Evaluating the Effectiveness of Voluntary Programs:

Did Ohio's ToxMinus Program Affect Participants' TRI Emissions?

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

December 31, 2013

[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

[2] 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.

Abstract: In this paper we evaluate whether Ohio's Tox-Minus program 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 difference-in-difference approach to compare emissions before and after the program. This is done using both the simple difference in emissions between 2003 and 2011 and a fixed-effects, panel regression. We also examine whether simply being invited to the program, regardless of participation, had an impact. To form an appropriate comparison for our participants, we use a propensity score matching estimation techniques based on pre-participation attributes to select a comparison group. Our results suggest that there may be two effects at play. First, it may be that participation in the Tox-Minus program produced a significant percent reduction in air emissions, possibly because many facilities expressed their reduction goals in percentage terms. Second, it appears that being invited to the program, regardless of whether a facility joined the Tox-Minus program, produced a significant decline in the absolute level of air emissions. A sensitivity analysis including the ratio of Clean Air Act emissions to total emission offers additional support for this second effect.

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

JEL codes: Q53; Q58

I. 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 to "enhance [its] image as an environmentally proactive, yet economically competitive state" (Ohio EPA n.d.). In this paper, we examine whether the Tox-Minus program 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. Currently, the U.S. EPA has almost 40 voluntary initiatives targeting issues ranging from air quality to pollution prevention to energy and climate change (US EPA 2013b). In addition, there are many state and local voluntary programs designed to address environmental issues, although, to our knowledge, no comprehensive list state and local levels exists. These voluntary approaches have been used to complement existing regulations and as a substitute when environmental regulations are not in place.

One 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. We would expect further emission reductions only when the private benefits of program participation outweigh its costs. Researchers have hypothesized ways in which 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, M., W. Quimio, and D. Bojilova 1998, Konar, S. and M. Cohen 2001). 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 et al. 2000, Maxwell et al 2000, Brouhle et al. 2009) or to cover up behavior that has received negative attention, a form of "green-washing" (Harrison 1999, Kim and Lyon 2011).

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 the state's goal to reduce Ohio's 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, ClimateWise, Strategic Goals Programs).[3] 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. 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). 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 attributable to the program. Kim and Lyon (2011) find that participants in the Department of Energy's voluntary Greenhouse Gas Registry increase emissions (although they report the opposite), 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.

[3] See de Vries et al. (2012) for a thorough overview of the theoretical and empirical literature on voluntary environmental programs.

Other studies find less evidence that national-level voluntary programs encouraged additional emission reductions. For instance, King and Lenox (2000) and Welch et al. (2000) find that Responsible Care and ClimateWise, respectively, did not encourage larger improvements in environmental performance for participants relative to non-participants. Instead, participants may actually have performed worse. Likewise, Morgenstern et al. (2007) observe that participants in ClimateWise saw only temporary improvements in environmental performance that disappeared after 1 to 2 years. Rivera et al. (2006) find no difference in the environmental performance of participating and non-participating ski slopes in the first five years of the Sustainable Slopes Program. Khanna and Keller (2005) are not able to evaluate whether the WasteWise 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 (2010) 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. While the program is broad-based –

applying to many industries – it differs from 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 ToxMinus program. Finally, we also have emissions information for facilities that were invited to join but did not, as well as facilities that were not invited to join ToxMinus, allowing us to compare participants with eligible but non-participating facilities in Ohio.

The paper is organized as follows. Section 2 describes the Ohio ToxMinus Program. 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 and 7. Section 8 contains some sensitivity analyses and section 9 concludes.

II. Ohio's ToxMinus Program

In 2007, the Ohio Environmental Protection Agency invited 100 of the top emitters (as of 2005) to join its new Tox-Minus program.[4] 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 program, including 44 invited facilities and 9 additional facilities (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 are required to specify their own (voluntary) reduction goal with a five-year time frame starting from a 2007 baseline, although facilities are allowed more time to meet their goal, if necessary.[5] Emissions reduction goals have been expressed in different ways by participating facilities. For instance, some facilities express 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 identify 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 provide annual written reports describing progress toward reducing their releases each year beginning

[4] The top 100 was determined 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 from the top 100 because of their limited ability to affect how much waste they receive from other facilities.
[5] All program information about Tox Minus is taken from http://web.epa.state.oh.us/ocapp/tox-minus.html.

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). 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 receive public recognition of their participation. The Ohio EPA actively promotes facility success stories in the Tox-Minus Initiative through its annual program report, media reports, and its website. The Ohio EPA also offers facilities technical assistance. This can 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 promises that any information gathered during the site visit will not be shared with inspection or enforcement programs.

III. Empirical Approach

We begin by examining whether participants in the Tox-Minus program reduce TRI emissions more than a similar set of non-participants. We use a two stage evaluation process. First, we use a 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 program 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 possible that the Tox-Minus program resulted in greater TRI emission reductions from the 100 emitters invited to the program, regardless of participation. We explore this possibility through a difference-in-difference estimation that accounts for both invited status and participation.

In both 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 program have a very different age, industry, or size profile than difference-in-differences on its own may not yield convincing estimates of the Tox-Minus program's effect on emissions. Introducing a first stage to refine the comparison group can help mitigate this concern.

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 facilities propensity score. When the propensity score is within a defined distance, treated and untreated observations are considered a match – this means that the 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) and Blackman et. al (2010b) use similar approaches in their examination of the United States' ClimateWise and Mexico's Clean Industry programs, respectively.[6] 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 program is introduced. Emissions in year *t* by facility *i* are denoted $emissions_{it}$. *ToxMinus* (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 (2007-2011). 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, (*ToxMinus*Post Policy)*, we have a

[6] See also Fowlie et al. (2012) for an example of this method applied in a different environmental context, the evaluation of Southern California's RECLAIM NOx trading program.

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

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

The difference-in-difference estimator of interest is the parameter β_3. This estimate is the difference between the change in emission 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 program (TM=0) over the same time period. That is,

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

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 *ToxMinus*, so it drops out of the specification (see Bennear and Olmstead 2008). The basic panel fixed-effects model is

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

In a second specification for both models (1) and (3) above, we add a difference-in-difference estimator to explore the possibility that an invitation to join – even if the facility chooses not join Tox-Minus - also affects emissions post policy. We define the dummy variable, *Invited* to captures pre-policy differences between invited and non-invited facilities. We then interact this variable with the dummy variable, *Post-Policy*, which is equal to 1 when for an invited facility in the second period. The long difference model is

$$emissions_{it} = \alpha + \beta_1\ ToxMinus + \beta_2\ Post\text{-}Policy + \beta_3 ToxMinus*Post\text{-}Policy + \beta_4\ Invited + \beta_5 Invited*Post\text{-}Policy + \beta_Z\ Z + e_{it} \quad (4)$$

In this case, we have two difference-in-difference estimates, one for the program, β_3, and one for being invited, β_5.

$$B_3 = (\underline{emissions}_{TM=1,PP=1} - \underline{emissions}_{TM=1,PP=0}) - (\underline{emissions}_{TM=0,PP=1} - \underline{emissions}_{TM=0,PP=0}) \quad (5)$$

$$B_5 = (\underline{emissions}_{Inv=1,PP=1} - \underline{emissions}_{Inv=1,PP=0}) - (\underline{emissions}_{Inv=0,PP=1} - \underline{emissions}_{Inv=0,PP=0}) \quad (6)$$

Note that in the panel specification (with fixed effects and a time trend) both *ToxMinus* and *Invited* drop out of the regression since they do not vary over time. The covariates, Z, are interacted with the post-policy dummy in the panel regression to avoid having them drop out as well. The model is then

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

$$+\, \beta_Z\, (Z * Post\text{-}Policy) + \beta_t\, Time + a_i + e_{it} \quad (7)$$

Future iterations of this paper will explore whether being invited and participating in Tox-Minus interact, but we have not done so for this version of the paper.

For comparison purposes, we run a difference-in-differences without first limiting the sample based on propensity score matching. For our other covariates, we include a dummy for whether a facility was invited to join Tox-Minus, county attainment status for particulate matter, and a dummy for manufacturing to control for some basic pre-treatment differences.

IV. Data and Variables

We rely on facility and geographic data from several data sets. Basic information on invited and participating facilities in the Tox-Minus program is available through online program materials (Ohio EPA 2007; 2008). Using name and address, these facilities are matched to Toxics Release Inventory (TRI) emissions data for reporting years 2003 to 2011 (US EPA 2013c). This matching exercise produced data for 93 of the 100 invited facilities, 48 of the 53 participants, and 274 non-participants. We also collect TRI emissions data for large emitting facilities not invited to participate in the Ohio Tox-Minus program. TRI data includes emissions to various media (e.g., air, water), but we use total air emissions as our main dependent variable in the difference-in differences equation and restrict our sample to facilities that report to the TRI in at least 2003 and 2011. While facilities are free to establish Tox-Minus pollution reduction goals that are not associated with TRI air emissions, air emission are the most consistently and widely reported information available in the TRI. We also collect the proportion of total air emissions stemming from Clean Air Act regulated pollutants. This is because these chemicals, which presumably pose a higher risk to human health, may receive increased scrutiny from EPA and, potentially, greater reductions by polluting facilities.

We collect several pre-program and facility characteristics that we expect would affect a facility's propensity to join Tox-Minus from the TRI. These are explored as candidate independent variables to inform the propensity score matching process. These pre-program and facility variables include: a facility's primary standard industry classification code; pre-program (2000-2002) air emissions;[7] and the change in emissions between 2000 and 2003. We also consider 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.

We also collect a number of independent variables from other EPA and state databases for consideration for both the propensity score matching and the difference-in-difference models. For instance, we collect the number of inspections and enforcement actions that occurred at a facility between 2000 and 2005 from EPA's Enforcement and Compliance History Online (ECHO) database (US EPA 2013a). Our hypothesis is that a facility that already has a fair amount of regulatory oversight may be more likely to join the Tox-Minus program 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 2005 from EPA's Greenbook of Nonattainment Areas for Criteria Pollutants (US EPA 2012).[8] Facilities in non-attainment counties may be subject to greater emission control requirements or face greater regulatory scrutiny than facilities in attainment counties. To capture the possibility that demographic characteristics play a role in a facility's decision to join Tox-Minus, we collect county level and zip code level data from the 2000 U.S. Census. Not all variables are available at both the county and zip code level. Data includes population (individuals, households, and families); income (per capita, household, and family); median age; percentage of the housing that is owner occupied; and percentage of the population that is white, black, asian, female, age 62 or older, and employed. The attention paid to environmental issues is proxied by the percentage of the facility's Congressional District that voted for the Democratic candidate in the 2006 U.S. House of Representatives election (Ohio Secretary of State 2012).

---

[7] 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. The change in emissions from 2000 to 2005 indicates whether the facility was already reducing emissions prior to the program. Such a facility may be more likely to join the Tox-minus program to get credit for already planned emission reductions or because it is a "good actor that is already making efforts to lower its environmental footprint.

[8] File "PHISTORY" last downloaded from http://www.epa.gov/oaqps001/greenbk/data_download.html, Data as of December 14th, 2012.

V. Summary Statistics

In Table 1 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 274 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 were invited to join the program than non-participants (85 percent vs. 19 percent). Likewise, participants' total air emissions are almost 4 times higher than non-participants, on average, in 2002 and 2007. 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 2007 and 2011: participant emissions are only 2 times higher than non-participants, on average, in 2011. 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 program.

Fewer of the facilities that participate in the Tox-Minus program 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 less likely to have received a full or partial inspection by EPA in 2002. Facilities look fairly similar across the two samples with regard to national air quality attainment status: about half of participating and non-participating facilities are located in a county out of attainment for particulate matter in 2005. Finally, while political affiliation and household incomes in the surrounding community also appear similar, on average, across participating and non-participating facilities, 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 program, we use a propensity score matching technique to select a sample of non-participants with similar pre-program characteristics. (This matching estimation is discussed in detail below.) The summary statistics for the matched sample of 48 non-participating facilities is listed in column 4 of Table 1. 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. For instance, both samples have similar average total air emissions in 2002 and 2007, as well as average number of inspections in 2002.

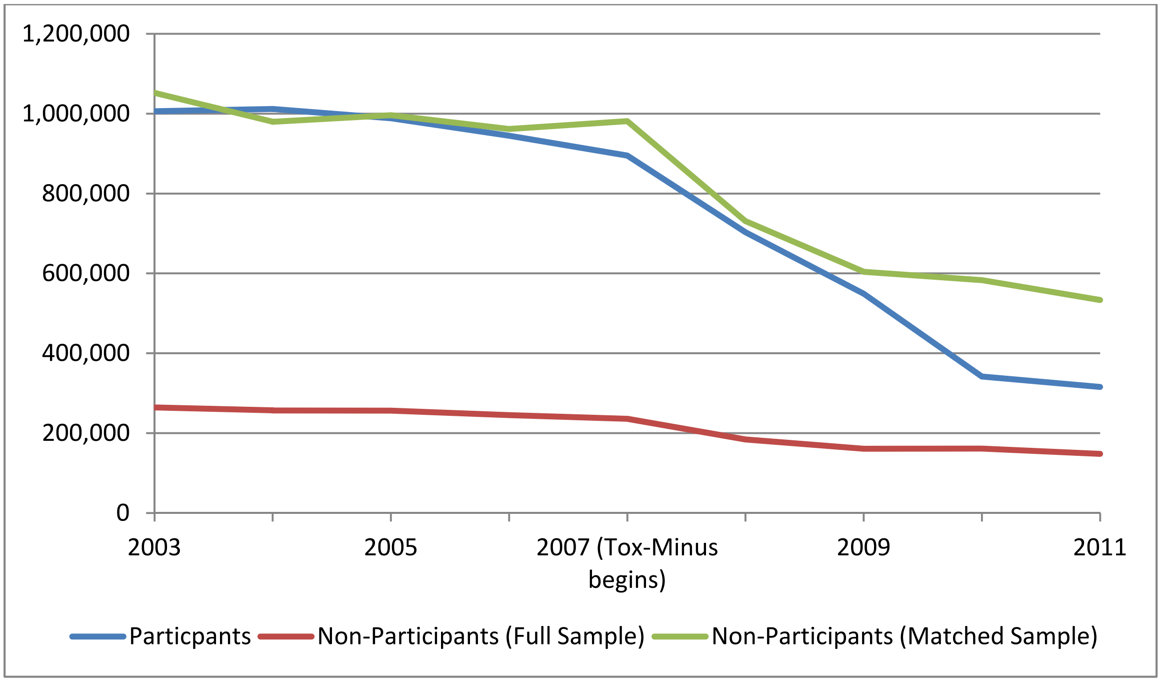
Table 1: Summary Statistics for Tox-Minus Participants and Non-Participants

Variables	Participants (48 observations)	Non-Participants (274 observations)	Matched Non-Participants (48 observations)
Invited to Tox-Minus	0.85 (0.36)	0.19 (0.39)	0.71 (0.46)
Total Air Emissions – lbs (2003)	1,006,004 (2,560,128)	264,246 (1,106,591)	1,051,941 (2,233,663)
Total Air Emissions – lbs (2007)	895,146 (2,189,006)	235,926 (1,047,523)	981,220 (2,224,442)
Total Air Emissions – lbs (2011)	315,474 (647,012)	148,205 (568,354)	533,262 (1,091,529)
Manufacturing facility	0.77 (0.42)	0.91 (0.29)	0.67 (0.48)
Inspections – partial or full (2002)	0.96 (1.43)	0.65 (1.23)	0.92 (1.67)
PM non-attainment county (2005)	0.49 (0.47)	0.52 (0.49)	0.45 (0.49)
County population (2000)	261,355 (334,636)	361,806 (429,971)	265,187 (355,689)
County household income (2000)	39,464 (5,677)	41,312 (6,352)	40,671 (7,082)
Voting democrat (2006)	0.54 (0.15)	0.52 (0.15)	0.50 (0.12)

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 19 percent in the full sample of non-participants. The percent of facilities classified as manufacturing is now 67 percent, which is somewhat lower than among participants. Likewise, county attainment status for particulate matter is also somewhat lower (indicating more counties in attainment) for non-participating facilities than it was previously. Both samples look quite similar with regard to demographic characteristics.

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 program 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) 2003-2011 by Tox-Minus Participation



VI. Naïve Results

We begin by running a set of "naïve" difference-in-difference regressions using Tox-Minus participants and the full sample of non-participants. Table 2 present the estimation results for several specifications: a cross-section using the long difference from 2003 to 2011 with the dependent variable defined in terms of level of total air emissions (equation (1)), and panel fixed effects with a time trend and robust standard errors (equation (3)).[9] We present both specifications twice: in the first instance we account for whether a facility is invited to participate in the Tox-Minus program; in the second instance, we also add the invited dummy interacted with the post-policy variable (equations (4) and (7)).

The cross-sectional results in Table 2 (the second and third column) indicate that total air emissions for Tox-Minus participants are not statistically different from the mean over the 2003 to 2011 time period (that is, the coefficient for *Participant* is not significant). Nor do facilities – both participating and non-participating - exhibit a statistically significant decrease in total air emissions once the Tox-Minus program is in place (the coefficient on *Post-Policy* is not significant). However, as expected due to

[9] We also examined long differences for levels and log total air emissions for 2004 and 2011, and for 2005 and 2011. The sign and significance of the independent variables did not change from what is presented in Table 2 for 2003 and 2011.

targeting of highly polluting facilities by the Ohio EPA, facilities invited to join the Tox-Minus program have emissions that are statistically different and higher than the mean. Manufacturing facilities have emissions that are statistically significant and lower than the mean, and attainment status does not have a statistically significant effect.

Table 2: Naïve Difference-in- Difference Estimation Results for Level of Total Air Emissions

	Cross-Sectional Long Difference		Panel Fixed Effects	
Variables				
Participant	165,986 (163,311)	-27,494 (174,579)		
Post-Policy	-113,183 (82,034)	-1,808 (89,541)	-815,894*** (289,413)	-677,970*** (266,610)
Participant*Post-Policy	-577,387*** (211,973)	-190,249 (246,880)	-280,303 (172,152)	-105,982 (205,632)
Invited	633,275*** (100,840)	925,398*** (139,576)		
PM attainment	130,534 (78,770)	130,763 (78,275)		
Manufacturing	-1,202,262*** (124913)	-1,202,391 *** (123,234)		
Invited*Post-Policy		-583,626*** (194,115)		-280,562*** (101,456)
PM attainment * Post-Policy			-77,280 (56,588)	-85,210 (56,798)
Manufacturing * Post-Policy			923,309*** (317,998)	835,256*** (302,163)
Time Trend			-15,594*** (6,242)	-15,647*** (6,241)
Constant	1,161,973*** (141,338)	1,105,931*** (141,682)	390,197*** (26,058)	390,319*** (25,853)
Overall Adj. R-Squared	0.24	0.25	0.03	0.04
Within R-Squared			0.17	0.18
Between R-Squared			0.25	0.30

Looking at the difference-in-difference estimator (the coefficient on *Participant*Post-Policy*), we find that participants decreased total air emissions significantly more than non-participants in the post-policy period in the first cross-sectional specification. However, once we control for the possibility that invited facilities – regardless of participation – have differential emissions in the post-policy period, this result disappears. Instead, we find that invited facilities reduced total air emissions by more than non-invited facilities once the Tox-Minus program is in place (that is, the coefficient on *Invited*Post-Policy* is statistically significant).

In the panel specifications in Table 2 (the fourth and fifth columns), we find that facilities in the post-policy period exhibit a statistically significant decrease in total air emissions. However, *Participant*Post-Policy* is not significant in either specification. In other words, all facilities exhibit a downward trend in total air emissions once the Tox-Minus program is in place, but Tox-Minus participants do not exhibit a differential trend from this overall effect. When *Invited*Post-Policy* is included we find a result consistent with the cross-sectional result: invited facilities reduced total air emissions by more than non-invited facilities in the post-policy period. Post policy attainment status is not significant in either specification but manufacturing facilities in the post policy period have significantly higher total air emissions.

Table 3 includes three of the same specifications as Table2 but with the dependent variable expressed in terms of log emissions. These regressions perform somewhat less well than those expressed in levels. Consistent with the results in Table 2, the cross-sectional regressions in Table 3 indicate that manufacturing facilities and those invited to join Tox-Minus have total air emissions that are statistically different from the mean over the 2003 to 2011 time period. However, unlike the previous results, all facilities reduced emissions in the post-policy period relative to 2003. Once again, the interaction between participation and the post-policy period (the difference-in-difference estimator) is not significant, but now the interaction between invited and the post policy period is also not significant. Curiously, in the panel specification, the difference-in-difference estimator (that is, the coefficient on *Participant*Post-Policy*) is signficiant, suggesting a difference in the log of air emissions for participants in the post-policy period. None of the other variables of interest are significant in this panel specificiation.

Table 3: Naïve Difference-in- Difference Estimation Results for Ln (Total Air Emissions)

	Cross-Sectional Long Difference		Panel Fixed Effects	
Variables				
Participant	0.38 (0.47)	0.43 (0.51)		
Post-Policy	-0.65*** (0.25)	-0.68*** (0.27)	0.12 (0.30)	0.11 (0.36)
Participant*Post-Policy	-0.97 (0.62)	-1.07 (0.72)	-0.84** (0.41)	-0.86** (0.38)
Invited	2.41*** (0.30)	2.33*** (0.41)		
PM attainment	-0.15 (0.24)	-0.15 (0.24)		
Manufacturing	-2.33*** (0.36)	-2.32*** (0.36)		
Invited*Post-Policy		0.16 (0.57)		0.03 (0.19)
PM attainment * Post-Policy			0.14 (0.20)	0.14 (0.20)
Manufacturing * Post-Policy			-0.36 (0.27)	-0.35 (0.29)
Time Trend			-0.08*** (0.02)	-0.08*** (0.02)
Constant	11.25*** (0.42)	11.27*** (0.42)	10.05*** (0.07)	10.05*** (0.07)
Overall Adj. R-Squared	0.23	0.23	0.01	0.01
Within R-Squared			0.11	0.11
Between R-Squared			0.01	0.01

VII. Matched Results

While the panel specification for logged air emissions indicate a statistically significant effect of the Tox-Minus program, the level air emissions do not demonstrate a similar result. However, there is some summary evidence to suggest that simply being invited to join program produces an impact on post-policy air emissions, even if facilities elect not to participate in the Tox-Minus program. 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 program, we re-run the difference-in-difference regressions 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 4 presents the results of this propensity score matching estimation. We predict the likelihood of participation in the Tox-Minus program using the following pre-program characteristics: total air emissions in 2002, number of inspections in 2002, whether a facility is classified as manufacturing, and a set of interactions between a dummy for whether a facility is invited and 2002 air emissions, 2002 inspections, and the manufacturing dummy variable.[10]

Table 4: Propensity Score Matching Estimation to Select Sample of Non-Participants

Variables	Coefficient Estimate (Standard Error)	Percent Bias Before Match	Percent Bias After Match
Total Air Emissions (2002)	2.55e-6* (1.47e-6)	38.9***	-1.0
Inspections (2002)	0.02 (0.14)	22.7	3.1
Manufacturing facility	-1.18*** (0.33)	-36.3***	28.5
Invited * Air emissions	-2.47e-6* (1.46e-6)	39.3***	-0.5
Invited * Manufacturing	1.71*** (0.28)	112.7***	19.5
Invited * Inspections	0.06 (0.16)	62.3***	5.6
Constant	-0.82*** (0.27)	--	--

*** indicates a coefficient that is significant at the 1 percent level, while * indicates significance at the 10 percent level.

We also experimented with including 2002 enforcement actions but found no common support across the two samples. Other iterations included whether a facility is classified as a large quantity generator – a proxy for facility size – and a square term for total air emissions in 2002. While the inclusion of 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 reporting them for simplicity's sake.[11]

[10] 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.

[11] 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 specify 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

While we do not want to over-emphasize the coefficient estimates in the propensity score matching estimation, we find it useful to at least confirm that they have the expected signs. We find that higher air emissions or number of inspections in 2002 increases the likelihood that a facility will participate in the Tox-Minus program. 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 Tox-Minus program. Among invited facilities, those that have been inspected more often are more likely to join. The only counter-intuitive result is that invited facilities with larger air emissions are less likely to join the program. That said, the coefficient estimate is smaller in magnitude than the positive coefficient on total air emissions.

It is also instructive to compare the potential bias[12] in the unmatched sample to that in the matched sample. Column 3 indicates substantial bias in a number of the key pre-program characteristics. The matching estimation eliminates statistically significant bias for the included variables.

The results for several specifications of the difference-in-difference estimation using the matched sample are presented in Table 5. We only present the models that include *Invited*Post-Policy*. When Tox-Minus participants are matched to the nearest non-participant based on pre-program characteristics we note several interesting similarities and differences when compared to the results from the naïve difference-in-difference regressions (Tables 3 and 4). First, the post-policy dummy is no longer significant in any of the regressions. Second, our main variable of interest, *Participant*Post-Policy*, is still negative and significant in the panel regression using logged total air emissions as the dependent variable. Third, when the invited dummy is interacted with the post-policy dummy it is not significant for the cross-sectional regressions, but it continues to be significant for the panel regression using level of total air emissions as the dependent variable.

As with the naïve difference-in-difference regressions, the dummy variable indicating that a facility was invited to join the Tox-Minus program continues to be positive and significant in the two cross-sectional regressions. Likewise, manufacturing continues to be significant and negative in both cross-sectional

---

and Rubin 2007), 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.

[12] 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.

specifications, and positive and significant for the post-policy period in the panel regression when total air emissions are expressed in levels.

Table 5: Difference-in-Difference Estimation Using Matched Sample

	Cross-Section – Long Difference		Panel Fixed Effects	
Variables	**Level of Total Air Emissions**	**Ln (Total Air Emissions)**	**Level of Total Air Emissions**	**Ln (Total Air Emissions)**
Participant	-107,214 (345,540)	-0.38 (0.54)		
Post-Policy	-15,983 (537,875)	-0.003 (0.87)	-412,509 (286,302)	0.25 (0.59)
Participant*Post-Policy	-68,395 (487,286)	-1.16 (0.77)	-112,850 (258,667)	-0.82** (0.38)
Invited	1,328,831*** (421,013)	2.53*** (0.66)		
PM attainment	269,022 (264,566)	-0.14 (0.42)		
Manufacturing	-1,379,326*** (277,833)	-1.82*** (0.44)		
Invited*Post-Policy	-709,689 (589,366)	-0.52 (0.95)	-602,440*** (220,463)	0.27 (0.38)
PM attainment * Post-policy			-198,987 (194,653)	-0.26 (0.49)
Manufacturing * Post-policy			1,113,438*** (452,047)	-0.48 (0.35)
Time Trend			-46,965*** (20,215)	-0.10*** (0.03)
Constant	909,737*** (431,012)	11.51*** (0.68)	1,078,026*** (82,723)	12.16*** (0.13)
Overall Adj. R-Squared	0.17	0.19	0.01	0.07
Within R-Squared			0.22	0.16
Between R-Squared			0.23	0.04

Of the results presented above, we consider those based on the matched sample and panel regression techniques (columns four and five in Table 5) to be the most defensible. However, the differences in results between the logged and level specifications are somewhat unexpected. When expressed in levels, we find no evidence of a differential trend in total air emissions in the post-policy period generally or between participants and non-participants during this time period. On its own, this might lead us to conclude that Tox-Minus did not lead to further reductions in air emissions among those that joined the voluntary program. However, there does appear to be some evidence that facilities invited to join Tox-Minus reduced total air emissions more than non-invited facilities in the post-policy period. This suggests that facilities may have changed their emissions in response to increased attention by the regulator - signaled by an invitation to join Tox-Minus – even when they choose not to directly participate in the program. The panel regression using logged total air emissions tells a somewhat different story. When specified in this fashion, facilities that joined Tox-Minus show significantly lower logged emissions in the post-policy period, indicating that the program *did* have a direct effect on participating facilities. However, in this specification, an invitation to join Tox-Minus does not seem to matter in the post-policy period.

One possible way to make sense of what may – at first – appear to be somewhat conflicting results is to look back at the Tox-Minus program itself. A review of the program indicates that the majority of facilities (57 percent) expressed their Tox-Minus goals in percentage terms. If we interpret the change in logged emissions as a measure of the percentage change, then it is possible that the Tox-Minus program had a measurable effect on *percent reductions* in emissions of participants but that these reductions did not translate into statistically lower *levels* of emissions. Likewise, 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, but they may react to the invitation letter as a warning to reduce their overall level of air emissions that are reported to the TRI.

VIII. Sensitivity Analyses

To further explore these results, we conduct a number of sensitivity analyses. To ensure that the differences between the logged and level specifications are not an artifact of outliers in emissions, we also re-ran the matching estimation including a dummy variable that identifies observations as potential outliers based on the multivariate BACON (blocked adaptive computationally efficient outlier nominators) algorithm (Billor et al. 2000). While the outlier dummy is positive and statistically significant in the matching estimation, it did not meaningfully affect the log or level difference-in-difference estimations.

We also explore whether 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 2009-2011 as the progress report period and a term that interacts this progress report dummy with participation are insignificant.

Regulatory Attention

To investigate whether the degree of regulatory attention received by facilities from the EPA over the 2003–2011 time period also affects emissions, we ran specifications that examined TRI emissions subject to the Clean Air Act (CAA). If facilities viewed ToxMinus 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. On the other hand, 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 ToxMinus.

To evaluate the first question we run our preferred 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 second and third column of Table 6. The signs and significance of the independent variables do not change from those of the original specifications reported in Table 5, except that *Participant*Post-Policy* is only significant at the 10% level. In other words, Clean Air Act emissions and total air emissions exhibit similar trends over the sample period (i.e., the time trend is negative and significant), and the sign and significance of the post policy period dummy and its interaction with participation remain unchanged from the total air emission specifications.

To analyze the second question, we add the lagged 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 fixed effect regression results in the fourth and fifth column of Table 6. We find that facilities that are more highly regulated have significantly lower emissions in the post-policy period regardless of participation in ToxMinus when the dependent variable is expressed in levels. Importantly, the interaction between invited and post-policy remains statistically significant even after accounting for the potential role of regulation. The ratio of CAA to total emission is not significant in the post policy period when total air emissions are expressed in logs. The coefficient on *Participant*Post-Policy* is now insignificant in both models. In other works, facilities that joined Tox-Minus do not have statistically different emissions in percentage terms than non-participants in the post-policy period. Again, this offers some support to the hypothesis that attention from regulatory agencies may drive reductions in total emissions.

Table 6: Regulatory Attention Difference-in-Difference Panel Estimation Using Matched Sample

	Panel Fixed Effects		Panel Fixed Effects	
Variables	**Level of CAA Air Emissions**	**Ln (CAA Emissions)**	**Level of Total Air Emissions**	**Ln (Total Air Emissions)**
Post-Policy	-467,284 (269,328)	-0.06 (0.85)	-26,264 (258,009)	0.30 (0.78)
Participant*Post-Policy	-45,479 (245,432)	-0.84* (0.45)	-57,847 (254,041)	-0.81 (0.42)
Invited*Post-Policy	-574,004*** (212,810)	0.60 (0.52)	-616,133*** (219,128)	0.27 (0.36)
PM attainment*Post-policy	-177,450 (189,348)	-0.87 (0.55)	-152,504 (185,134)	-0.25 (0.47)
Manufacturing*Post-policy	1,125,534** (426,263)	0.07 (0.50)	1,099,348*** (438,683)	-0.48 (0.35)
CAA Ratio*Post Policy			-627,728 *** (250,693)	-0.08 (0.82)
Time Trend	-42,754** (19,605)	-0.14*** (0.04)	-47,354*** (20,260)	-0.10*** (0.03)
Constant	868,466*** (81,133)	11.46*** (0.15)	1,078,985*** (81,960)	12.16*** (0.13)
Overall Adj. R-Squared	0.01	0.03	0.01	0.07
Within R-Squared	0.22	0.14	0.23	0.16
Between R-Squared	0.23	0.00	0.25	0.04

Regression discontinuity

Our results indicate that the Tox-Minus program 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.

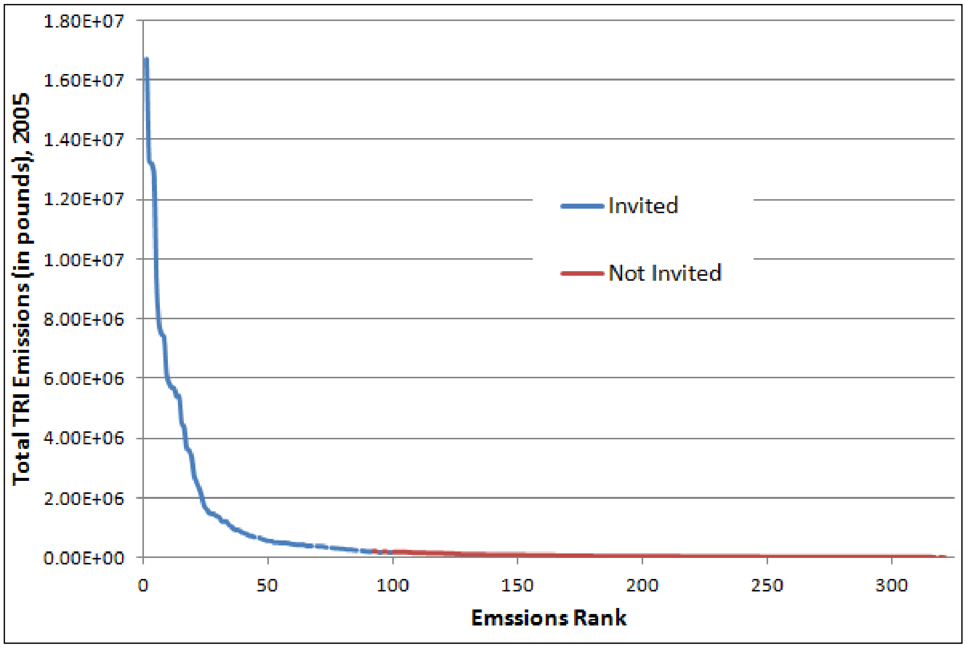
The regression discontinuity approach relies on the fact that the cutoff point is an arbitrary point and external to the facilities' decisions. It is likely that those just above the cutoff and those below it are similar with regard to pre-treatment characteristics. Only the ad hoc nature of the cutoff 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 ToxMinus. When the cutoff point is strictly determined – that is, the entire treated group lies on one side of a cutoff point and the entire untreated lie on the other – we have a sharp discontinuity design. If, instead, 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. Although the program claims to have invited the 100 of the top emitters, based on 2005 total emissions, we cannot exactly replicate the invited list of facilities using TRI emission. There appears to have been some other criteria used when deciding to invite facilities other than the strictly stated criteria reported by the Ohio EPA. Therefore, we have a fuzzy regression discontinuity design.

However, 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. However, the further from the cutoff point the more likely that assignment to one side of the cutoff or the other is not random. ToxMinus facilities received invitations mainly based on the magnitude of their 2005 TRI emissions: if they were in the top 100 they received a letter. To assess whether we have a sufficient number of ToxMinus facilities with emissions close to the cutoff, we plot the total TRI emissions in 2005 by rank in Figure 2.

The figure shows a cutoff between invited (in blue) and not invited facilities (in red) near the 100th rank, but it also shows two other things. First, the cutoff is not sharp. Even without limiting the sample to the points near the 100th rank, a blending between invited and non-invited facilities can be seen around the cutoff. Second, total emissions across the sample are very non-linear. This makes finding an acceptable sample for a regression discontinuity analysis difficult. It is very likely that the pre-treatment characteristics of facilities are not random the greater the difference in total emissions.

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 one end and twice these emissions on the 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 a robust regression discontinuity analysis comparing invited to non-invited facilities is not feasible.

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



IX. **Conclusion**

Voluntary programs such as Ohio's Tox-Minus program attempt to achieve the laudable goal of encouraging environmental improvement without imposing formal regulation, and the raw data might suggests 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 315,500 pounds in 2011. Average emissions of the 274 non-participants in our sample dropped from around 264,000 pounds to 148,000 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 more subtle interpretation.

Expanding the naïve approach to control for the post-policy emissions of those invited to the program and running a panel fixed-effects model with this cofactor suggests that it is the invitation to participate that may be motivating the decline in emissions and not participation in the Tox-Minus program itself. In other words, it may be that attention by the Ohio EPA and the fact that the facility was explicitly recognized as one of the top 100 emitters were the motivating factors for reducing emissions. Running these same models using the natural logarithm of air emissions as the dependent variable does not produce the same results for those invited to the program, but, in the panel model, indicates an effect for program participation.

One problem with these conflicting results is that it may not be appropriate to compare the 48 Tox-Minus participants with all 274 non-participants in our sample. To address this concern, we undertake a propensity score matching exercise, matching each treated facility with its most similar non-treated facility based on a probit regression of pre-treatment characteristics. This results in a sample of 96 facilities, 48 of which were Tox-Minus participants and 48 of which were non-participants. Summary statistics and a measure of bias suggest a very close match between these two groups. Running the cross-sectional and panel fixed effect difference-in-difference models on this matched sample produces similar results to the naïve model. In our preferred panel fixed-effects model, the invited group had a statistically significant decline in air emission levels in the post-policy period but not in logged emissions. In contrast, program participants showed a significant decline in logged emissions, but not in absolute levels.

Our interpretation of these results is that there may be two effects at play. First, it may be possible to claim that participation in the Tox-Minus program produced results. In our preferred model, participants show a significant reduction in logged emissions. It may be that because the majority of these facilities expressed their reduction goals in percentage terms, the program itself may have had a measurable effect on that percent. Second, it appears that simply being invited to the program produced an impact as well, regardless of actual participation. This second effect produces a significant decline in the level of emissions, but not necessarily in the percent reduction, as simply being invited does not require the facility to set an explicit emission goal.

Sensitivity analysis offers additional support for the existence of this second effect. Including the ratio of a facility's Clean Air Act emissions to its total emissions on the right hand side of the panel regression suggests a statistically significant decline in the level of air emissions as the ratio of Clean Air Act emissions rises. In addition, the invited group continues to show a significantly negative effect in the post-policy period. This suggests that both being invited to join the Tox-Minus program and having relatively more emissions subject to regulation, both a measure of regulatory attention, produced significant declines in the level of emissions.

References

Arora, S. and T. Cason (1996). Why Do Firms Volunteer to Exceed Environmental Regulations? Understanding Participation in EPA's 33/50 Program. *Land Economics* 72, 413-432.

Arora, S. and S. Gangopadhyay (1995). Toward a Theoretical Model of Voluntary Overcompliance. *Journal of Economic Behavior and Organization* 28(3), 289-309.

Bi, X., and M. Khanna (2012). "Reassessment of the Impact of the EPA's Voluntary 33/50 Program on Toxic Releases." Land Economics 88: 341-361.

Billor, N., A. Hadi and_P. Velleman (2000). BACON: Blocked Adaptive Computationally Efficient Outlier Nominators. Computational Statistics & Data Analysis 34: 279-298.

Blackman, A. and J. Boyd (2002). The Economics of Tailored Regulation: Will Voluntary Site-Specific Performance Standards Necessarily Improve Welfare? *Southern Economic Journal* 169(2), 309-326.

Blackman, Darley, Lyon, and Wernstedt (2010a). "What Drives Participation in State Voluntary Cleanup Programs? Evidence form Oregon." *Land Economics* 86(4): 785-799.

Blackman, A., B. Lahiri, W. Pizer, M. Rivera Planter, C. Muñoz Piña (2010b). Voluntary environmental regulation in developing countries: Mexico's Clean Industry Program. Journal *of Environmental Economics and Management* 60: 182-192.

Bennear, L., and S. 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, L., C. Griffiths, and A. 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, L., and S. 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.

Davies, J., J. Mazurek, N. Darnall, and K. McCarthy (1996). *Industry Incentives for Environmental Improvement: Evaluation of U.S. Federal Initiatives*. Washington, DC: Global Environmental Management Initiative.

de Vries, F., A. Nentjes, and N. Odam (2012). Voluntary Environmental Agreements: Lessons on Effectiveness, Efficiency and Spillover Potential. *International Review of Environmental and Resource Economics* 6: 119-152.

Fowlie, Holland, and 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.

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.

Hamilton, J. (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, K. (1999). Talking With the Donkey: Cooperative Approaches to Environmental Protection. *Journal of Industrial Ecology* 2(3), 51-72.

Henriques, I. and P. Sadorsky (1996). The Determinants of an Environmentally Responsive Firm: An Empirical Approach. *Journal of Environmental Economics and Management* 30(3), 381-95.

Horowitz, M. (2004). Electricity Intensity in the Commercial Sector: Market and Public Program Effects. *Energy Journal* 25(2), 115-137.

Innes and 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, M. and. L. 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.

Khann, M., and A. Keller (2005). Free Riding in Voluntary Environmental Programs: The Case of the U.S. EPA WasteWise Program. *Policy Sciences* 38: 91-106.

Khanna, M., W. Quimio, and D. Bojilova (1998). Toxics Release Information: A Policy Tool for Environmental Protection. *Journal of Environmental Economics and Management* 36(3), 243-266.

King, A. and M. Lenox (2000). Industry Self-Regulation Without Sanction: The Chemical Industries' Responsible Care Program. *Academy of Management Journal* 43(4), 698-716.

Konar, S. and M. Cohen (2001). Does the Market Value Environmental Performance? *Review of Economics and Statistics* 83(2), 281-89.

Lutz, S., T. Lyon, and J. Maxwell (2000). Quality Leadership When Regulatory Standards Are Forthcoming. *Journal of Industrial Economics* 48(3), 331-348.

Maxwell, J., T. Lyon, and S. Hackett (2000). Self-Regulation and Social Welfare: The Political Economy of Corporate Environmentalism. *Journal of Law and Economics* 43(2), 583-618.

Morgenstern, Pizer, and Shih (2007). Evaluating Voluntary U.S. Climate Programs: The Case of ClimateWise in *Reality Check: The Nature and Performance of Voluntary Environmental Programs in the United States, Europe and Japan*. Eds., R. Morgenstern and W. Pizer. Washington, DC: RFF Press.

Ohio EPA (n.d.). "Tox-Minus" Accessed on December 20, 2013. http://web.epa.ohio.gov/ocapp/tox-minus.html.

Ohio EPA (2007). "Alphabetical Listing of 100 of the Top Toxic Release Inventory Reporters for 2005" Accessed on December 23, 2013. http://web.epa.ohio.gov/ocapp/tox-minus/Tox-Minus%20Listing%20of%20100.pdf.

Ohio EPA (2008). " Listing of Ohio Facilities Participating in Ohio EPA's Tox-Minus Initiative" Data as of February 29, 2008. Accessed on December 23, 2013. http://web.epa.ohio.gov/ocapp/tox-minus/080229%20Tox%20Minus%20Updated%20Participant%20List.pdf.

Ohio EPA (2009). *Tox-Minus 2008 Activities Report*. March. http://web.epa.ohio.gov/ocapp/tox-minus/Tox-minus%20report%20for%20web.pdf.

Ohio EPA (2010). *Tox-Minus 2009 Activities Report*. June. http://www.epa.ohio.gov/Portals/41/tox-minus/100608%20Tox%20Minus%202009%20Activities%20Report.pdf.

Ohio Secretary of State (2012). “Election Results”. Accessed on November 8, 2012. http://www.sos.state.oh.us/elections/Research/electResultsMain.aspx.

Rivera, J., P. de Leon, and C. Koerber (2006). Is Greener Whiter Yet? The Sutainable Slopes Program After Five Years. The Policy Studies Journal 34: 195- 221.

Rosenbaum, P. and D. Rubin (1985). “Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score” The American Statistician. 39(1) (Feb): 33-38.

Segerson, K. and T. Micelli (1998). Voluntary Environmental Agreements: Good or Bad News for Environmental Protection? *Journal of Environmental Economics and Management* 36(2), 109-130.

US EPA (2012). " Data Export Downloads | Green Book | US EPA" Last modified December 14, 2012. Accessed March 21, 2013. http://www.epa.gov/oaqps001/greenbk/data_download.html.

US EPA (2013a). " Enforcement and Compliance History Online | US EPA" Accessed December 23, 2013. http://echo.epa.gov/.

US EPA (2013b). "List of Programs | Partnership Programs | US EPA" Last modified March 8, 2013. Accessed December 20, 2013. http://www.epa.gov/partners/programs/index.htm.

US EPA (2013c). " Toxics Release Inventory (TRI) Program | US EPA" Last modified December 16, 2013. Accessed December 23, 2013. http://www2.epa.gov/toxics-release-inventory-tri-program.

Vidovic and 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 and 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.

Welch, E., A. Mazur, and S. 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.