NCEE Working Paper

Evidence of How Electronic Reporting and Automated Auditing Affects Regulatory Compliance and Environmental Performance

Wayne Gray, Ron Shadbegian, and Ann Wolverton

Working Paper 26-04
July 2026

U.S. Environmental Protection Agency
National Center for Environmental Economics
https://www.epa.gov/environmental-economics

NCEE
NATIONAL CENTER FOR ENVIRONMENTAL ECONOMICS

Evidence of How Electronic Reporting and Automated Auditing Affects Regulatory Compliance and Environmental Performance

Wayne Gray (Clark University), Ron Shadbegian (Appalachian State University) and Ann Wolverton (U.S. EPA)

ABSTRACT

We investigate whether, by providing better and more timely information, the first mandatory online reporting and auditing program for wastewater discharge releases in the U.S. improved facility compliance and environmental performance, while also allowing state regulators to more effectively monitor and enforce regulations. Examining reporting programs with automated feedback features can also offer insights into the potential for artificial intelligence-based tools to improve compliance and environmental performance. Difference-in-difference results suggest that the program resulted in significant increases in the completeness of reporting and reductions in discharges but also greater reported violations. We find that effects are larger for minor dischargers and publicly owned facilities. We also find evidence consistent with the more efficient targeting of inspections by state authorities towards plants with a history of recent noncompliance, which could be a potential mechanism driving these results.

Key Words: auditing, electronic reporting, environmental compliance, wastewater discharges

JEL Codes: Q53, Q58

It has become increasingly common at both the state and federal levels to move to online reporting, and at times to include automated auditing of environmental compliance information. The purported benefits of this type of reporting program include time and cost savings, increased accuracy and transparency of information, and improved compliance. However, we are unaware of any formal evaluations of the extent to which a move to an electronic reporting and auditing program has delivered these benefits, particularly in an environmental context. Better understanding of whether such programs deliver tangible compliance or environmental benefits can also shed light on the potential for artificial intelligence (AI) tools that automate reporting requirements to deliver similar types of benefits. The potential for AI tools to reduce compliance burden and improve environmental performance has received increasing attention. For instance, in response to Executive Order 14179, Removing Barriers to American Leadership in Artificial Intelligence, the U.S. Environmental Protection Agency has begun piloting AI for recordkeeping and enforcement targeting (U.S. EPA, 2025).[1]

After substantial outreach and training, Ohio moved to an electronic reporting and auditing system at the end of 2007 for the reporting of wastewater discharges. At the time, it was the only state with a mandatory program of this kind in the United States. The online reporting and auditing system guided reporting plants in Ohio how to enter wastewater discharge information via a step-by-step “easy-to-use data entry wizard.” After submission, an automated auditing tool reviewed the data and sent an email to the reporting plant within 24 hours if its discharges were above allowable limits to alert the plant that it was either in violation of its permit or had committed a potential data entry error. The plant was then given a period of time in which to correct any data entry errors. The emailed information also instructed the plant on what to do if there was an actual permit violation.

This online reporting and auditing tool was expected to deliver improved facility compliance for two main reasons. First, it was expected to improve data quality through reduced manual error by allowing regulators to cross-check data quickly with permitted discharge levels to identify potential errors or compliance problems and give timely feedback to reporting plants to rectify the problem. Second, if reporting plants perceived an increase in the overall ability of state environmental authorities to effectively monitor and enforce regulations, either directly through the auditing mechanism or indirectly through improved targeting of traditional compliance tools (e.g., inspections), they may have put additional effort into improving compliance. For example, reporting plants may have taken actions to reduce the probability that they are out of compliance. Even facilities already in compliance or less likely to be audited may have responded to a perceived increase in monitoring or enforcement capabilities. Empirical research on the effect of monitoring and enforcement activity suggests that it has historically resulted in

[1] AI-based predictive enforcement tools have also been used by the Securities and Exchange Commission, Centers for Medicare and Medicaid Services, and the Internal Revenue Service (Engstrom, et al. 2020).

substantial emissions reductions (see Gray and Shimshack (2011) and Shimshack (2014) for overviews of this literature).

Our paper contributes to the existing literature on enforcement activity and environmental performance in the following ways. First, to the best of our knowledge, we are the first to examine empirically the degree to which an online reporting and auditing tool affects environmental compliance and performance. Second, we explore how improvements in compliance and performance vary by facility type, with differential effects between major and minor dischargers and publicly owned and other facilities. Minor dischargers, in particular, appear to achieve proportionately greater improvements in environmental compliance and performance from the switch to the online reporting and auditing tool. Third, we find suggestive evidence that the efficiencies gained by moving to an online tool with automated auditing features allow regulators to adjust their inspection and enforcement mechanisms to better target non-compliers.

In this paper, we compare wastewater discharging facilities in Ohio with similar facilities in the nearby state of Indiana, which did not have a mandatory online reporting and auditing program for wastewater discharges for the period 2005-2012. We use difference-in-difference estimation techniques to estimate the combined direct and indirect effects of the online reporting and auditing program on Ohio facility compliance behavior. We find that the online reporting and auditing program increased the filing of complete monthly discharge reports but also resulted in more reported violations by Ohio facilities compared to untreated facilities in Indiana. The online reporting and auditing program is also associated with improved environmental performance (i.e., lower discharge-limit ratios for total suspended solids (TSS) and biological oxygen demand (BOD)) by Ohio facilities in the post-policy period, though these results are not always statistically significant. Using a coarsened exact matching algorithm to balance the distributions of covariates between the treatment and control group, we find consistent, statistically significant evidence that the online reporting and auditing tool improves environmental performance. Finally, utilizing a Harrington-style framework, we find evidence that suggests higher reported violations may be driven, in part, by more efficient targeting of inspections and other enforcement-related activities on the part of Ohio state authorities.

This paper is organized as follows. Section II discusses the literature on the compliance benefits of enhanced auditing mechanisms. The more limited literatures on online reporting with or without an auditing mechanism and on how machine-learning tools can improve the targeting of inspections are also briefly summarized. Section III discusses the online reporting and auditing tool used by Ohio facilities to report wastewater discharges. Section IV describes our methodology for empirically investigating whether and how the online reporting and auditing of wastewater discharges affects compliance behavior and environmental performance. A discussion of how the comparison states were selected is in section V. The data used for this research are discussed in section VI. Section VII presents summary statistics. Results are presented in section VIII. Finally, section IX concludes.

1. Existing Literature on Potential Compliance Benefits of Auditing Mechanisms

While there is a broad literature on the effects of reporting requirements generally, the published literature examining the benefits of moving to an online reporting and auditing system is more limited. To date, we have found published literature that describes the time savings and reduced errors from moving to online reporting and auditing of environmental information and a few studies that examine whether auditing mechanisms – not necessarily through an online reporting tool – generally improve environmental compliance. In addition, a small but growing literature examines how machine-learning tools can be used to improve the targeting of inspections, which is potentially relevant when considering how access to better and more timely information could improve regulators' ability to leverage existing compliance and enforcement tools, given limited resources. We briefly summarize these literatures below.

1.1 Online reporting and auditing of environmental information

Evidence of the benefits of online reporting and auditing mechanisms in the environmental context is mostly qualitative in nature, but it is still instructive. These papers point to increased efficiency in processing and reviewing emissions data – an audit-like function - so that information is made available more quickly for facilities regarding compliance and the public regarding potential environmental hazards. Whether this translates into measurable improvements in compliance, and thus environmental outcomes, is not well-studied.[2] [3]

[2] Brennear and Olmstead (2008) use a difference-in-difference approach to examine the impact of information disclosure through annual consumer confidence reports (CCRs) on drinking water violations. No auditing mechanism is included but they examine the role that information conveyed via paper or electronic mechanisms plays. The authors examine the effect of direct mailing relative to lesser requirements on utilities serving small populations. They find that medium and large utilities reduced total violations by 30-44 percent and more severe health violations by 40-57 percent as a result of direct mailing and, at times, electronic posting of data.

[3] According to the Internal Revenue Service (IRS), online reporting simplified its ability to cross-reference data against other sources, allowing it to catch and correct errors more efficiently. The IRS noted that the error rate for electronically filed returns was less than 1 percent, compared to an error rate for paper returns of about 20 percent (e.g., IRS 2010). Filing tax returns online also expedited processing of tax payments and refunds. One paper studied the implications of online filing for the earned income tax credit (EITC), which was substantially under-utilized by qualifying households in the early 2000s. Making use of differences across states, they found that online filing had a significant and positive effect on EITC claims (Kopczuk and Pop-Eleches 2007). Evidence of the potential benefits from medical trials of the use of technology that incorporates an auditing mechanism is suggestive of benefits but inconclusive. While the use of electronic diaries by participants suffering from chronic pain to record their symptoms resulted in substantially improved compliance with protocols when paired with reminders to record information in a timely manner and feedback on compliance (Stone et al. 2003), other studies did not replicate similar improvements in compliance (Green et al. 2006; Blondin et al. 2010).

A 2006 Environmental Council of the States funded project reviewed the quality and efficiency of five technologies used by the U.S. EPA to facilitate the online exchange of information for states, Tribes and local agencies.[4] The project found that automating the exchange of data has allowed states to dramatically improve the quality, timeliness, and availability of environmental data (Environmental Council of States 2006). The U.S. EPA noted that use of online reporting for Toxic Release Inventory (TRI) emissions data has reduced errors relative to what was observed for manual entry and facilitated easier detection of remaining errors via automated audits instead of relying on more costly methods including site visits (U.S. EPA 1997).[5] TRI data were made available to the public more quickly due to decreased time required to compile, verify, and analyze data (Karkkainen 2001). Online reporting of SO_2 allowance and emissions data in the U.S. cap-and-trade system for electric utilities was credited with improving processing and verification. It allowed the U.S. EPA to reduce the time to reconcile actual emissions with the quantity of allowances held from up to 5 days to 24 hours for the vast majority of transactions. Likewise, the emissions tracking system enabled utilities to submit quarterly reports electronically, which allowed the U.S. EPA to perform automated compliance checks, provide feedback, and resolve any compliance issues quickly (Perez Henriquez 2004).[6]

1.2 Environmental compliance via self-auditing mechanisms

A few papers explored whether self-auditing mechanisms improve environmental compliance. In these cases, auditing was voluntary and initiated by the facility. It was not usually combined with an online mechanism to facilitate quick feedback between the reporting plant and the regulator. Pfaff and Sanchirico (2000) investigated the incentives for improved compliance under various self-auditing mechanisms. Audits allowed a facility to gather information on its compliance with environmental regulations and provided an opportunity to correct violations more quickly than if the audit had not occurred. Using a theoretical model, the authors demonstrated that facilities may not have had incentives to perform an audit unless penalties for non-compliance were conditioned on the level of effort put into the audit.[7]

[4] These are Air Quality System (AQS); Resource Conservation and Recovery Act (RCRA); Safe Drinking Water Information System (SDWIS); Toxics Release Inventory (TRI); and Electronic Discharge Monitoring Report (eDMR).
[5] EPA's Information Streamlining Plan projected dramatic reductions in the costs of reporting, storing, and analyzing conventional forms of information as a result of a transition to electronic reporting.
[6] Using a randomized field experiment, Benami, et al. (2023) find that sending customized email (though not automated) reminders to facilities holding National Pollutant Discharge Elimination System (NPDES) permits under the Clean Water Act significantly improved on-time reporting of environmental information via EPA's electronic reporting system.
[7] See Mishra et al. (1997) and Friesen (2006) for other theoretic considerations of self-auditing policies.

Two papers investigated whether self-auditing policies that granted facilities a reprieve from inspections or penalties, when a violation was discovered and subsequently corrected, resulted in improved future compliance. Building on prior work by Stafford (2005, 2007), Evans et al. (2011) examined the role that environmental self-audits by manufacturing facilities in Michigan played for Resource Conservation and Recovery Act (RCRA) compliance. They found that larger facilities and those subject to more stringent regulations were more likely to perform an audit, while facilities that had a history of non-compliance were less likely to perform an audit. However, audits appeared to have no significant effect on the likelihood that a facility was inspected or on its RCRA compliance record in the long run. Khanna and Widyawati (2011) examined voluntary self-audits to evaluate compliance with the Clean Air Act for a sample of S&P 500 firms. Consistent with Evans et al., they found that larger facilities and those facing greater liability under Superfund were more likely to perform an audit. Facilities with a history of prior inspections and those with better compliance records were also more likely to perform an environmental audit. Similarly, Khanna and Widyawati found that facilities in states that grant immunity from penalties if a violation is discovered during a self-audit (such as Michigan) did not have statistically different compliance behavior. Facilities in states with audit privilege policies (where a firm can choose not to disclose the result of an audit as long as a violation is corrected) had statistically lower compliance rates in the short run, the opposite of what was intended. Both papers used a two-stage empirical framework to control for endogeneity between the decision to self-audit and compliance status.

1.3 Leveraging machine-learning and other automated tools to improve compliance and enforcement targeting

A few papers have explored how machine learning and other automated tools can lead to improved targeting of compliance and enforcement tools. Hino, et al. (2018) found that leveraging a machine learning model to predict facilities at greatest risk of failing inspection (based on industry, location, and inspection history) increased the number of detected Clean Water Act single-event violations sevenfold. When they imposed constraints on the algorithm to account for budget and other limitations on the way inspections can be conducted by states, they still found a doubling of detected violations over the status quo approach. Similarly, results from a national field test of a machine-learning model that targeted inspections at facilities more likely to violate hazardous waste regulations increased an inspection's "hit rate" by almost 80 percent (Buchsbaum, et al 2024). Finally, Johnson, et al. (2023) found that targeting OSHA inspections to locations likely to experience the highest number of serious injuries based on accident history or predicted injuries from a machine-learning algorithm could have avoided up to twice as many serious injuries compared to when inspections were randomly assigned.

2. State Online Reporting and Auditing of DMR Data

While the Clean Water Act delegates authority to the states to regulate surface water discharges, it requires that entities discharging wastewater hold a National Pollution Discharge Elimination System (NPDES) permit and monitor and report their discharges to the states. Prior to 1999, monitoring data were collected by states only via paper forms. Ohio was the first to implement electronic reporting in 1999 - a form could be filled out electronically and then emailed to the state agency - followed a short time later by Florida and Michigan.[8] Both Ohio and Michigan were also early adopters of improved online reporting systems that incorporated auditing mechanisms to allow state regulators to quickly provide feedback to regulated entities as well as transfer Discharge Monitoring Report (DMR) data to the U.S. EPA, though only Ohio's system was mandatory over the period of study.

The voluntary adoption of online reporting of DMR data by states suggests the existence of potential net benefits. In particular, cost savings for governments in administering the system and for regulated entities in submitting data were noted. Anecdotal evidence points to possible compliance benefits associated with moving to such a system. For instance, Michigan, Florida, and Ohio all noted improved data accuracy through the elimination of manual data entry errors and improved overall program effectiveness due to an enhanced ability to more quickly analyze data, assess compliance, and act on this information.[9] A case study in Ohio noted a marked improvement in the accuracy of its compliance data after implementation of the online reporting system, with the number of errors falling from approximately 50,000 to 5,000 per month (U.S. EPA 2011).

Prior to the implementation of its current system, Ohio collected 70 percent of its wastewater discharge information via fillable forms available through a downloadable software program called SwimWare. While SwimWare required entering information via computer, it bore little resemblance to modern online reporting systems and was fairly labor intensive as it required a high degree of technical knowledge to run (U.S. EPA 2011).[10] Forms filled out using SwimWare were emailed to the state agency, which meant the system did not allow for easy updating to correct data errors or to reflect changes in permit conditions. In addition, a common complaint

[8] Twenty-three states reportedly allowed online reporting of DMR data as of 2011.Source: email exchange with the U.S. EPA's Office of Enforcement and Compliance Assurance in early 2012.

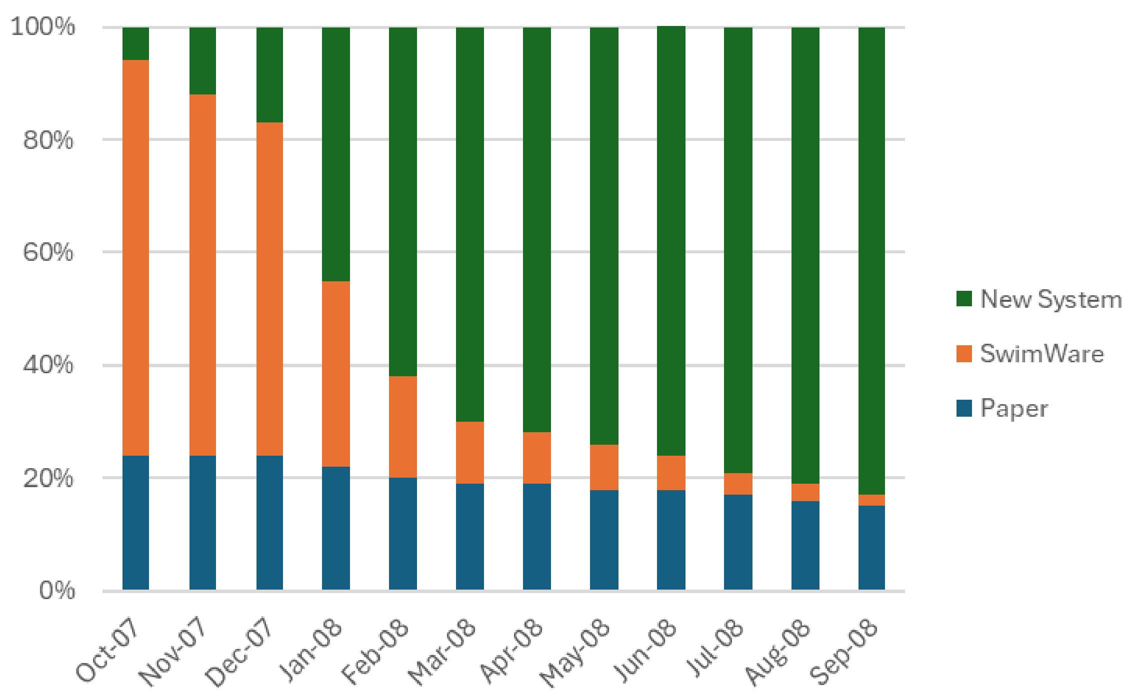
[9] See posted questions and answers on online reporting for wastewater facilities for these two states. For Michigan, see https://secure1.state.mi.us/e2rs/config/helpdoc/e2rs_faq.pdf, and for Florida, see http://www.dep.state.fl.us/water/wastewater/wce/edmrqa.htm. Accessed 03/07/2011.

[10] SwimWare required regulated entities to build NPDES permit requirements into the software, which proved challenging for many, and emailed to the state regulator in a zipped file (Enfotech News 2008).

for the remaining 30 percent of DMR data still submitted via paper reporting was that data were illegible or incomplete.

After doing substantial outreach and training, Ohio introduced its new online electronic reporting and auditing system for monthly and semi-annual DMR reports in late 2007. An "easy-to-use data entry wizard" guided regulated entities through the software and was paired with an automated compliance tool that sent an email to permit holders, within 24 hours, if they reported discharges above allowable limits. If the regulated entities made a data entry error, they could quickly correct it; otherwise, the emailed information instructed the regulated entity on what to do if there was an actual permit violation (Enfotech News 2008).

Figure 1: Percent of Ohio Facilities Submitting by Paper, SwimWare, or New System



Source: Enfotech News, 2008.

Figure 1 presents information on Ohio's phase-out of SwimWare and introduction of the new online reporting auditing system between October 2007 and September 2008 (paper reporting continued to be offered as an option during the transition).[11] The Ohio EPA rolled out the new online reporting and auditing system by introducing it to a new district each month over five months beginning in November 2007(U.S. EPA 2011). While SwimWare was still the dominant

[11] The initial number of entities reporting discharges in Ohio in October 2007 was 3,214. The number of entities reporting peaked at 3,308 in January 2008 and hit its lowest point of 3,146 in September 2008 (though the previous month had 3,237 entities reporting).

form of reporting at the end of 2007, it was only used by 11 percent of regulated entities by March 2008. By September 2008, only 2 percent of reporting facilities still used SwimWare and 83 percent of facilities were using the new online reporting and auditing system. By 2011, all permit holders reportedly used the new online system (EPA 2016).

3. Empirical Approach

To discern whether the new online reporting and auditing program improved environmental compliance and performance for Ohio wastewater dischargers, we take advantage of the fact that a nearby state did not adopt similar mandatory requirements over the period of study. Specifically, we examine whether there was a significant improvement in DMR submission rates, a decline in the number of reported violations, and a decline in the degree of non-compliance as measured by a facility's discharge-limit ratio, between facilities in Ohio that were required to report via the new system after 2007 and facilities in the nearby state of Indiana where use of an online system was not required. While we cannot distinguish between differences in reported compliance that resulted from reduced reporting errors versus those that were due to genuine compliance improvements, both are manifestations of an improved compliance oversight system. Unlike Evans et al. (2011) and Khanna and Widyawati (2011), we do not need to control for endogeneity between the audit and compliance decision since auditing is an automatic feature of the new online reporting system introduced in Ohio.

Our main empirical strategy uses a difference-in-difference estimation method to examine compliance behavior before and after the online reporting and auditing system was introduced in Ohio as a function of whether a facility is in the treatment state where online reporting is mandatory or the control state without any online reporting or auditing. While Ohio already had introduced an electronic system in 1999, descriptions of the older (i.e., SwimWare) and newer systems suggest that they are fundamentally different and that the main mechanisms through which we would expect the new system to affect compliance behavior - namely, the ability to audit and provide real time feedback - were not present in the old system.

One appeal of the difference-in-difference approach is that it differences out facility heterogeneity that is constant over time, including unobserved heterogeneity, and therefore does not have to be accounted for in the regression. This feature greatly reduces the number of explanatory variables we need to include in the analysis. For instance, we do not need to include factors such as the year a facility was established, its industry, and its size (if it has not changed substantially over time) as alternate explanations for changes in compliance.

A potential limitation of this approach is that it assumes that the same model applies to plants in Ohio and the nearby state for time-varying control variables. If conditions in Ohio and the nearby state vary too widely, that can bias our results. To at least partially address this concern, we use a coarsened exact matching algorithm (Iacus et al. 2012) to match treated Ohio facilities with similar facilities in the nearby state. CEM reduces the likelihood of substantial omitted variable bias by controlling non-parametrically for observed pre-treatment differences between facilities in Ohio and Indiana, while potentially making them more comparable in terms of unobserved characteristics (Iacus et al., 2012). We specifically match Ohio facilities to comparable Indiana facilities based on four indicator variables: whether it is publicly owned (these are mostly publicly owned treatment facilities but also includes other state, county, and municipal facilities); its major or minor status;[12] whether its initial NPDES permit is less than twenty years old; and whether it had three or fewer inspections over 2003-2005.

Another potential challenge with the difference-in-difference approach is that it greatly reduces the amount of variation in the explanatory variables that can then be used to explain variation in compliance behavior across facilities in these two states. In general, anything that removes large amounts of variation in the explanatory variables has the potential to magnify measurement error. However, since the explanatory variable of interest – being covered by the online reporting and auditing program – depends only on the state and date of the observation, it seems unlikely that measurement error is a significant issue.

We begin with a simple model in equation (1) where compliance in year t by facility i is denoted $compliance_{it}$. The variable α captures the average compliance of plants in the control state. *Ohio* is a dummy variable that is set to 1 when a facility is in the treatment group (i.e., located in Ohio), which captures any pre-policy differences between facilities in Ohio and those in the control group. *Post Policy* is a time dummy variable that is set to 1 in the post-policy period (2008-2012), which captures any general factors that result in changes in facility compliance behavior over time in all states apart from the online reporting and auditing program in Ohio. When we interact these two variables, the dummy variable *Ohio*Post Policy* is equal to 1 when a facility is in the treatment group (i.e., required to submit its DMR report through Ohio's online system) in the second period. Finally, e_{it} is defined as the residual error term.

$$compliance_{it} = \alpha + \beta_1 \, Ohio + \beta_2 \, Post\ Policy + \beta_3 \, Ohio*Post\ Policy + e_{it} \qquad (1)$$

A second regression – in equation 2 - takes advantage of the panel nature of our data set by adding a facility-specific fixed effect, a_i. Following Bennear and Olmstead (2008), we include a

[12] POTW dischargers are categorized as major if they have a design flow of one million gallons per day or greater or that serve a population of 10,000 or more or that cause significant water quality impacts. Non-POTW dischargers are categorized as major based on a NPDES worksheet based on the number of points they accumulate. For more information, see https://www.epa.gov/system/files/documents/2025-09/pwm_2010_edits_2025_06.pdf .

quadratic time trend, t and t^2.[13] The inclusion of facility fixed effects means we can no longer independently identify the coefficient on *Ohio*, so it drops out of the specification:

$$compliance_{it} = \beta_2 Post\ Policy + \beta_3\ Ohio{*}Post\ Policy + a_i + t + t^2 + e_{it} \qquad (2)$$

4. Data and Selection of Comparison States

We extract monthly data on DMR reporting rates and compliance status between 2005 and 2012 from EPA's Integrated Compliance Information System (ICIS). While these data are self-reported, we think it unlikely that facilities in Ohio (or in the comparison state) deliberately misreported discharges prior to the introduction of the online reporting and auditing program. Shimshack and Ward (2008) note that the Clean Water Act allows for criminal action to be taken directly against employees that deliberately misreport facility discharge information, including the possibility of jail time. We chose 2012 as the end year for our analysis due to the growing national focus on adopting an electronic system to report wastewater discharges. The U.S. EPA proposed electronic reporting of DMRs in 2013 and promulgated the regulation in 2015. Beginning in December 2016, the U.S. EPA required that most facilities submit DMRs electronically.

We compare Ohio to a nearby state in which use of an online reporting and auditing tool is not mandated for wastewater discharges during the study period. The comparison state is chosen from adjacent states where online reporting was either not available or only available on a voluntary basis. We limit our choice of a comparison state to neighboring states to help control for exogenous differences across states that might be correlated with reporting of wastewater discharges. Ohio is bordered by Michigan, Indiana, Kentucky, West Virginia, and Pennsylvania.

We rule out states that are vastly different in terms of the overall level of economic activity or the regulated universe (e.g., West Virginia and Kentucky). Michigan is eliminated from consideration as a comparison state because, while its system was voluntary over the study period, the version it put into place in 2010 had many of the same auditing features as the Ohio system. This leaves Indiana and Pennsylvania as potential comparison states. Both states are slightly smaller than Ohio in the pre-policy period in terms of overall economic activity, but both have a similar mix of manufacturing sectors (e.g., motor vehicles, petroleum refining, iron and steel, plastics). Ohio and Pennsylvania are similar in terms of land area and population size, while Indiana is somewhat smaller in size. However, Indiana is in the same EPA region as Ohio and therefore likely shares the same general approach and priorities for Federal enforcement. Moreover, Indiana required wastewater discharge reports to be filed on a monthly basis for both major and minor dischargers, similar to Ohio during the period of study, whereas Pennsylvania

[13] There are very few qualitative differences between our results in specifications with year fixed effects and a quadratic time trend. All results are available upon request.

only required major dischargers to file monthly discharge reports, which would limit the sample size dramatically.[14] Therefore, we use facilities in Indiana as our control group in the main specification.

Facilities listed as inactive or that do not submit monthly discharge monitoring reports (DMRs) in 2005 – 2012 are dropped from the data set to ensure that we do not artificially inflate the counts of non-compliance by including facilities that are not required to report. There are a few instances where a facility is listed as inactive but continues to submit monthly DMRs. We have included these facilities in our data set. We also limit the universe of possible comparison facilities to those with data reported in the first and last years of the data set, 2005 and 2012, to ensure changes in discharges are not due to plants that open after 2007 but before the end of our study period or that permanently close prior to 2012.

5. Compliance and Other Variables

Our analysis focuses on several measures of the degree of compliance with federal Clean Water Act requirements. *Percent Violation* is defined as the percent of DMRs submitted by a facility each month that resulted in a reported violation. *Percent Reported* is defined as the percent of a facility's permits for which discharges were fully reported each month. *Violation Dummy* is defined as equal to 1 if a facility has at least one reported violation in a given month, and zero otherwise. Finally, *BOD Discharge-Limit Ratio* is defined as the ratio of average monthly wastewater discharges of biochemical oxygen demand (BOD) to the Federal BOD discharge limit for that facility. We define a comparable measure for total suspended solids, *TSS Discharge-Limit Ratio*. A value less than or equal to 1 demonstrates that a facility is under its limit ratio, on average, while a value greater than 1 indicates that a facility is above its limit ratio, on average, in a given month. In general, we assume that average monthly values less than 1 are highly correlated with compliance, while values greater than 1 highly correlate with lack of compliance with the Federal standard.

Many facilities had both quantity and concentration-based limits for TSS and BOD. We observe that the quantity-based discharge-limit ratio included very large outliers, facilities whose discharges were drastically higher than the limit. While the typical facility emitted well below its discharge limit for both TSS and BOD, the mean was substantially higher than 1. Taken at face value, the data indicates that the average facility discharged at 157 times over its BOD limit and 142 times over its TSS limit. Looking at the 99th percentile of the distribution, facilities were about four and six times over their limits, respectively, still well below the mean values. The largest values in the dataset for BOD and TSS, however, were extremely large and appear to be errors.

[14] Federal policy did not require monthly DMRs be submitted by minors during our period of study, though states were given the option of submitting this information to EPA when they collected it. For more information, see https://www.epa.gov/sites/default/files/2020-03/documents/owm0205.pdf .

For instance, at the tail of the distribution there is a facility with BOD discharges more than 488,000 times greater than its limit. Other researchers, including Earnhart (2004) and Horowitz (2006), noted this problem with quantity-based discharge limit information, and chose to rely only on the concentration-based limits and discharge information for their analyses. We follow past practice and rely on the concentration-based limit and discharge information. As with previous studies, we find that concentration information is available for most of the observations in our dataset subject to quantity-based discharge limits. While some outliers are still apparent in the concentration discharge-limit ratio data, they do not drive a large wedge between the mean and median and, as such, are potentially consistent with non-compliant behavior (e.g., they are 6 or 10 times the limit instead of hundreds of thousands of times the limit).

The independent variables are dummy variables representing the post-policy time period, *Post Policy*, facilities in Ohio, *Ohio*, and the interaction between them to represent facilities subject to the online reporting and auditing program in Ohio in the post-policy period, *Ohio*Post Policy*. We leverage information on the month and year when an individual facility first reported using the new system to allow *Post Policy* to vary by facility. To allow *Post Policy* to enter the panel regressions, we use a quadratic time trend instead of year fixed effects.

6. Summary Statistics

Table 1 presents the mean and standard deviation for key variables in Ohio and Indiana. First, note that the number of Ohio facilities in the data set that are required to submit DMRs is larger than those in Indiana. There are almost 1,600 Ohio facilities in our data set for 2005 and 2012, while just over 800 facilities in Indiana are required to submit DMRs in these years. Second, facilities in Ohio tend to have fewer reported violations (*Percent Violation*) but also submit a lower percent of complete discharge reports (*Percent Permit*) than facilities in Indiana both before and after the online reporting and auditing program was introduced for Ohio facilities at the end of 2007. Third, we do not observe any pre-treatment trends in these variables across Ohio and Indiana: The percent reported violations declined slightly from 2005 to 2008.[15] The percent of facilities' filed discharge reports that were complete were directionally different in Ohio and Indiana but stayed roughly constant in both states between 2005 and 2008 - they increased by about 4% in Ohio and declined by 2% in Indiana. In the post-policy period, reported violations continued to decline in both states. However, the percent of facilities' filed discharge reports continued to increase in Ohio between 2008 and 2012 and stay roughly constant in Indiana (declines by less than 1%).

[15] See also the event study analysis at the end of section 7.

Table 1: Summary Statistics for Facilities in Ohio and Indiana

	Ohio	Indiana
Dependent Variables		
Percent Violation, 2005	2.73 (7.8)	6.59 (17.8)
Percent Violation, 2008	2.05 (5.9)	5.92 (17.5)
Percent Violation, 2012	2.19 (9.1)	3.32 (13.1)
Percent Reported, 2005	65.98 (27.0)	88.98 (28.4)
Percent Reported, 2008	68.82 (26.9)	87.34 (28.2)
Percent Reported, 2012	74.86 (24.1)	85.29 (30.7)
Mean BOD Discharge to Limit Ratio, 2005	0.45 (0.83)	0.33 (1.73)
Mean BOD Discharge to Limit Ratio, 2008	0.47 (4.37)	0.29 (0.49)
Mean BOD Discharge to Limit Ratio, 2012	0.36 (0.63)	0.24 (0.41)
Mean TSS Discharge to Limit Ratio, 2005	0.59 (1.4)	0.42 (0.71)
Mean TSS Discharge to Limit Ratio, 2008	0.59 (8.32)	0.42 (1.19)
Mean TSS Discharge to Limit Ratio, 2012	0.41 (0.74)	0.37 (1.05)
Independent Variables		
Number of Major Facilities	203	137
Number of Publicly Owned Facilities (mostly POTWs)	742	402
Number of Facilities in Manufacturing	157	27
Total Number of Facilities in 2005, 2012	1,584	826

The summary statistics for the BOD and TSS mean discharge limit ratios show that facilities in Ohio and Indiana tend to be well under the allowable discharge limit, though Ohio facilities tend to have slightly larger discharge-limit ratios on average. We see a noticeable dip in the average TSS and BOD discharge-limit ratios for Ohio facilities in 2012 relative to the 2005 and 2008 levels, larger than the dip observed in Indiana. Finally, a similar proportion of DMRs filed in Ohio and Indiana are from major (13 percent versus 17 percent) or publicly owned facilities (47 percent versus 49 percent). However, manufacturing facilities (denoted by SIC codes 2000-3999) comprise a larger percent of DMRs in Ohio compared to Indiana (10 percent versus 3 percent) over our study period.

7. Results

We now turn to the results from a series of difference-in-difference specifications. Table 2 describes results for *Percent Violation* and *Violation Dummy*, measures of a facility being out of compliance with wastewater discharge regulations. Table 3 reports results for *Percent Reported*, the percent of a facility's permits for which discharges were fully reported. Tables 4 and 5 present results for *TSS Discharge-Limit Ratio* and *BOD Discharge-Limit Ratio*, respectively.

In Tables 2 and 3, columns 1 and 2 present pooled regression results using ordinary least squares for the full and CEM samples, respectively. The robust standard errors are clustered at the plant-level. Using the same dependent variable, columns 3 and 4 present results for the full and CEM sample panel fixed effect regressions, respectively. Columns 5 and 6 present results from a panel logit model when the dependent variable is converted from *Percent Violation* to an indicator dummy, *Violation Dummy*.[16] [17] Note that about 13,000 facility-by-month observations drop from the panel logit specifications due to lack of variation in compliance status over time – i.e., facilities with no violations in any year or violations in every year.

Table 2: Regression Results for Percent Violation and Violation Dummy

	Pooled		**Panel Fixed Effects**			
	Percent Violation		Percent Violation		Violation Dummy	
	(1)	(2)	(3)	(4)	(5)	(6)
Ohio*Post Policy	1.16*** (0.09)	1.02*** (0.14)	1.17*** (0.22)	1.01*** (0.33)	0.877 (0.545)	0.959* (0.558)
Ohio	-3.42*** (0.08)	-3.02*** (0.13)				
Post Policy	-0.423** (0.08)	-0.52*** (0.21)	-0.44* (0.23)	-0.54 (0.35)	-2.50** (0.70)	-3.12*** (0.72)
Quadratic time trend	N	N	Y	Y	Y	Y
Sample	**Full**	**CEM**	**Full**	**CEM**	**Full**	**CEM**
Obs	226,870	226,870	226,870	226,870	213,901	213,901

* denotes significance at the 10 percent level; ** denotes significance at the 5 percent level; *** denotes significance at the 1 percent level. Columns (5) and (6) report marginal effects (multiplied by 100) for the logit models.

Consistent with the summary statistics, *Percent Violation* is negatively related to *Post Policy*, reflecting the general decline in reported violations in Indiana after 2008. We find that *Percent Violation* is significant and negatively related to Ohio facilities before 2008, compared to those in Indiana. Our key policy variable, *Ohio*Post Policy*, is positively related to *Percent Violation*,

[16] Converting violations into percent terms addresses the challenge that data on violations is often dominated by zeros. Bennear and Olmstead (2008) addressed this issue by running panel Poisson and negative binomial models.
[17] To avoid estimating misleading average partial treatment effects in nonlinear models we evaluate the marginal effects only for Ohio facilities during the treatment period (N=89,322) - see Curran et al. (2026).

implying that compliance among facilities covered by Ohio's online reporting and auditing program improved by less than the improvement for facilities in Indiana. The coefficient magnitude suggests an increase of about 1 percent in reported violations, about one-third of the 2005 violation rate in Ohio. In the logit specifications, while the finding that reported violations have declined by less at Ohio facilities subject to the online reporting and auditing program in the post-policy period compared to facilities in Indiana continues to hold, it is no longer statistically significant.

Table 3: Regression Results for Percent Reported

	Pooled		Panel Fixed Effects	
	(1)	(2)	(3)	(4)
Ohio*Post Policy	7.41*** (0.21)	6.93*** (0.33)	6.95*** (0.44)	6.98*** (0.58)
Ohio	-20.92*** (0.002)	-15.91*** (0.003)		
Post Policy	-1.57*** (0.18)	-4.20*** (0.43)	-3.92*** (0.33)	-4.08 *** (0.45)
Quadratic time trend	N	N	Y	Y
Sample	Full	CEM	Full	CEM
Obs.	226,870	226,870	226,870	226,870

* denotes significance at the 10 percent level; ** denotes significance at the 5 percent level; *** denotes significance at the 1 percent level.

The model specifications for *Percent Reported,* the percent of a facility's required monthly DMRs that are completely reported, in Table 3 are presented in a similar order to those in Table 2, omitting the logit models since large numbers of zero values are not an issue in this case. The higher *Percent Reported,* the better a facility's compliance record. While *Post Policy* and *Ohio* are both negative and statistically significant in all specifications, Ohio facilities that reported via the online reporting and auditing tool in the post-policy period had a significantly higher increase in completed reports compared to facilities in Indiana, with a difference of 7 percent, about one-tenth of the initial rate in Ohio.

We also examine facility wastewater discharges relative to the allowable limit as a more direct indicator of the impact of the reporting and auditing program in Ohio on environmental outcomes. Table 4 presents the regression results for TSS, while Table 5 presents the results for BOD. Columns 1 and 2 present pooled regression results for the full and CEM samples, respectively, while columns 3 and 4 present the panel fixed effect results.

Table 4: Regression Results for TSS Discharge-Limit Ratio

	Pooled		Panel Fixed Effects	
	(1)	(2)	(3)	(4)
Ohio*Post Policy	-0.229*** (0.08)	-0.248*** (0.09)	-0.209** (0.09)	-0.232** (0.10)
Ohio	0.261*** (0.06)	0.304*** (0.07)		
Post Policy	0.039 (0.08)	0.135* (0.08)	0.116 (0.10)	0.135 (0.11)
Quadratic time trend	N	N	Y	Y
Sample	Full	CEM	Full	CEM
Obs.	171,795	171,795	171,795	171,795

* denotes significance at the 10 percent level; ** denotes significance at the 5 percent level; *** denotes significance at the 1 percent level.

The results in Table 4 indicate that *Ohio*Post Policy* has a significantly negative effect on the *TSS Discharge-Limit Ratio* in both the pooled and panel fixed effect regressions. In other words, the introduction of the online reporting and auditing program in Ohio is associated with declining discharge-limit ratios for TSS (i.e., reduced water pollution) relative to facilities in Indiana. The reduction of 0.23 is about one-third of the average value in Ohio in 2005. In the pooled regressions *Ohio* is positive and significant, indicating that Ohio facilities tend to have higher discharge-limit ratios than facilities in Indiana in the pre-treatment period. The dummy variable for *Post Policy* is also positive showing that discharges in Indiana increased after 2008 but are never statistically significant.

A similar set of regressions are presented in Table 5 for BOD discharge-limit ratios. While all the variables retain the same sign as in the TSS regressions in Table 4, many are less significant or no longer significant. *Ohio*Post Policy* is still negatively related to the discharge-limit ratio, but it is only nominally significant in the pooled regressions and has smaller coefficients than in the TSS results. The reduction of 0.075 is about one-sixth of the Ohio mean in 2005. In other words, the introduction of the online reporting and auditing program in Ohio is associated with lower discharge-limit ratios for BOD, but the effect is much weaker than for TSS. We also find that *Ohio* is positive and significant, indicating that Ohio facilities have higher discharge-limit ratios than those in Indiana before the policy. The *Post Policy* dummy is still always positive but remains statistically insignificant.

Table 5: Regression Results for BOD Discharge-Limit Ratio

	Pooled		Panel Fixed Effects	
	(1)	(2)	(3)	(4)
Ohio*Post Policy	-0.085** (0.04)	-0.095** (0.04)	-0.065 (0.04)	-0.075* (0.04)
Ohio	0.194*** (0.03)	0.213*** (0.04)		
Post Policy	0.076 (0.03)	0.078** (0.04)	0.067 (0.05)	0.069 (0.06)
Quadratic time trend	N	N	Y	Y
Sample	Full	CEM	Full	CEM
Obs.	142,846	142,846	142,846	142,846

* denotes significance at the 10 percent level; ** denotes significance at the 5 percent level; *** denotes significance at the 1 percent level.

Heterogeneity in Effects Across Ohio Facilities

We also estimate triple difference specifications to explore heterogeneity in effects by facility type. For the first triple-different specification, we define a dummy variable equal to one to indicate a minor facility, *Minor*. One hypothesis is that minor facilities that previously faced far less oversight will adjust compliance behavior relatively more in response to the auditing component of the new system, as it represents a large shift from previous practice. Major facilities generally already face greater scrutiny from regulators, even absent the online reporting and auditing program. However, with more reliable and readily available data through the online reporting system, an alternate hypothesis is that major facilities will further improve compliance compared to other facilities.

Table 6 presents the results for the minor facility triple difference for the panel fixed effect specifications using the CEM sample.[18] Minor facilities in the post policy period in Indiana (*Minor*Post Policy*) do not demonstrate statistically significant improvements in outcomes compared to major facilities, except in the case of *Percent Reported* where there is a small positive improvement in the complete filing of monthly reports. For *Percent Violation* the coefficient on *Ohio*Minor*Post Policy* is not statistically significant. However, it is positive and statistically significant for *Percent Reported*, indicating that the requirement to report via the electronic and auditing tool in Ohio resulted in an additional improvement in complete monthly reporting for minor facilities in the post policy period over and above the improvement for major

[18] Results for the pooled specifications based on the CEM sample are available upon request. In the pooled regressions, the coefficients on *Ohio*Post Policy* and *Minor*Ohio*Post Policy* are of similar signs and magnitudes as in the panel fixed effects regressions but are always statistically significant.

facilities, as identified by the positive and statistically significant coefficient on *Ohio*Post Policy*. Given the similar coefficient magnitudes, minor facilities had about twice the improvement of majors.

Table 6: Triple Difference Regression Results Accounting for Minor Facilities

	Panel Fixed Effects			
	Percent Violation (1)	Percent Reported (2)	TSS Discharge-Limit Ratio (3)	BOD Discharge-Limit Ratio (4)
Ohio*Post Policy	0.676*** (0.182)	4.38*** (1.14)	-0.033 (0.024)	-0.017 (0.015)
Ohio*Post Policy*Minor	0.387 (0.410)	2.98** (1.29)	-0.234* (0.122)	-0.070 (0.060)
Post Policy	-0.087 (0.279)	-6.27*** (1.05)	-0.052 (0.042)	0.024 (0.025)
Minor*Post Policy	-0.516 (0.447)	2.49** (1.20)	0.097 (0.090)	0.054 (0.046)
Quadratic time trend	Y	Y	Y	Y
Sample	CEM	CEM	CEM	CEM
Obs.	226,870	226,870	171,795	142,846

* denotes significance at the 10 percent level; ** denotes significance at the 5 percent level; *** denotes significance at the 1 percent level.

In the case of TSS and BOD discharges, it appears that the significantly lower TSS and BOD discharge-limit ratios after the introduction of the online reporting and auditing program in Ohio found in Tables 4 and 5 are mainly driven by improvements at minor facilities - the coefficients on *Minor*Ohio*Post Policy* are of similar magnitude and significance to those on *Ohio*Post Policy* in Tables 4 and 5, while the coefficients on *Ohio*Post Policy* in Table 6 are notably smaller and no longer statistically significant. Overall, it appears that the reporting requirement had a greater impact on minor facilities than on majors, though the differences are not always significant.

For the second triple difference specification, we define a dummy variable equal to one to indicate government-owned facilities, *Public*. Publicly owned facilities, including waste-water treatment plants managed by a municipality, operate under a different incentive structure than privately held facilities or those operated by private and government partnerships, though it is hard to predict how this may affect compliance after the online reporting and auditing program was introduced.[19]

[19] We also explore using industry dummies in several of the regressions, though these results are not reported in the paper. Including industry dummies had no effect on the qualitative nature of our results.

Table 7: Triple Difference Regression Results Accounting for Publicly Owned Facilities

	Panel Fixed Effects			
	Percent Violation (1)	Percent Reported (2)	TSS Discharge-Limit Ratio (3)	BOD Discharge-Limit Ratio (4)
Ohio*Post Policy	-0.199 (0581)	2.97* (1.70)	-0.107** (0.051)	-0.057 (0.070)
Ohio*Public*Post Policy	1.51** (0.695)	5.19*** (1.77)	-0.148 (0.136)	-0.019 (0.087)
Post Policy	0.850 (0.649)	-3.80** (1.70)	-0.001 (0.056)	0.060 (0.079)
Public*Post Policy	-1.74** (0.759)	-0.445 (1.83)	0.163* (0.096)	0.010 (0.087)
Quadratic time trend	Y	Y	Y	Y
Sample	CEM	CEM	CEM	CEM
Obs.	226,870	226,870	171,795	142,846

* denotes significance at the 10 percent level, ** denotes significance at the 5 percent level while *** denotes significance at the 1 percent level.

Table 7 presents the results for the public facility triple difference for the panel fixed effect specifications using the CEM sample. Results for the pooled specifications based on the CEM sample are available upon request.[20] Public facilities in the post policy period in Indiana (*Public*Post Policy*) do not demonstrate statistically significant differences in outcomes compared to non-public facilities, except in the case of *Percent Violations* where they have fewer reported violations in later years. For the outcome variables related to environmental outcomes, *TSS Discharge-Limit Ratio* and *BOD Discharge-Limit Ratio*, the coefficient on *Ohio*Public*Post Policy* is not statistically significant. The coefficient on *Percent Reported* is positive and significant, indicating that publicly owned facilities in Ohio see a relative improvement in complete monthly reporting in the post policy period beyond the improvement experienced for all Ohio facilities. Taken together, publicly owned facilities are responsible for more of the overall significantly larger increase in violations and in improved reporting, as well as larger reductions in TSS and BOD discharges, although the latter differences are not statistically significant. In the case of *Percent Violation*, the increase in reported violations after the introduction of the online reporting and auditing program in Ohio appears to be mainly driven by what is occurring at publicly owned facilities. Overall, the reporting requirement had a greater impact on publicly owned facilities, though these differences are not significant for the discharge-related measures.

[20] In the pooled regressions, the coefficients on *Ohio*Post Policy* are of similar sign and magnitude but are always statistically significant, while *Public*Ohio*Post Policy* is of similar sign, magnitudes and statistical significance as in the panel fixed effects regressions.

Robustness of Results

As a sensitivity check, we examine the robustness of our results to how we define the post-policy period. Some facilities did not begin using the online reporting and auditing tool until mid-2008. Even when facilities began to use it, they may not have immediately realized that they had a potential compliance issue due to unfamiliarity with the new system. If this is the case, we may not see behavioral change until 2009. When we redefine the post policy period as 2009-2012, Table 8 shows that the results for each key outcome are consistent with those presented in Tables 2-5 where 2008-2012 is defined as the post-policy period. The sign, magnitude and statistical significance of the coefficient on *Ohio*Post Policy* is consistent with previous results.

Table 8: Regression Results When Post-Policy Period Is Defined as 2009-2012

Panel Fixed Effects				
	Percent Violation (1)	Percent Reported (2)	TSS Discharge-Limit Ratio (3)	BOD Discharge-Limit Ratio (4)
Ohio*Post Policy	0.982*** (0.303)	5.87*** (0. 525)	-0.205*** (0.075)	-0.058* (0.031)
Post Policy	-0.567*** (0.268)	- 0.064 (0.352)	0.037 (0.034)	-0.0004 (0.01)
Quadratic Time Trend	Y	Y	Y	Y
Sample	CEM	CEM	CEM	CEM
Obs.	226,870	226,870	171,795	142,846

* denotes significance at the 10 percent level, ** denotes significance at the 5 percent level while *** denotes significance at the 1 percent level.

What about traditional compliance mechanisms?

A key question when interpreting the regression results is whether any other policy changes in Ohio coincided with the introduction of the online reporting and auditing program. One possibility is that Ohio also made substantial changes to traditional compliance mechanisms within the same time frame. If this is the case, we could be misattributing improvements in compliance to the online reporting and auditing program. It is worth noting, however, that these two mechanisms are not necessarily independent. If the online reporting and auditing program allows for better targeting of existing resources (e.g. inspecting facilities are expected to be in non-compliance more often), one could argue that any resulting improvement in compliance through traditional mechanisms is indirectly attributable to the new program.[21]

We first examine attributes of Ohio's inspection regime over the 2005 – 2012 timeframe (Table 9). While the number and types of inspections conducted in Ohio fluctuate somewhat year to year, we do not see evidence of a drastic shift in resources between the pre (2005 – 2007) and post policy (2008 - 2012) periods, which argues against the idea that enhanced inspection activity in Ohio is mistakenly being attributed to the online reporting and auditing program. The number of inspections conducted in Ohio hovered around 1,000 between 2006 and 2008 and again in 2010 and 2011. The number of unique facilities that were inspected increased until 2007 and then fell slightly from that point forward and did not increase again until 2010. While the number of major facilities inspected in 2009 and 2012 was slightly larger than in 2008, it was on par with pre-policy trends. Also, the number of inspections at major facilities remained roughly constant between 2007 and 2012 with a slight dip in 2008 and 2012.

Table 9: Summary Statistics for 2005 - 2012 Inspections in Ohio

Year	Total Number of Inspections	Number of Facilities Inspected	Number of Major Facilities Inspected	Number of Inspections at Majors
2005	638	464	117	240
2006	1,095	761	154	346
2007	1,005	792	143	242
2008	957	754	137	220
2009	937	699	146	255
2010	1,006	785	152	248
2011	977	785	149	232
2012	870	706	149	224

Examining a mechanism by which the e-reporting program may lead to increased violations

A key remaining question is whether the increase in reported violations in Ohio in the post-policy period is consistent with the possibility of more effective targeting of inspection and enforcement

[21] There are many papers in the regulation-compliance literature (e.g., Gray and Shadbegian 2005) that find that regulators tend to inspect facilities with past violations more often than facilities expected to be in compliance.

activities by state authorities, made possible by more complete and accurate reporting of DMR information and the freeing up of resources no longer needed to administer the reporting system.[22] Harrington (1988), in response to the observation that environmental fines tend to be quite low yet most entities still comply with regulations, develops a dynamic repeated game model that allows an environmental regulator to markedly increase the expected long-run penalty for non-compliance by creating two categories of regulated plants - compliers and non-compliers. Because the compliers are well-behaved, they are rarely inspected. The non-compliers face more enforcement activity. A non-complier can earn its way to "complying" status by demonstrating a successful record of compliance over some period of time, while a complier caught violating regulations can be moved to the non-complier category and inspected more frequently.

To formally examine whether moving to the online reporting and auditing program in Ohio may have enabled state regulators to more effectively target inspections and enforcement activity in the post-policy period towards plants with a history of recent violations (i.e., non-compliers), we estimate a Harrington-style model for complying and non-complying plants that report wastewater discharges. To separate plants into "compliers" and "non-compliers" we use the distribution of the number of reported violations from the previous year. If a plant's reported violations are at or exceed the 90th percentile, that plant is designated as a non-complier; otherwise the plant is designated a complier for the current year.

We estimate the following model:

$$Enforcement_{it} = \Upsilon + \delta_1\, Ohio{*}Post\ Policy + \delta_2\, Non\text{-}Complier + \delta_3\, Ohio{*}Non\text{-}Complier + \delta_4\, Ohio{*}Post\ Policy{*}Non\text{-}Complier + \delta_5\, Post\ Policy + \delta_6\, Post\ Policy{*}Non\text{-}Complier + \delta_7 + t + t^2 + e_{it} \quad (3)$$

where $enforcement_{it}$ is either the number of inspections or enforcement actions directed at plant i in period t, and non-complier equals 1 if the plant's reported violations are at the 90th percentile or higher in the previous year. Enforcement is a count variable, so we estimate equation (3) with a fixed effects Poisson model. Note that in a fixed-effect model, facilities that receive none of that enforcement type during the entire period drop out of the analysis.

[22] A case study noted that no full-time staff was needed to maintain the new system: "I used to dedicate over 40 hours a week to this program answering hundreds of e-mails and phone calls by permit holders. After the deployment of Ohio [electronic DMR] system, I've been able to reduce my level of effort 75 percent to just 10 hours a week." (U.S. EPA 2011).

Table 10: Triple Difference Results to Identify Possible Post Period Enforcement Targeting

Panel Fixed Effects				
	Number of State Inspections		Number of State Enforcement Actions	
	(1)	(2)	(3)	(4)
Ohio*Post Policy	-0.138*** (0.034)	-0.378*** (0.040)	-0.621*** (0.091)	-0.478*** (0.097)
Ohio*Post Policy*Non-Complier	0.292*** (0.106)	0.328*** (0.125)	0.056 (0.234)	-0.580*** (0.293)
Post Policy	-0.237*** (0.041)	-0.128*** (0.047)	-1.30*** (0.119)	-1.34*** (0.125)
Non-Complier	0.084 (0.058)	0.038 (0.075)	-0.090 (0.161)	-0.491** (0.232)
Ohio*Non-Complier	-0.021 (0.079)	0.002 (0.093)	0.293 (0.209)	0.693*** (0.268)
Post Policy*Non-Complier	-0.153** (0.077)	-0.169* (0.101)	0.330* (0.174)	0.857*** (0.247)
Quadratic time trend	Y	Y	Y	Y
Sample	Full	CEM	Full	CEM
Obs.	16,350	16,350	9,734	9,734

* denotes significance at the 10 percent level; ** denotes significance at the 5 percent level; *** denotes significance at the 1 percent level.

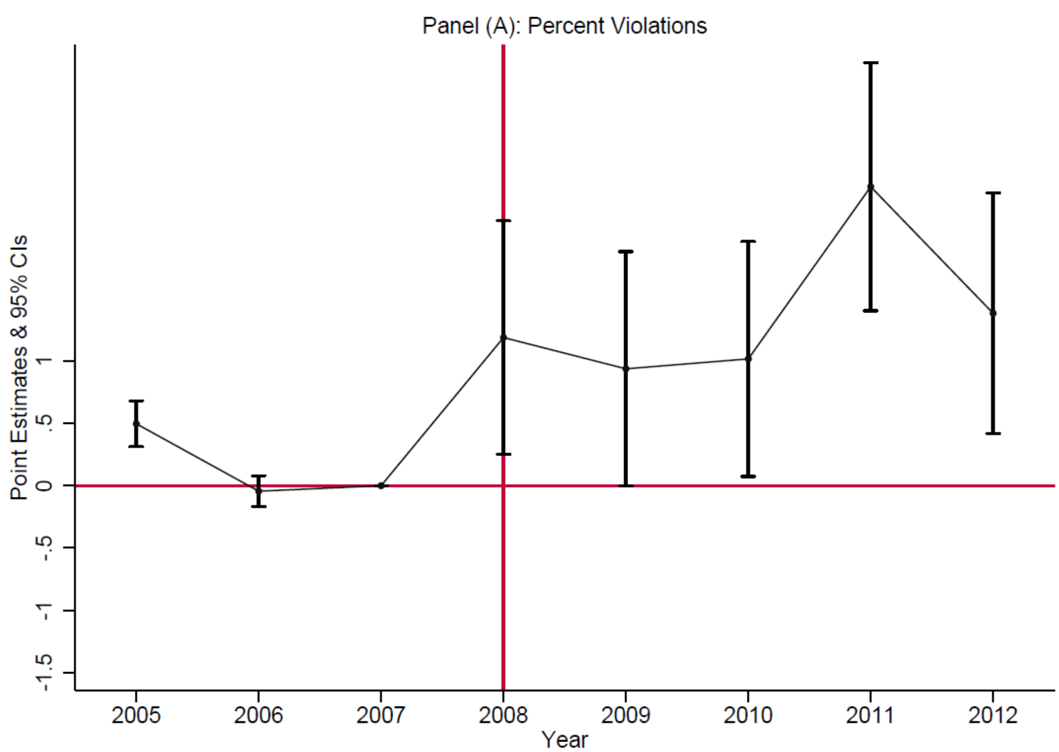
Table 10 indicates that there was a significant decline in the number of inspections and enforcement actions in the post policy period in Indiana for both compliers (*Post Policy*) and non-compliers (*Post Policy*Non-Complier*). In Ohio, the decline was even larger for inspections and enforcement actions against compliers (*Ohio*Post Policy*), but this was at least partly offset by a significant increase in inspections targeted to non-compliers (*Ohio*Post Policy*Non-Complier*). Conducting more inspections at non-complying facilities and fewer at complying facilities is consistent with regulators more effectively targeting their inspections in the post-policy period. That said, we find some evidence that state regulators in Ohio directed relatively fewer enforcement actions towards non-compliers in the post-policy period; this effect is only statistically significant in the CEM model.

Assessing Parallel Trends Assumption

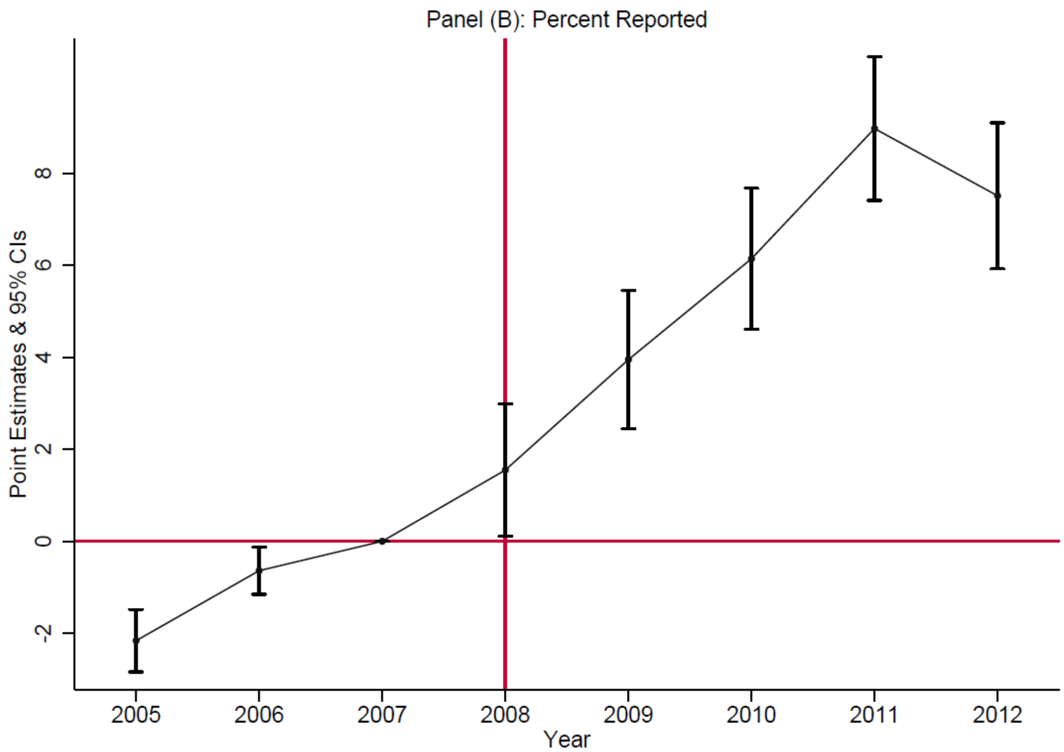
The difference-in-difference estimator requires the identifying assumption that, absent the introduction of the online reporting and auditing program in Ohio, average *Percent Violation*, *Percent Reported*, *TSS Discharge-Limit Ratio* and *BOD Discharge-Limit Ratio* for Ohio and Indiana facilities exhibited similar trends over time prior to the adoption of the e-reporting program in Ohio (i.e., prior to 2008). To assess whether we see parallel pre-treatment trends, we conduct an event study analysis where we estimate a variant of equation (2) that includes 3-year incremental lead and 4-year lag indicator variables for *Ohio*Post Policy.* The estimated "lead" coefficients capture the incremental difference in the outcome of interest between Ohio facilities – the treated group - and Indiana facilities - the control group - and allow for this association to vary over time relative to the adoption of the online reporting and auditing program in Ohio. We ue 2007, the period before adoption, as our base year. The solid vertical line depicts the beginning of the treatment period in 2008.

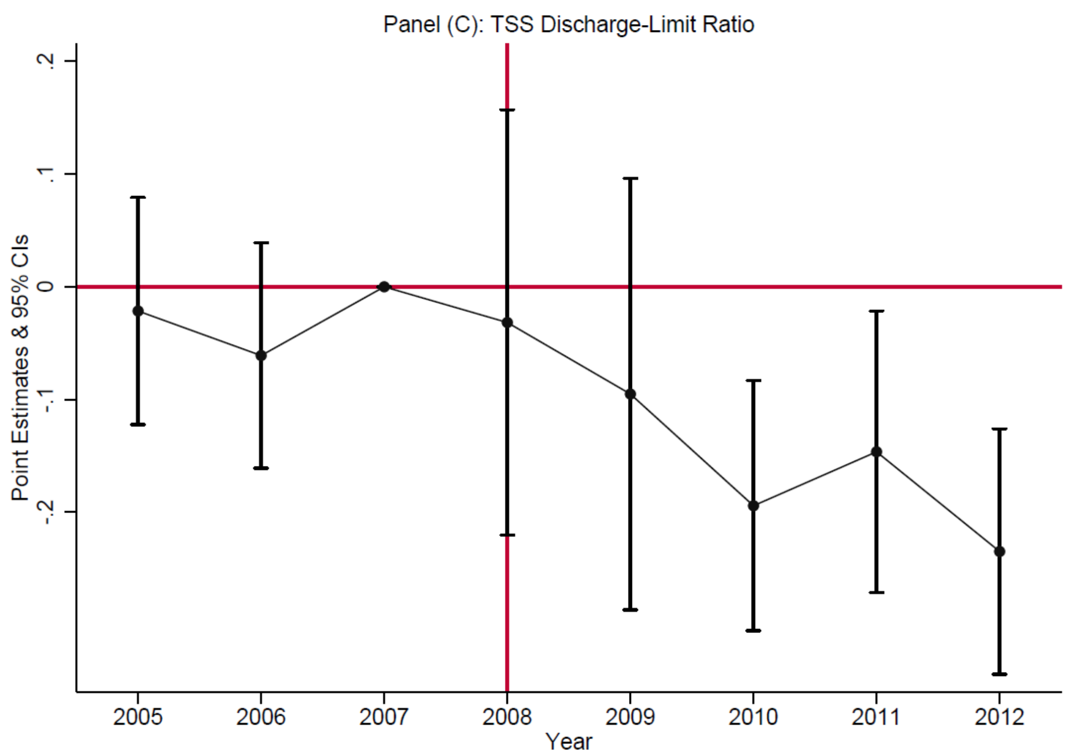
Panels A – D in Figure 2 graphically present the results for all four outcome variables for the CEM sample.[23] As can be seen in each case there are no discernable trends during the pre-treatment period – while some individual years show significant differences, we do not see consistently significant pre-treatment differences combined with a clear trend in the year-to-year changes. These results provide us with greater confidence that the estimated effect of the online reporting and auditing program on the compliance behavior of Ohio facilities is not biased by pre-existing trends. These results are also consistent with the results reported in Tables 2 – 5, but also demonstrate persistent and, for three of the four outcome variables, consistently statistically significant effects over time in the post-policy period.

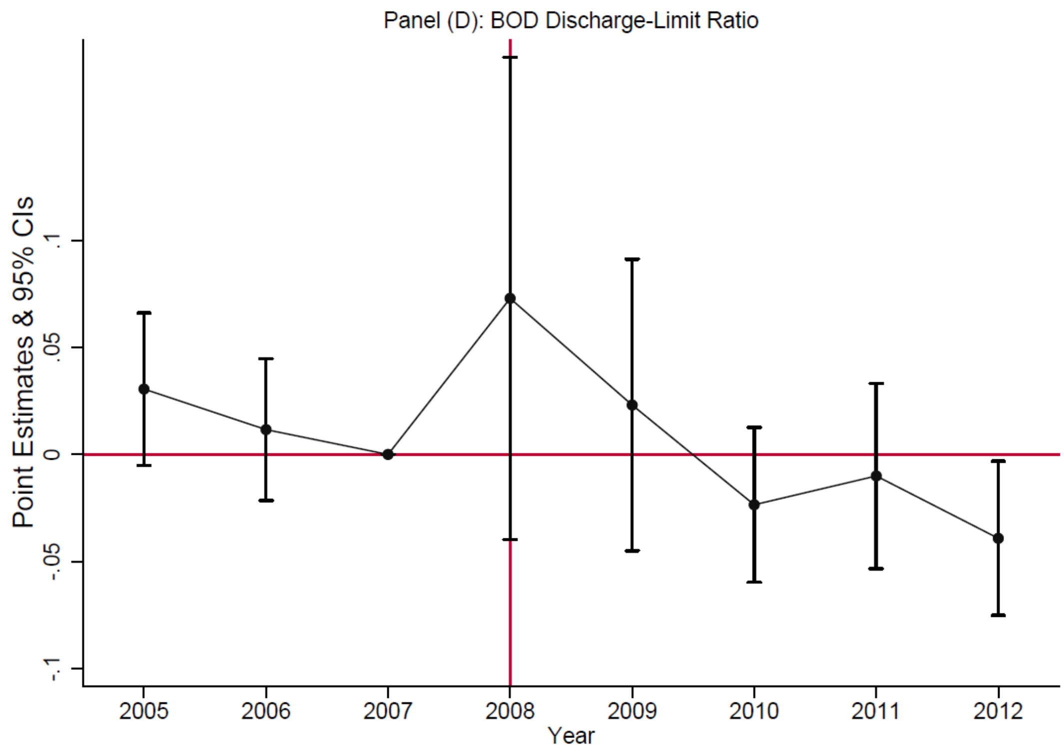
Figure 2 – Event Study Analysis of Parallel Trends (CEM Sample)



[23] The results for the full sample are nearly identical.


Panel (B): Percent Reported
Point Estimates & 95% CIs
-2
0
2
4
6
8
2005
2006
2007
2008
2009
2010
2011
2012
Year


Panel (C): TSS Discharge-Limit Ratio
Point Estimates & 95% CIs
-.2
-.1
0
.1
.2
2005
2006
2007
2008
2009
2010
2011
2012
Year



VIII. Conclusion

The results from this study provide evidence that the mandatory state-level online reporting and auditing program for wastewater discharges in Ohio had a positive effect on the compliance behavior and environmental performance of its facilities compared to similar facilities in Indiana. Using a difference-in-difference estimation strategy, we find that the online reporting and auditing program resulted in more complete filed reports but also greater reported violations for Ohio facilities compared to untreated facilities in Indiana. Moreover, the online reporting and auditing program is associated with lower discharge-limit ratios for both TSS and BOD, though these reductions are larger for TSS than for BOD and the latter are not always statistically significant. We also find that these program impacts are larger for minor dischargers and for publicly owned facilities. Leveraging the Harrington model, we find suggestive evidence that higher reported violations may be due in part to more efficient targeting of inspections on the part of state authorities.

While this analysis examines the effects of a specific state-level mandatory online reporting and auditing program, the results may provide insights into the nature of possible improvements in compliance and enforcement that can be gained by leveraging tools that automate and provide direct feedback to regulated facilities. The U.S. Environmental Protection Agency finalized a nationally mandated online reporting tool for wastewater discharges in 2015 in which states continue to have the option to leverage their own tools to collect this information and pass it along to the EPA. The specific features embedded in these tools may therefore influence the degree to which these state-level systems affect facility behavior and result in improvements in compliance. Likewise, the results of this study also point to potential dividends that may be gained from integrating automated auditing features via machine learning or AI into existing reporting tools at the national or state level.

References

Benami, E., N. Jo, E. Ragnauth, and D. Ho (2023). “Drop a Line, Submit on Time? Randomized Tailored Reminders Improve Pollution Reporting Timelines.” Working paper.

Blondin, J., K. Abu-Hasaballah, H. Tennen, and R. Lalla (2010). “Electronic versus Paper Diaries: A Pilot Study of Concordance and Adherence in Head and Neck Cancer Patients Receiving Radiation Therapy.” *Head and Neck Oncology* 2: 29-33.

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.

Buchsbaum, J., M. Greenstone, and K. Meckel (2024). “Can Machine Learning Improve the Environment? Evidence from a National Field Test Targeting Hazardous Waste Inspections” Working paper. April.

Cameron, C., and P. Trivedi. *Microeconometrics Using Stata*. College Station, TX: Stata Press, 2009.

Curran, B., B.D. Meyer, and D. Wu (2026). “Misleading Estimates from Nonlinear Models with a Binary Outcome” Working Paper No. 35366. National Bureau of Economic Research.

Enfotech News (2008). “Ohio Success Story – Computer Technologies Help Environmental Compliance.”

Engstrom, D., D. Ho, C. Sharkey, and M. Florentino-Cuéllar. 2020. *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*. Administrative Conference of the United States (ACUS).

Environmental Council of the States (2006). *Environmental Information Exchange Network: Exchange Network – Return on Investment and Business Process Analysis Final Report*. http://www.exchangenetwork.net/benefits/ROIReport_v1.0.pdf

Evans, M., L. Liu, and S. Stafford (2011). “Do environmental audits improve long-term compliance? Evidence from manufacturing facilities in Michigan.” *Journal of Regulatory Economics* 40: 279-302.

Friesen, L. (2006). “The Social Welfare Implications of Industry Self-Auditing.” *Journal of*

Environmental Economics and Management 51: 280-294.

Gray, W. and R. Shadbegian (2005). “When and Why do Plants Comply? Paper Mills in the 1980s.” *Law and Policy*, 27, 2005, 238-261.

Gray, W., and J. Shimshack (2011). “The Effectiveness of Environmental Monitoring and Enforcement: A Review of the Empirical Evidence.” *Review of Environmental Economics and Policy* 5: 3-24.

Green, A., E. Rafaeli, N. Bolger, P. Shrout, and H. Reis (2006). “Paper or Plastic? Data Equivalence in Paper and Electronic Diaries.” *Psychological Methods*, 11, 1: 87 – 105.

Harrington, Winston (1988) “Enforcement Leverage when Penalties are Restricted,” *Journal of Public Economics* 37: 29 53.

Hino, M., E. Benami, and N. Brooks (2018). “Machine Learning for Environmental Monitoring.” Nature Sustainability 1: 583-588.

Iacus, S.M., King, G. & Porro, G., (2012). “Matching for Causal Inference Without Balance Checking: Coarsened Exact Matching.” Political Analysis 20(1), 1-24. doi:10.1093/pan/mpr013.

Internal Revenue Service (2010). “IRS E-File: It’s Safe, It’s Easy, It’s Time.” January 15. Available at: https://www.irs.gov/pub/irs-news/ir-10-004.pdf .

Johnson, M., D. Levine, and M. Toffel (2023). “Improving Regulatory Effectiveness through Better Targeting.” American Economic Journal: Applied Economics 15(4): 30-67.

Karkkainen, B. (2001). “Information as Environmental Regulation: TRI and Performance Benchmarking, Precursor to a New Paradigm?” *Georgetown Law Journal* 89: 257.

Khanna, M., and D. Widyawati (2011). “Fostering Regulatory Compliance: The Role of Environmental Self-Auditing and Audit Policies.” *Review of Law and Economics* 7(1), article 7.

Kopczuk, W., and C. Pop-Eleches (2007). “Electronic Filing, Tax Preparers, and Participation in the Earned Income Tax Credit.” *Journal of Public Economics* 91: 1351-1367.

Mishra, B., D. Newman, and C. Stinson (1997). “Environmental Regulations and Incentives for Compliance Audits.” *Journal of Accounting and Public Policy* 16; 187-214.

Perez Henriquez, B. (2004). "Information Technology: The Unsung Hero of Market-Based Environmental Policies." *Resources.* Resources for the Future.

Pfaff, A., and C. Sanchirico (2000). "Environmental Self-Auditing: Setting the Proper Incentives for Discovery and Correction of Environmental Harm." *Journal of Law, Economics, and Organization* 16 (1): 189-208.

Shimshack, J. (2014). "The Economics of Monitoring and Enforcement." Annual Review of Resource Economics 6: 339-360.

Shimshack, J. and M. Ward (2008). "Enforcement and Over-Compliance." *Journal of Environmental Economics and Management* 55: 90-105.

Stafford, S. (2005). "Does Self Policing Help the Environment?" *Vermont Journal of Environmental Law* 6: 1-22.

Stafford, S. (2007). "Should You Turn Yourself In? The Consequences of Environmental Self-Policing." *Journal of Policy Analysis and Management* 26 (2): 305-326.

Stone, A., S. Shiffman, J. Schwartz, J. Broderick, and M. Hufford (2003). "Patient Compliance with Paper and Electronic Diaries." *Controlled Clinical Trials* 24: 182-199.

U.S. Environmental Protection Agency (EPA) (1997). *Information Streamlining Plan*. Environmental Protection Agency.

U.S. EPA (2011). Case Study: Ohio Environmental Protection Agency's Electronic Discharge Monitoring Report (eDMR) System Reaches 99% Adoption. EPA305-F-10-001.

U.S. EPA (2025). *Artificial Intelligence Strategy for OMB Memorandum M-25-21*. https://www.epa.gov/system/files/documents/2025-10/epa_ai_strategy_plan_final2025.pdf .