	-19,912***	-20,075***	-20,066***	-0.098***	-0.098***	-0.098***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	409,995***	410,432***	410,412***	10.092***	10.092***	10.092***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Within R-Squared	0.1699	0.1864	0.1868	0.1242	0.1115	0.1245
Between R-Squared	0.2442	0.3018	0.2991	0.0293	0.0903	0.0361
Overall R-Squared	0.0255	0.0356	0.035	0.0053	0.0025	0.0043

*** Significant at the 1 percent level, ** Significant at the 5 percent level, * Significant at the 10 percent level.

Unlike previous results, the triple difference results (third and sixth columns) that account for both participation and invited status in the post-policy period now find a statistical difference between participant and non-participant emissions in the post-policy period but only when total air emissions are expressed in levels. The results for the other interaction terms in the triple difference retain the same sign and significance as in the cross-sectional specifications: Invited facilities reduce total air emissions by significantly more than non-invited facilities in the post-policy period when total air emissions are expressed in levels. The triple difference (the coefficient on *Participant*Invited*Post-Policy*) continues to be insignificant. None of the interaction terms is significant in the logged specification.

Consistent with the results in Table 3, the regressions in Table 4 also indicate that facilities in areas out of attainment status have higher, but not statistically higher, emissions in the post-policy period.

Manufacturing facilities have total air emissions that are statistically different from the mean in the post policy period when emissions are expressed in levels. However, unlike the previous results, this result in no longer statistically significant when total air emissions are expressed in logs.

7. Matched Results

As the summary statistics make clear, there are a number of distinct pre-program differences between participants and non-participants, calling into question whether the full sample of non-participants represent a defensible counterfactual. To more carefully investigate whether the naïve results are due to other factors affecting participants and non-participants during the same time period as the Tox-Minus Initiative, we re-run the triple difference regressions that account for both invited status and participation in Tox-Minus using a sample of non-participants that are matched to participant pre-program characteristics. We use a propensity score matching technique to select the appropriate sample of non-participants.

Table 5 presents the results of this propensity score matching estimation. We predict the likelihood of participation in the Tox-Minus Initiative using the following pre-program characteristics: total average air emissions in 2002-2004, average annual number of inspections in 2002-2004, whether a facility is classified as manufacturing, and a dummy for whether a facility is invited interacted with 2002-2004 average air emissions, 2002-2004 average inspections, and the manufacturing dummy, respectively.[13]

We experimented with including enforcement actions but found no common support across the two samples. Other iterations included whether a facility is a large quantity generator – a proxy for facility size – and a square term for total air emissions. While including these variables did not introduce bias into the matching estimator and were at times statistically significant, they also do not meaningfully affect the difference-in-difference estimation so we have omitted them for simplicity's sake.[14]

[13] The simple, non-interacted dummy of whether a facility is invited to join Tox-Minus is perfectly collinear with at least one of these variables and drops out of the regression.

[14] We also examined whether the difference-in-difference estimates are sensitive to matching with vs. without replacement or the use of a caliper for propensity score matching. We specified a caliper - the maximum distance for a positive match - that is the 0.25 standard deviation of the logit transformation of the propensity score (Stuart and Rubin, 2008), as well as one more precise and one less precise caliper. The difference-in-difference estimation is not significantly affected by differences in matching introduced by these sensitivities.

Table 5: Propensity Score Matching Estimation to Select Sample of Non-Participants

Variables	Coefficient Estimate	Percent Bias Before Match	Percent Bias After Match
Average Total Air Emissions (2002-2004)	3.950E-06** (0.05)	38.7*** (0.00)	9.9 (0.66)
Average Inspections (2002-2004)	-0.070 (0.57)	43.1*** (0.00)	-9.1 (0.69)
Manufacturing facility	-1.067*** (0.00)	-34.2*** (0.01)	16.9 (0.49)
Invited * Avg Total Air Emissions (2002-2004)	-3.880E-06** (0.05)	39.2*** (0.00)	10.5 (0.64)
Invited * Manufacturing	1.663*** (0.00)	116.7*** (0.00)	9.9 (0.68)
Invited * Avg Inspections (2002-2004)	0.147 (0.29)	94.1*** (0.00)	-2.6 (0.92)
Constant	-0.894*** (0.00)		

*** Significant at the 1 percent level, ** Significant at the 5 percent level, * Significant at the 10 percent level.

While we do not want to over-emphasize the coefficients in the propensity score matching estimation, we find it useful to confirm that they have the expected signs. We find that higher average air emissions increase the likelihood that a facility will participate in Tox-Minus. Manufacturing facilities are less likely to participate (compared to electric utilities and other facility types). However, manufacturing facilities that are invited to participate are more likely to join. The average number of inspections – either overall or among invited facilities – is not statistically significant and therefore does not appear to affect the likelihood of Tox-Minus participation. The only counter-intuitive results is that invited facilities with larger air emissions are less likely to join the program. It is also instructive to compare the potential bias[15] in the unmatched sample to that in the matched sample. Column 3 indicates substantial bias in key pre-program characteristics. The matching estimation eliminates statistically significant bias for the included variables, as indicated by a t-test for a difference in the means of the matched sample.

[15] The percent bias is calculated using the Stata command "pstest" and is described as the percent difference of the sample means in the treated and non-treated groups as a percentage of the square root of the average of the sample variances. The command references Rosenbaum and Rubin (1985) for more details.

The results for the triple difference estimation for total air emissions in levels and logs using the matched sample are presented in Table 6. Cross-sectional long difference results are presented in the first two columns, while panel regression results are presented in the last two columns. When Tox-Minus participants are matched to the nearest non-participant based on pre-program characteristics most of the results remain unchanged from the naïve triple difference regressions in Tables 3 and 4. Focusing on the interaction terms, one of our main variables of interest, *Participant*Post-Policy*, is still insignificant in most of the specifications, and continues to be negative and significant in the panel regression using levels of total air emissions as the dependent variable. Unlike the unmatched results the invited dummy interacted with the post-policy dummy is now not significant for either cross-sectional regression, though it continues to be significant for the panel regression using level of total air emissions as the dependent variable. As in the unmatched results, the triple difference coefficient (*Invited*Participant*Post-Policy*) is always statistically insignificant. Coefficients for other variables included in the regressions also retain the same sign and significance as the previous results.

Of the results presented above, we consider those based on the matched sample and panel regression techniques that use the triple difference specification (columns three and four) to be the most defensible. However, the differences in results between the logged and level specifications are somewhat unexpected. When expressed in logs, we find no evidence of a differential trend in total air emissions in the post-policy period between either invited and non-invited facilities or participant and non-participant facilities during the post policy period. On its own, this might lead us to conclude that the Tox-Minus Initiative did not lead to further reductions in air emissions. However, the coefficients on *Participant*Post-Policy* and *Invited*Post-Policy* are both negative and significant in the panel specification when total air emissions are expressed in levels. This suggests that facilities invited to join Tox-Minus may have reduced total air emissions more than non-invited facilities in the post-policy period. In other words, facilities may have reduced their emissions in response to increased attention by the regulator - signaled by an invitation to join Tox-Minus – even when they chose not to directly participate in the program. Likewise, facilities that joined the Tox-Minus Initiative also reduced their emissions significantly more than non-participants when total air emissions are expressed in levels.[16]

[16] To further explore these results, we conducted a number of sensitivity analyses. To ensure that the differences between the logged and level specifications are not an artifact of outliers in emissions (which would have a greater effect on levels), we ran the matching estimation including a dummy variable that identifies observations as potential outliers based on the multivariate BACON (blocked adaptive computationally efficient outlier

Table 6: Difference-in-Difference Estimation Using Matched Sample (p-values in parentheses)

	Cross-Sectional Long Difference		Panel Fixed Effects	
	Total Air Emissions (lbs) (192 obs)	ln[Total Air Emissions (lbs)] (190 obs)	Total Air Emissions (lbs) (948 obs)	ln[Total Air Emissions (lbs)] (944 obs)
	(1)	(2)	(3)	(4)
Participant (β_1)	122,078 (0.69)	-0.084 (0.89)		
Post-Policy (β_2)	-285,640 (0.58)	-1.644 (0.12)	-266,990 (0.23)	0.331 (0.38)
Participant * Post-Policy (β_3)	659,642 (0.39)	0.334 (0.83)	-579,489** (0.05)	-0.771 (0.27)
Invited (β_4)	1,204,831*** (0.00)	2.047*** (0.01)		
Invited*Post-Policy (β_5)	-231,137 (0.70)	0.916 (0.44)	-708,172*** (0.01)	0.027 (0.94)
Participant * Invited * Post-Policy (β_6)	-1,019,282 (0.19)	-1.739 (0.27)	397,768 (0.14)	-0.102 (0.91)
PM attainment	-172,620 (0.60)	0.113 (0.86)		
Manufacturing	-1,369,682*** (0.00)	-1.989*** (0.00)		
PM attainment * Post-Policy			88,287 (0.55)	-0.972 (0.36)
Manufacturing * Post-Policy			1,062,892** (0.02)	-0.221 (0.49)
Time Trend			-48,937** (0.01)	-0.118 (0.00)
Constant	935,809** (0.01)	11.678	978,797*** (0.00)	11.943*** (0.00)
Adj. R-Squared	0.1729	0.1639		
Within R-Squared			0.1973	0.1917
Between R-Squared			0.2447	0.0192
Overall R-Squared			0.0126	0.0511

*** Significant at the 1 percent level, ** Significant at the 5 percent level, * Significant at the 10 percent level.

One possible way to make sense of what may at first appear to be somewhat conflicting results is to re-examine the Tox-Minus Initiative itself. A review of the program indicates that the majority of participants (57 percent) expressed Tox-Minus goals in percentage terms. These facilities also appear to

nominators) algorithm (Billor, Hadi, and Velleman, 2000). While the outlier dummy is positive and statistically significant in the matching estimation, it did not meaningfully affect the log or level estimations.

have made larger reductions in emissions relative to participants that set their goals using other metrics.[17] Unfortunately, this does not help explain why *Participant* Post-Policy* is statistically significant for levels but not logged total air emissions. If anything, we would expect the opposite.

However, invited facilities that did not participate in Tox-Minus may not pay attention to their percent reduction in emissions, since they are not required to set explicit emission goals. They may instead react to the invitation letter as a warning to reduce their overall level of air emissions that are reported to the TRI. If we interpret the change in logged emissions as a measure of the percentage change, then our interpretation is consistent with the finding that the Tox-Minus Initiative may have had a measurable effect on the emission levels of invited facilities even though these reductions did not translate into a statistically significant percent reduction in emissions.

8. Regulatory Attention

To investigate whether the degree of regulatory attention received by facilities from the EPA from 2003 to 2012 affects emissions, we also ran specifications that examined TRI emissions subject to the Clean Air Act (CAA). Brouhle et al. (2009) find that the threat of regulation is a significant factor in explaining voluntary emission reductions for both participants and non-participants. In this case, we may find that highly regulated facilities reduce emissions more in the post policy period regardless of participation in Tox-Minus. On the other hand, if facilities viewed Tox-Minus participation as a way to forestall an increase in stringency, then participants may have lowered TRI emissions by more – particularly of emissions more closely monitored by the EPA – in the post policy period.

To evaluate the first question we add the ratio of CAA to total air emissions as an independent variable in the original specification to capture the possibility that regulatory pressure may induce reductions in emissions separate from those associated with the voluntary program in the post-policy period. We report the triple difference fixed effect panel regression results in the first and second columns of Table 7. We find that facilities that are more highly regulated have significantly lower emissions in the post-policy period regardless of participation in Tox-Minus when the dependent variable is expressed in levels (the first column). Importantly, the interaction between invited and post-policy remains statistically significant even after accounting for the potential role of regulation, though *Participant*Post-Policy* is

[17] While intriguing, our sample size precludes investigating this difference by further restricting our sample to only those that set goals in percentage terms. If we did this, we would only have 27 participants in our sample.

now only significant at the 10% level. The ratio of CAA emissions to total emission is not significant in the post policy period when total air emissions are expressed in logs. This result suggests that once the degree of regulatory attention is accounted for in the regression participation in the Tox-Minus Initiative did not significantly affect total air emissions, which is consistent with the Brouhle et al. (2009) finding. Being invited to the program – which may be a signal of increased regulatory attention - still significantly reduced total air emission levels of both participants and non-participants.

To analyze the second question, we run our preferred triple difference panel fixed effect regressions in both levels and logs using CAA emissions reported to the TRI as the dependent variable instead of total TRI air emissions. These results are reported in the third and fourth columns of Table 7. In this case, we find that participants reduced total air emissions subject to the Clean Air Act in the post policy period by significantly more than non-participants regardless of whether they are specified in level or log terms. Viewed in conjunction with the results reported in columns two and three, this may suggest that Tox-Minus participation and degree of regulatory attention interact in a way that is not completely accounted for when the *CAA Ratio * Post Policy* variable is added to the regression. The interaction between invited and post-policy remains statistically significant only in the levels specification.

If participation is viewed as a way to guard against a possible regulatory threat, it is also possible that the act of supplying information to regulators for annual progress reports, the first of which was released in 2009, had a differential effect on participant emissions. We might expect this to be the case if facilities want to demonstrate to regulators that they are making progress towards their emission goals. Non-participants would not have information revealed to the regulator in the same way and therefore would not face a comparable incentive to reduce emissions. However, we find no evidence that this is the case. Both a dummy variable that defines 2009-2012 as the progress report period and a term that interacts this progress report dummy with participation are insignificant.

Table 7: Regulatory Attention Difference-in-Difference Panel Estimation Using Matched Sample

	Total Air Emissions (lbs) (948 obs)	ln[Total Air Emissions (lbs)] (944 obs)	Total CAA Air Emissions (lbs) (948 obs)	ln[Total CAA Air Emissions (lbs)] (902 obs)
	(1)	(2)	(3)	(4)
Post-Policy (β_2)	63,004 (0.75)	0.174 (0.77)	-277,593 (0.18)	0.432 (0.35)
Participant * Post-Policy (β_3)	-491,484* (0.07)	-0.809 (0.24)	-565,759** (0.04)	-1.328* (0.06)
Invited * Post-Policy (β_5)	-742,376*** (0.01)	0.047 (0.90)	-679,130** (0.01)	-0.109 (0.83)
Participant *Invited * Post-Policy (β_6)	377,064 (0.15)	-0.098 (0.91)	440,442 (0.10)	0.265 (0.72)
CAA Ratio * Post-Policy	-494,793** (0.02)	0.232 (0.77)		
PM attainment * Post-Policy	104,214 (0.50)	-0.978 (0.36)	109,734 (0.41)	-1.142 (0.30)
Manufacturing * Post-Policy	1,033,617** (0.02)	-0.207 (0.48)	1,045,765** (0.02)	0.320 (0.40)
Time Trend	-49,748*** (0.01)	-0.118*** (0.00)	-44,508*** (0.02)	-0.138*** (0.00)
Constant	980,820	11.943*** (0.00)	763,152*** (0.00)	11.150*** (0.00)
Within R-Squared	0.2101	0.1925	0.1912	0.1394
Between R-Squared	0.2668	0.0251	0.2253	0.0017
Overall R-Squared	0.0153	0.0549	0.0087	0.0171

*** Significant at the 1 percent level, ** Significant at the 5 percent level, * Significant at the 10 percent level.

Lyon and Maxwell (2007) suggest that one may be able to isolate effects on participant behavior in the initial years of the program, since they have access to technical assistance via the voluntary program first. If non-participants learn by watching the participant peer group, as the information spreads to and is utilized by non-participants any difference in emission reductions across the two groups would be expected to decrease over time. Alternatively, invitees may view a letter inviting them to join Tox-Minus as a signal that they could be targeted for future regulation. In either case, we might expect some initial differences in emission reductions between participants and non-participants or between invitees and non-invitees, that are targeted by the Tox-Minus Initiative. However, because a regulatory threat would target all emitters, such differences would dissipate over time. We test for this possibility by interacting *Participant* and *Invited* with separate year dummies for each year Tox-Minus was in place. We find no evidence to suggest that either participating or invited plants reduced emissions by significantly more in the early years of the program.[18]

9. Regression discontinuity

Our results indicate that the Tox-Minus Initiative resulted in greater TRI emission reductions from the 100 emitters invited to the program, regardless of participation. Because we have an identifiable - though fuzzy - cutoff point between treated (invited) and non-treated (non-invited) facilities based on total TRI emissions, we evaluated whether we could use a regression discontinuity approach to examine the robustness of our results that increased attention by the regulator - signaled by an invitation to join Tox-Minus – changed TRI emissions behavior.

The regression discontinuity approach relies on the fact that the cutoff point is an arbitrary point and external to the facilities' decisions. It is likely that those just above the cutoff and those below it are similar with regard to pre-treatment characteristics. Only the ad hoc nature of the cutoff point separates the treated from the untreated. This allows us to use non-invited facilities close to the threshold as a reasonable counterfactual for what would have occurred to invited facilities absent Tox-Minus. When the probability of being included in the treatment group increases in the vicinity of the cutoff point, but there is a blending of treated and untreated observations at the cutoff point, the discontinuity is said to be fuzzy. In our case, although the program claims to have invited the top 100 emitters based on 2005 total emissions, we cannot exactly replicate the invited list of facilities using TRI emissions. There appears to have been some other criteria used to invite facilities other than the criteria reported.

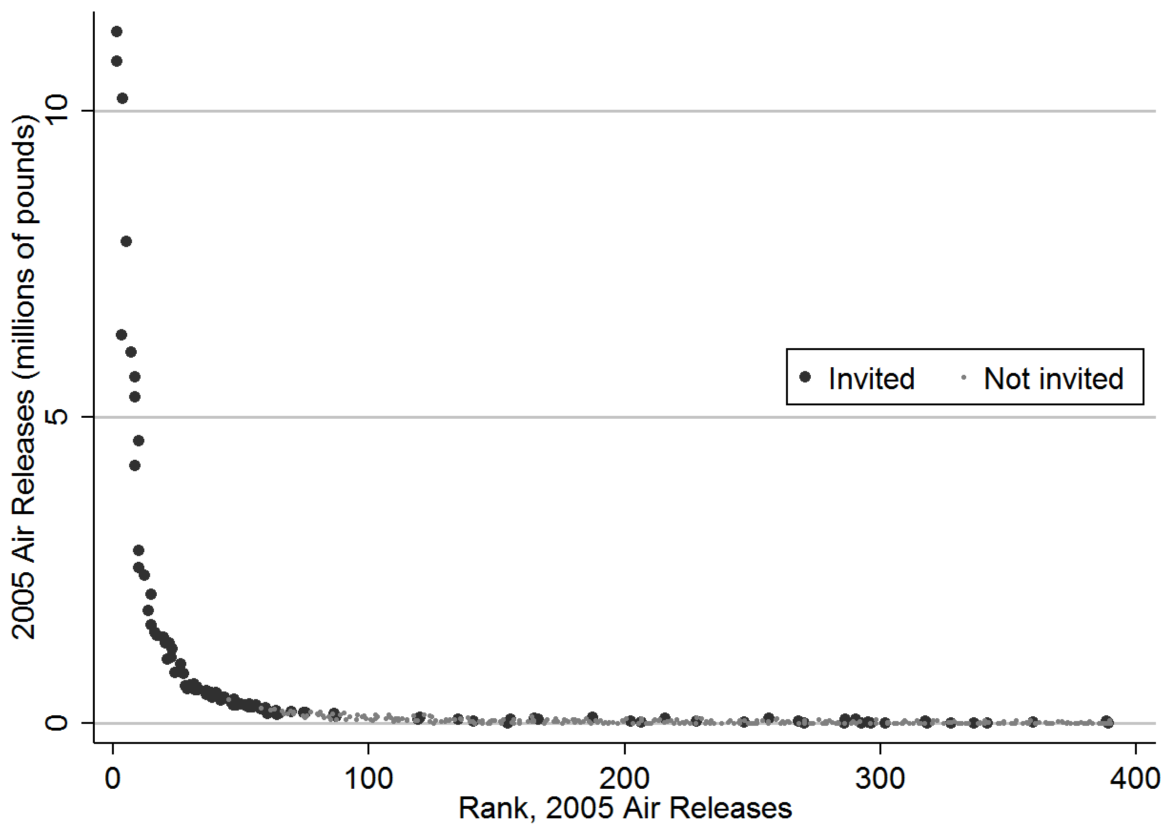
In addition, a regression discontinuity approach is only valid for treated facilities reasonably close to the cutoff point. What qualifies as reasonably close is largely an empirical determination. The further from the cutoff point the more likely that assignment to one side of the cutoff or the other is not random. To assess whether we have a sufficient number of Tox-Minus facilities with emissions close to the cutoff,

---

[18] Alternatively, a firm may join the program to signal to stakeholders that it is acting to reduce its negative impact on the environment (and perhaps should not be subject to new regulation). However, since the program uses readily accessible public TRI information, a facility doesn't necessarily need to formally join the program to signal that it is reducing emissions. As such, higher emitters may have joined the program to strengthen their signal to stakeholders and the regulator, while lower emitting non-participants would not have seen the value of joining because their signal to stakeholders and the regulator via TRI is already clear.

we plot the total TRI emissions in 2005 by rank in Figure 2. The figure shows invited (the large black dots) and not invited facilities (the smaller gray dots). Total emissions in 2005 for the 100th ranked facility in our sample was around 183,000 pounds. Using a bracket of 91,500 to 366,000 pounds, which is half of the emissions of this 100th facility as one end and twice these emissions for the other end, only 21 invited facilities remain in the sample, only 10 of which participated in the program. Centering the cutoff on the emissions of the 93rd facility (since we only have 93 invited facilities), produces a bracket of 101,000 to 404,000 pounds, and a sample of 61 facilities, 25 of which were invited to join Tox-Minus and 13 of which chose to participate. Given the nonlinear nature of the emissions data, it is our opinion that the sample is too small to allow for a robust regression discontinuity analysis comparing invited to non-invited facilities in a way that would shed light on the overall effectiveness of the Tox-Minus Initiative.[19]

Figure 2: Rank of Total TRI Emissions (in pounds), 2005



10. Conclusion

---
[19] The use of total TRI releases does not change this conclusion.

Voluntary programs such as Ohio's Tox-Minus Initiative attempt to achieve the laudable goal of encouraging environmental improvement without imposing formal regulation, and the raw data might suggest that Ohio succeeded. Total air emissions by the 48 Tox-Minus participants in our sample dropped from an average of over one million pounds in 2003 to around 312,000 pounds in 2012. Average emissions of the 272 non-participants in our sample dropped from around 264,000 pounds to 97,500 pounds. A naïve cross-sectional difference-in-difference model that controls for whether a facility is invited to join Tox-Minus, county attainment status for particulate matter and manufacturing facilities indicates a statistically significant decline in the level of total air emissions for program participants. However, further analysis suggests a subtler interpretation.

Expanding the naïve approach to control for the post-policy emissions of those invited to the program suggests that it may be the invitation to participate that motivates the decline in emissions and not necessarily participation in the Tox-Minus Initiative itself. In other words, it may be that attention by the Ohio EPA and the fact that the facility was explicitly recognized as a top 100 emitter were the motivating factors for reducing emissions. Using a panel instead of a cross-sectional long difference we find that both participation and invited status are significant in the post policy period when levels of total air emissions is used as the dependent variable. However, these results disappear when the natural logarithm of air emissions is used as the dependent variable.

One problem with these conflicting results is that it may not be appropriate to compare the 48 Tox-Minus participants with all the non-participants in our sample. To address this concern, we match each treated facility with its most similar non-treated facility based on a probit regression of pre-treatment characteristics. Summary statistics and a measure of bias suggest a very close match between these two groups. Running the panel fixed effect models on this matched sample produces similar results to the panel version of the naïve model: both the invited group and Tox-Minus participants had a statistically significant decline in air emission levels in the post-policy period but not in logged emissions.

The invited group continues to show a significantly negative effect in the post-policy period only in the levels regressions. One possible explanation for the difference in the invited results across level and log specifications is that many of these facilities do not necessarily set emission reduction goals, only participants do. It is possible that invited facilities concentrated on reducing emission levels to minimize regulatory attention but since they were not required to set an explicit emissions reduction goal (the majority of which were stated in percentage terms) this did not necessarily translate into a noticeable difference in the percent reduced.

Additional analyses that focus on regulatory attention as proxied by the ratio of CAA to total air emissions a facility reports to the TRI resolves the discrepancy between levels and logs for participants in the post policy period. Including the ratio of a facility's Clean Air Act emissions to its total emissions on the right hand side of the regression suggests a statistically significant decline in the level of air emissions as the ratio of Clean Air Act emissions rises regardless of participation. Directly controlling for the extent of CAA regulation in this way renders Tox-Minus participation consistently insignificant in the post policy period. However, we find that participants reduced emissions by significantly more than non-participants in the post policy period for both the level and log specifications when we restrict the dependent variable to only CAA emissions.

References

Arora, Seema, and Timothy Cason. 1996. Why do firms volunteer to exceed environmental regulations? Understanding participation in EPA's 33/50 program. *Land Economics* 72: 413-432.

Arora, Seema, and Shubhashis Gangopadhyay. 1995. Toward a theoretical model of voluntary overcompliance. *Journal of Economic Behavior and Organization* 28(3): 289-309.

Bi, Xiang, and Madhu Khanna. 2012. Reassessment of the impact of the EPA's voluntary 33/50 program on toxic releases. *Land Economics* 88: 341-361.

Billor, Nedret, Ali S. Hadi and Paul F. Velleman. 2000. BACON: Blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis* 34: 279-298.

Blackman, Allen and James Boyd. 2002. The economics of tailored regulation: Will voluntary site-specific performance standards necessarily improve welfare? *Southern Economic Journal* 169(2), 309-326.

Blackman, Allen, Thomas P. Lyon, Kris F. Wernstedt, and Sarah R. Darley. 2010a. What drives participation in state voluntary cleanup programs? Evidence from Oregon. *Land Economics* 86(4): 785-799.

Blackman, Allen, Bidisha Lahiri, William Pizer, Marisol Rivera Planter, Carlos Muñoz Piña. 2010b. Voluntary environmental regulation in developing countries: Mexico's Clean Industry program. Journal *of Environmental Economics and Management* 60: 182-192.

Bennear, Lori, and Sheila Olmstead. 2008. The impacts of the 'right to know': Information disclosure and the violation of drinking water standards. *Journal of Environmental Economics and Management* 56 (2): 117-130.

Brouhle, Keith, Charles Griffiths, and Ann Wolverton. 2009. Evaluating the role of EPA policy levers: An examination of a voluntary program and regulatory threat in the metal-finishing industry. *Journal of Environmental Economics and Management* 57 (2): 166-181.

Bui, Linda T. M., and Samuel Kapon. 2012. The impact of voluntary programs on polluting behavior: Evidence from pollution prevention programs and toxic releases. *Journal of Environmental Economics and Management* 64: 31-44.

Carrión-Flores, Carmen, Robert Innes, and Abdoul G. Sam. 2013. Do voluntary pollution reduction programs (VPRs) spur or deter environmental innovation? Evidence form 33/50. *Journal of Environmental Economics and Management* 66(3): 444-459.

Davies, Terry, Janice Mazurek, Kieran McCarthy, and Nicole Darnall. 1996. *Industry incentives for environmental improvement: Evaluation of U.S. federal initiatives*. Washington, DC: Global Environmental Management Initiative.

de Vries, Frans P., Andries Nentjes, and Neil Odam. 2012. Voluntary environmental agreements: Lessons on effectiveness, efficiency and spillover potential. *International Review of Environmental and Resource Economics* 6: 119-152.

Delgado, Michael S., Jessica L. Harriger, and Neha Khanna. 2015. The value of environmental status signaling. *Ecological Economics* 111: 1-11

Delgado, Michael S., and Neha Khanna. 2015. Voluntary pollution abatement and regulation. *Agricultural and Resource Economics Review* 44 (1): 1-20.

Delmas, Magali, and Arturo Keller. 2005. Free riding in voluntary environmental programs: The case of the U.S. EPA WasteWise program. *Policy Sciences* 38(2): 91-106.

Ferris, Ann E., Ronald J. Shadbegian, and Ann Wolverton. 2014. The effect of environmental regulation on power sector employment: Phase I of the Title IV SO2 trading program. *Journal of the Association of Environmental and Resource Economists* 1(4): 521-553.

Fowlie, Meredith, Stephen P. Holland, and Erin T. Mansur. 2012. What do emissions markets deliver and to whom? Evidence from southern California's NOx trading program. *American Economic Review* 102 (2): 965-993.

General Accounting Office (GAO). 1994. *Status of EPA's Efforts to Reduce Toxic Releases*. Washington, DC: GAO.

GAO. 1997. *Information on the Results of Four of EPA's Voluntary Climate Change Programs*. Washington DC: GAO.

Gamper-Rabindran, Shanti. (2006). Did the EPA's voluntary industrial toxics program reduce emissions? A GIS analysis of distributional impacts and by-media analysis of substitution. *Journal of Environmental Economics and Management* 52(1): 391-410.

Greenstone, Michael. 2004. Did the Clean Air Act cause the remarkable decline in sulfur dioxide concentrations? *Journal of Environmental Economics and Management* 47(3): 585-611.

Hamilton, James T. 1995. Pollution as news: Media and stock market reactions to the Toxic Release Inventory data. *Journal of Environmental Economics and Management* 28: 98-113.

Harrison, Kathryn. 1999. Talking With the donkey: Cooperative approaches to environmental protection. *Journal of Industrial Ecology* 2(3): 51-72.

Henriques, Irene, and Perry Sadorsky. 1996. The determinants of an environmentally responsive firm: An empirical approach. *Journal of Environmental Economics and Management* 30(3): 381-95.

Horowitz, Marvin J. 2004. Electricity intensity in the commercial sector: Market and public program effects. *Energy Journal* 25(2): 115-137.

Innes, Robert, and Abdoul G. Sam. 2008. Voluntary pollution reductions and the enforcement of environmental law: An empirical study of the 33/50 program. *Journal of Law and Economics* 51: 271-296

Khanna, Madhu, and. Lisa A. Damon. 1999. EPA's voluntary 33/50 program: Impact on toxic releases and economic performance of firms. *Journal of Environmental Economics and Management* 38(1): 1-28.

Khanna, Madhu, Wilma Rose H. Quimio, and Dora Bojilova. 1998. Toxics release information: A policy tool for environmental protection. *Journal of Environmental Economics and Management* 36(3): 243-266.

Kim, Eun-Hee, and Thomas P. Lyon. 2011. Strategic environmental disclosure: Evidence from the DOE's voluntary greenhouse gas registry. *Journal of Environmental Economics and Management* 61:311-326.

King, Andrew A., and Michael J. Lenox. 2000. Industry self-regulation without sanction: The chemical industries' Responsible Care program. *Academy of Management Journal* 43(4): 698-716.

Konar, Shameek, and Mark A. Cohen. 2001. Does the market value environmental performance? *Review of Economics and Statistics* 83(2): 281-89.

Kotchen, Matthew J. 2012. Climate policy and voluntary initiatives: An evaluation of the Connecticut Clean Energy Communities Program in *The design and implementation of U.S. climate policy*, D. Fullerton and C. Wolfram (Eds.). Chicago: University of Chicago Press, 145-154.

Lutz, Stefan, Thomas P. Lyon, and John W. Maxwell. 2000. Quality leadership when regulatory standards are forthcoming. *Journal of Industrial Economics* 48(3): 331-348.

Lyon and Maxwell. 2007. Environmental Public voluntary programs reconsidered. *Policy Studies Journal* 35(4): 723-750.

Maxwell, John W., Thomas P. Lyon, and Steven C. Hackett. 2000. Self-regulation and social welfare: The political economy of corporate environmentalism. *Journal of Law and Economics* 43(2): 583-618.

Morgenstern, Richard, William A. Pizer, and Jhih-Shyang Shih. 2007. Evaluating voluntary U.S. climate programs: The case of Climate Wise in *Reality check: The nature and performance of voluntary environmental programs in the United States, Europe and Japan*. R. Morgenstern and W. Pizer (Eds.). Washington, DC: RFF Press.

Mosier, Samantha, and Jonathan Fisk. 2013. Can local voluntary environmental programs 'work'? An examination of Fort Collins' Climate Wise program. *Environmental Management* 51(5): 969-987.

Ohio EPA. 2007. Ohio EPA Tox-Minus Initiative Implementation Strategy. September 17. Accessed on November 14, 2016. http://web.epa.ohio.gov/ocapp/tox-minus/Tox-Minus%20Implementation%20Strategy%20Final%209-17-07.pdf.

Ohio EPA. 2008. Listing of Ohio facilities participating in Ohio EPA's Tox-Minus Initiative. Accessed on November 12, 2016. http://web.epa.ohio.gov/ocapp/tox-minus/080229%20Tox%20Minus%20Updated%20Participant%20List.pdf.

Ohio EPA. 2009. *Tox-Minus 2008 activities report*. March. Accessed on November 12, 2016: http://web.epa.ohio.gov/ocapp/tox-minus/Tox-minus%20report%20for%20web.pdf.

Ohio EPA. 2010. *Tox-Minus 2009 activities report*. June. Accessed on November 12, 2016. http://www.epa.ohio.gov/Portals/41/tox-minus/100608%20Tox%20Minus%202009%20Activities%20Report.pdf.

Rivera, Jorge, Peter de Leon, and Charles Koerber. 2006. Is greener whiter yet? The Sustainable Slopes program after five years. *Policy Studies Journal* 34: 195- 221.

Rosenbaum, Paul R., and Donald B. Rubin. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* 39(1): 33-38.

Segerson, Kathleen, and Thomas J. Miceli. 1998. Voluntary environmental agreements: Good or bad news for environmental protection? *Journal of Environmental Economics and Management* 36(2): 109-130.

Sexton, Steven E., and Alison L. Sexton. 2014. Conspicuous conservation: The Prius halo and willingness to pay for environmental bona fides. *Journal of Environmental Economics and Management* 67(3): 303–317.

Stuart, Elizabeth A., and Donald B. Rubin. 2008. Best practices in quasi-experimental designs: Matching methods for causal inference in *Best practice in quantitative methods*. J. Osborne (Ed.). CA: Sage Publications, 155-176.

U.S. EPA. 2013. List of programs. Partnership programs. Accessed November 14, 2016. https://archive.epa.gov/partners/web/html/index-5.html.

Vidovic, Martina, and Neha Khanna. 2007. Can voluntary pollution prevention programs fulfill their promises? Further evidence from the EPA’s 33/50 program. *Journal of Environmental Economics and Management* 53: 180-195.

Vidovic, Martina, and Neha Khanna. 2011. Is voluntary pollution abatement in the absence of a carrot or stick effective? Evidence from facility participation in the EPA's 33/50 program. *Environmental and Resource Economics* 52 (3): 369-393.

Vidovic, Martina, Neha Khanna, and Michael S. Delgado. 2015. Third party certification and the effectiveness of voluntary pollution abatement programs: Evidence from Responsible Care. Working paper. http://web.ics.purdue.edu/~delgado2/VKD%202014.pdf

Welch, Eric W., Allan Mazur, and Stuart Bretschneider. 2000. Voluntary behavior by electric utilities: Levels of adoption and contribution of the climate challenge program to the reduction of carbon dioxide. *Journal of Policy Analysis and Management* 19(3): 407-425.