Can Collective Action Institutions Outperform the State? Evidence from Treatment of Abandoned Mine Drainage.

Harleman, Max and Weber, Jeremy G.

Georgia College and State University Department of Government and Sociology, University of Pittsburgh Graduate School of Public and International Affairs

22 December 2023

Can Collective Action Institutions Outperform the State? Evidence from Treatment of Abandoned Mine Drainage. *

Max Harleman and Jeremy G. Weber

June 25, 2024

Abstract

A core public administration literature seeks to understand whether decentralized collective action institutions will emerge to provide public goods, such as management of environmental resources. Few studies examine how they perform relative to the state at providing public goods, and they fail to account for the possibility that the state might self-select into providing public goods in the most challenging contexts. If it does, finding that the state performs worse than collective action institutions could reflect its more challenging context rather than differences in knowledge, skill, or motivation. We examine several quantitative measures of performance in remediating polluted water discharges from abandoned coal mines in Pennsylvania, a task sometimes done by the state and sometimes by nonprofit watershed associations. We find that the two types of institutions address discharges with generally similar water quality problems and build systems that yield similar initial improvements in water quality. However, watershed association systems better maintain effectiveness at reducing acidity and removing heavy metals over time. The findings suggest a role for sustained public investment in collective action institutions to address complex and enduring environmental problems.

Keywords: collective action, abandoned mines, water quality, decentralization

*Harleman: Department of Government and Sociology, Georgia College and State University, 410 West Greene Street Campus Box 18. Milledgeville, GA 31061. max.harleman@gcsu.edu. Weber: Graduate School of Public and International Affairs, University of Pittsburgh, 3601 Wesley W. Posvar Hall, Pittsburgh, PA 15260. jgw99@pitt.edu, 3601 Wesley W. Posvar Hall, Pittsburgh, PA 15260. We would like to thank Caroline Fettes and Irving Hu for providing research assistance.

1 Introduction

Private community members have sometimes voluntarily organized into "collective action institutions" to deliver public goods and services, such as the management of common-pool resources like pastures or surface water (North, 1990; Ostrom, 1990). Many scholars and practitioners have argued that collective action institutions provide public goods more efficiently and effectively than "the state," a term we use to refer to public institutions such as federal, state, and local governments. For example, proponents of watershed associations, one type of collective action institution, argue that they are more cost-effective and able to address problems outside the scope of centralized environmental regulation, such as non-point source pollution or habitat destruction (Lubell et al., 2002, p.149). On the other hand, collective action institutions may fail to achieve the success predicted by economic theory (Marshall, 2015; McCord et al., 2017), especially when they attempt to address inherently complex and dynamic policy problems like environmental remediation (North, 1994; Marshall, 2010; Dietz et al., 1212).

A body of empirical literature concludes that collective action institutions perform better than the state, but it has two notable limitations. The first is that many studies rely on case studies (e.g., D'Souza and Nagendra (2011); Mathenge et al. (2014); Wade (1988); Cleaver and Toner (2006); Klassen and Evans (2020); Frantzeskaki (2019); Foster and Iaione (2016)), some of which are chosen because the observed performance of the collective action institutions aligns with their theoretical advantages. Though valuable for their details, such case studies may have limited generalizability even to other collective action institutions in the same field and region. While the literature presents a well-developed theory on why community members organize to provide public goods (Ostrom, 1990, 1999), the second limitation is that it does not consider the contextual factors that might lead the state to provide them. This leaves open the possibility that the state might self-select into providing public goods in the most severe and challenging settings, in which case a finding that it performs worse than collective action institutions could reflect its more challenging context rather than differences in knowledge, skill, or motivation.

We overcome both limitations in our study of the treatment of abandoned mine drainage in

Pennsylvania, a problem that degrades tens of thousands of miles of streams in Appalachian states (Kruse Daniels et al., 2021). The scale of the problem and the two distinct funding sources that address it have produced many treatment systems built by different institutions. The institutional variation stems in large part from a Pennsylvania law that authorizes state grants to nonprofit watershed associations that propose systems, and a federal law that provides funds to the state Department of Environmental Protection to implement and oversee treatment. As a result, we have a large sample of treatment systems—61 of which were built by the state and 203 by a diversity of watershed associations (Figure 1).

Our rich data also permit testing for institutional selection, which motivates our first empirical question: do watershed associations (the collective action institutions) and the Department of Environmental Protection (the state) specialize in treating discharges of different sizes or severity? Our second empirical question considers cases where the watershed associations and the state address similar discharges and asks: do association systems outperform state systems as measured by monitoring activities, cost-effectiveness, and their initial, average, and long-term effectiveness at cleaning discharged water?

We find that state systems address slightly more acidic discharges. However, many discharges are similar across the two groups of systems, so we match each state system with one or more comparable association system. In the matched sample, the systems treat discharges with similar water quality problems, flow rates, and number of discharge points. Comparisons across the two groups reveal that watershed association systems perform as well as, if not better than, state systems across a variety of outcomes. The two groups of systems perform similarly at cleaning discharge water on average over their observed lifespans, but state systems decline in their ability to treat the discharge over time. Conversely, association systems appear to maintain their effectiveness, suggesting that local individuals better recognize a system's deterioration and act to maintain and upgrade it. State systems are formally monitored more frequently, as measured by the number of laboratory water quality tests taken at the discharge. But informal, on-the-ground monitoring may explain associations' superior maintenance. In our data, the typical association manages only three geographically concentrated systems, relative to over 60 dispersed systems managed by the state.

Our findings have contemporary policy relevance. The 2021 Infrastructure Investment and Jobs Act allocates a historic $11.3 billion for states and tribes to address abandoned mine problems over 15 years, and reauthorizes a federal tax on coal that funds abandoned mine reclamation (Legere, 2021). The increase in funding implies a need for substantial scaling up of reclamation efforts, as it represents a roughly three-fold increase in annual federal funding for abandoned mine problems (U.S. Department of the Interior, 2022). Our findings suggest that the state Department of Environmental Protection can continue to devolve treatment responsibilities to watershed associations. Because Pennsylvania's mine drainage problems are not known to be less complex than elsewhere, other state governments should not categorically rule out partnering with associations in their state. Given the complexity of treating mine drainage, the finding also suggests that the more than 2,600 watershed groups (U.S. Environmental Protection Agency, 2017) across the U.S. could be useful partners for addressing similar or simpler pollution problems.

Academically, the findings expand two related bodies of public administration literature. First, by considering a large sample of projects and by accounting for the state self-selecting into providing public goods in more complex settings, we add rigorous evidence to the literature on the relative performance of state and collective action institutions. This literature is exemplified by Ostrom (1965), who studied ground water management in the West Coastal Basin of Southern California in the 1960s. Using a comparative case study approach, Ostrom found that watershed associations were more effective than state-mandated efforts in other basins at enforcing rules to conserve water. Other studies find superior performance of collective action institutions using the cases of agriculture and irrigation systems (Shivakoti et al., 2005; Lam et al., 1998; Wade, 1988), surface water resources (D'Souza and Nagendra, 2011), public water supplies (Cleaver and Toner, 2006; Pahl-Wostl et al., 2012; Mathenge et al., 2014), forests (Hayes, 2006; Ghate and Nagendra, 2005), and urban public goods (Foster and Iaione, 2016; Frantzeskaki, 2019).

Second, we expand the empirical literature comparing the performance of public and private organizations in diverse settings. Public choice theorists have argued that profit-seeking can lead private firms to deliver the quantity and quality of public goods that best meet citizens' demand (Chubb and Moe, 1988). On the other hand, contract failure theorists have argued that public organizations may perform better and be perceived as more trustworthy in providing public goods for which citizens cannot easily assess quantity and quality (Hansmann, 1996). A vast empirical literature explores public versus private performance in the contexts of nursing homes (Broms et al., 2023), hospitals (Rushing, 1974), emergency services (Sobel and Leeson, 2006), and schools (Ballou and Podgursky, 1998; Chubb and Moe, 1988). The studies compare the quality of services before and after they are privatized, or compare public and private organizations that deliver similar goods and services. The approaches have produced mixed findings (Rainey and Chun, 2007).

A smaller literature finds differences in organizational and managerial characteristics across three sectors: public, private, and nonprofit (Lee and Wilkins, 2011; Koliba et al., 2011; Cohen, 2001; Rainey and Bozeman, 2000; Lan and Rainey, 1992). But Amirkhanyan et al. (2008, p.330) note that most studies of cross-sector performance group public and nonprofit organizations together. We know of only two studies that break out nonprofit and public organizations. Amirkhanyan et al. (2008) find similar performance across public and nonprofit nursing homes in terms of quality of care and accessibility for low-income individuals, and Johansen and Zhu (2013) find that public and nonprofit hospitals have similar responsiveness to legislative mandates. We expand upon this literature by being the first to compare nonprofit and public performance in an environmental context that includes very small nonprofit institutions. Most of the watershed associations in our sample are run by one or two staff members and a handful of volunteers, as opposed to large professionalized hospitals and nursing homes.

2 Abandoned Mine Drainage and Funding Mechanisms

2.1 Abandoned Mine Drainage

Since passage of the Federal Surface Mining Control and Reclamation Act of 1977, mining companies have had to provide financial commitments known as bonds to be allowed to mine. When abandoning a mine, they must comply with reclamation requirements or forfeit their bonds. Prior to the Act, companies could usually abandoned their mines with little to no repercussions. As a result, about half a million abandoned mines with no responsible owner are spread throughout 32 of 50 states in the U.S. (Glatzel and Gordon, 2018), and are unlikely to be addressed apart from public funding. Because abandoned mine drainage flows into streams and rivers, they affect common-pool resources widely enjoyed by outdoor enthusiasts including fishers and boaters.[1] In addition, discolored streams and rivers are a visual disamenity, and the acidity and metals can increase water treatment costs for downstream industrial or municipal users (Hansen et al., 2010).

The problem of abandoned mine hazards is severe in Pennsylvania, which has the largest inventory of unremediated abandoned mines (CRS, 2020), and where the state's Department of Environmental Protection reports that abandoned mine drainage impairs 5,524 miles of stream (Pennsylvania DEP, 2022). The degradation stems from water moving through a mined area and interacting with rock exposed to oxygen through mining. The interaction results in the water often having the same acidity as tomato juice, and containing high concentrations of metals such as iron and aluminum. When it flows into nearby streams, the drainage can kill fish and turn water orange as dissolved iron precipitates. Efforts to address abandoned mine drainage involve a diversity of treatment systems that all aim to reduce the acidity of the drainage, capture metals, and discharge cleaner water into the nearest stream. Most systems are "passive" treatment systems, which typically involve one or more retention ponds or wetlands lined with limestone to reduce acidity and metal contents before the water flows into nearby rivers and streams (U.S. Department of the In-

[1] In Ostrom's typology of goods, streams and rivers polluted by abandoned mine drainage are "common-pool resources" (Ostrom, 2005, p.24). This is because 1) clean water is fully *subtractable* because pollution by one user creates less clean water by other users, and 2) there is high *difficulty of excluding users* because there were no policies prior to 1977 to ensure remediation.

terior, 2003). While less expensive than active systems (water treatment plants), they still require large upfront costs to design, construct, and maintain, including costs to replenish limestone or to dredge wetlands that have filled with metals.

2.2 Funding Treatment in Pennsylvania

One major source of funding for treating abandoned mine drainage is set-aside money authorized by the 1990 and 2006 amendments to the federal Surface Mining Reclamation and Control Act. The set-aside provisions allocate money from the federal Abandoned Mine Land Fund, funded by a federal tax on coal, to state or tribal accounts to treat abandoned mine drainage. In the case of Pennsylvania, the Department of Environmental Protection assumes responsibility for the set-aside money (Pennsylvania DEP, 2023a).[2] This involves overseeing all phases of a treatment system's life: design, construction, operation, monitoring, and maintenance. The federal Office of Surface Mining Reclamation and Enforcement annually evaluates the Department's abandoned mine land program. The Department only expends the set-aside funds in qualified watersheds with a hydrologic plan documenting the health of the watershed, and its topography, human activities, and ecology (Pennsylvania DEP, 2016).

Another major source of funding comes from state watershed protection grants authorized by the Pennsylvania Environmental Stewardship and Watershed Protection Act. The Act directs the Department to provide grants to local governments and watershed associations for abandoned mine drainage treatment, among other activities.[3] The grants have funded work by a variety of watershed associations, which we describe in detail below. Although watershed grants may address restoration of the same streams as the set-aside program, the Department does not assume legal or maintenance responsibilities for systems built with watershed grants. The Department states that it prioritizes use of set-aside funds for operating and maintaining treatment plants or systems

[2] See also Chapter 4-130 of the Federal Assistance Manual of the Office of Surface Mining Reclamation and Enforcement (Office of Surface Mining Reclamation and Enforcement, 2023). The entity within the Pennsylvania Department of Environmental Protection that oversees the set-aside program is the Bureau of Abandoned Mine Reclamation.

[3] See section 6105(b)1(iii) of Pennsylvania General Assembly (1999).

"constructed by DEP, or operated by or on behalf of DEP" (Pennsylvania DEP, 2016).

To put a finer point on the implications of the funding streams, with federal set-aside funds the Department acts as a funding intermediary and a general contractor. It is accountable to its funder, the federal Office of Surface Mining. With the watershed grants, the Department is the funder, using only state money. It selects which proposals to fund and ensures that recipients comply with its reporting requirements, but it does not act as a general contractor by soliciting subcontractors or putting staff in the field to directly construct or monitor systems. Nor does it report activities and results to a higher authority.

State and association systems are likely built to similar standards because the subcontractors that design and construct systems serve both the state and associations. The Department has established guidelines that apply only to its own treatment systems, but they recommend them as a resource for organizations applying for grants (Pennsylvania DEP, 2016). Despite potentially similar design and construction quality, performance could vary greatly over time based on whether the associations or the state monitor the systems and address problems as they arise.

2.3 Watershed Associations

Various authors identify a movement in the late 20th century towards managing natural resources at the watershed level (Kenney, 1999; Born and Genskow, 2001). Tarlock (2000) notes that the focus on watershed-level planning came alongside the "emergence of grassroots organizations interested in conserving and restoring specific places" (p.188). The federal government supported the movement through the 1987 amendments to the Clean Water Act that authorized the Environmental Protection Agency to fund community-based efforts to fight non-point source pollution (Hardy and Koontz, 2008).

Watershed associations in Pennsylvania reflect this historical development. Westmoreland County, for example, reports seven main watershed associations, all of which were created between 1970 and 2000 (Westmoreland Conservation District, 2023). Most organizations, like the Babb Creek Watershed Association are staffed by volunteers with little to no compensation. Others are more professional. The Mountain Watershed Association, for example, has a full-time staff of 12 professionals. The associations support diverse activities that promote surface water quality and aquatic life in the watershed such as streambank restoration, fencing, and tree planting. Since abandoned mine drainage affects many of their watersheds, they have also supported treatment efforts, sometimes by pursuing large grants and taking full responsibility for design, construction, and maintenance of treatment systems. A repository of information on treatment systems in Pennsylvania lists roughly 30 different watershed associations or coalitions in its database (Datashed, 2023b). The associations are helped by the Eastern and Western Pennsylvania Coalitions for Abandoned Mine Reclamation, which have full-time, experienced staff that provide technical assistance and serve as a liaison with the Department of Environmental Protection.

3 Theory, Literature, and Hypotheses

3.1 Institutional Selection

A core literature in the field of public administration and management explores why decentralized collective action institutions voluntarily organize to provide public goods that improve local well-being. While the early literature explored community action to improve urban public goods including policing and domestic water supplies (Ostrom, 1965; Ostrom and Whitaker, 1973; Ostrom et al., 1973), its primary focus has been on the emergence of collective action institutions to manage common-pool environmental resources. The literature is exemplified by Lubell et al. (2002), who study the emergence of watershed associations in the U.S. as a dependent variable. With Ostrom (1990, 1999), they define a watershed as an action arena where the *action situation* and *characteristics of actors* interact to determine institutional rules that govern a natural resource system.[4] They find that watershed associations emerge when their benefits to actors are high because the environmental problem is severe and enforcement of government rules to protect the

[4]The action situation includes the initial states of natural resources, institutional arrangements, and usage patterns and their associated costs and benefits to resource users. Characteristics of actors include initial resource endowments across individuals and interest groups, along with their preferences and knowledge of the action arena.

watershed is weak, and when local incomes and public support for partnerships are high enough to overcome transaction costs.

There is comparatively little literature on why the state would provide public goods without an explicit legislative mandate. Just as collective action is more likely when state enforcement of environmental rules is weak, the state may choose to intervene when collective action institutions have failed to emerge. This perspective of the state filling a policy void aligns with Salamon (1987), who envisions state action as a response to a failure of collective action, rather than the traditional explanation of state action as a response to market failure (as noted by Amirkhanyan et al. (2008, p.329)).

Two empirical challenges make it difficult to hypothesize about whether collective action institutions or the state will provide a public good. First, researchers must measure and analyze the many variables that interact within an action arena to determine institutional selection—as many as 44 unique variables in the framework presented by Ostrom (2009, p.421). Second, institutional selection is a temporally ambiguous two-stage process in which the state may crowd out collective action institutions, or vice versa. While these challenges complicate our ability to precisely understand why the Department constructs systems at some discharges and watershed associations construct them at others, our empirical approach accounts for a form of institutional selection that may bias previous studies of the effect of institutional choice on performance: the severity and complexity of the environmental problem.

Considered alone, severity has an ambiguous relationship with institutional selection. On one hand, severe environmental problems raise the benefits to collective action. On the other, the state may choose to intervene where the benefits of its involvement are highest. For instance, the geologists, hydrologists, and construction experts within an environmental regulator may choose to remediate pollution when it is most pervasive or where technical details make cleanup most complex. This latter explanation sets our first hypothesis:

Hypothesis 1: *The state will build systems in more complex and severe settings compared to association systems, as measured by the number of contaminated inflows and the volume and contamination of water running into them.*

3.2 Institutional Performance

As noted in the introduction, a large literature compares collective action institution and state performance, and a closely related literature compares private, public, and nonprofit performance. We contribute to it with rigorous evidence based on a relatively large sample of state and collective action projects and by accounting for institutional self-selection that may have biased past studies. We also contribute by operationalizing the broad construct of environmental performance with four quantifiable measures: effectiveness, cost-effectiveness, monitoring, and maintenance.

Prior studies have considered the *effectiveness* of state and collective performance in diverse ways. Some have quantitatively operationalized effectiveness by coding case study evidence of environmental quality into ordinal scales (Lam et al., 1998; Shivakoti et al., 2005; Pahl-Wostl et al., 2012; Hayes, 2006). Others have qualitatively operationalized perceptions of effectiveness with surveys and interviews (D'Souza and Nagendra, 2011; Mathenge et al., 2014; Ghate and Nagendra, 2005), and still others have relied on secondary case study data (Foster and Iaione, 2016; Frantzeskaki, 2019; Cleaver and Toner, 2006; Klassen and Evans, 2020). In contrast, we use a continuous, quantitative measures of water quality observed at the project level over time. The mixed results from the private versus public performance literature discussed in the introduction, and the institutional details in Section 2.2 suggest that the two types of systems are likely built to similar standards. This prompts our second hypothesis:

Hypothesis 2: *The two types of systems will achieve similar initial and average lifetime effectiveness in terms of reducing acidity and removing heavy metals from discharge water.*

To our knowledge, few studies have compared *cost-effectiveness* in quantifiable, monetary terms across state and collective action institutions. State rules that require contractors to be registered with the state and force the selection of a "lowest responsible bid" may be capable of weeding out overly costly and unprofessional bidders. But they may also unintentionally weed out the most capable or cost-competitive contractors that sort towards privater buyers to avoid registering with the state and filling out time-consuming regulatory paperwork. While contracting requirements may elicit low and very cost-competitive bids, with our third hypothesis we expect this will be outweighed by greater in-kind contributions of labor by local watershed members:

Hypothesis 3: *Association systems will be more cost-effective than state systems, as measured by the cost of each gallon of water treated.*

After community members overcome first-order collective action problems to organize into functional institutions, Bates (1988) and Heckathorn (1989) note that they face the second-order collective action problems of *monitoring* and *maintaining* their performance. In the case of abandoned mine drainage, formal monitoring can involve costly field or laboratory tests of air, water, and soil. Maintenance can mean reconstructing existing treatment technologies. The Department states that its top priority for the use of federal abandoned mine drainage funding is to operate and maintain its own systems, even above developing new watershed plans and constructing new systems (Pennsylvania DEP, 2016, p.8). Therefore, with its annual stream of federal funds we expect it to outperform the monitoring and maintenance of associations that may run out of money, face challenges in securing grants, or experience deteriorating organization as members move away or age. These facts motivate our fourth and fifth hypotheses:

Hypothesis 4: *State systems will be monitored more frequently, as measured by the number of water quality readings taken at the systems over their observed lifespans*;

Hypothesis 5: *State systems will be better maintained, as measured by the change in their effectiveness at improving water quality from their first post-construction reading to their last reading*.

The number of water quality readings is only one operational definition of monitoring. We do not observe *informal monitoring*, such as passing by a system to see if effluent appears acidic, which may be carried out more often by local watershed associations. Several studies of forests have found that supplementing formal state monitoring regimes with monitoring by local user groups is necessary for successful resource management (Gibson et al., 2005; Coleman, 2009; Banana et al., 2007; Ghate and Nagendra, 2005; Ostrom and Nagendra, 2006). Therefore, if our findings contradict our fifth hypothesis that state systems will be better maintained, it could be explained by superior informal monitoring of the watershed associations that we do not directly observe.

4 Data and Sample

4.1 System Characteristics and Water Quality Problems

To test our five hypotheses, we use data on abandoned mine drainage treatment systems from Datashed (2023b). Datashed is an online repository that is funded, operated, and maintained by multiple nonprofit and governmental entities (Datashed, 2023a). It contains data on systems in Pennsylvania, including their funding sources, the parties involved in their construction, and a series of water quality variables from samples collected by state employees and association volunteers over time. Datashed provides a system-level report for each system. The report includes measures of influent and effluent water quality averaged across the entire lifetime of the system. Influent measures are from laboratory tests of water sampled at a system's inflow, and effluent measures are from laboratory tests of water sampled at a system's outflow into the nearest river or stream. The system-level averages do not permit tracking water quality over time. We therefore download data on the first and most recent reading from raw laboratory reports for a subset of systems with more than one influent and effluent reading.

Datashed contains system-level reports on 292 passive systems with both inflow and outflow measures. We drop systems without a pH reading at an inflow or outflow, those lacking a measurement of flow, four systems that inexplicably show a large drop in pH between inflow and outflow, and one system that is an extreme high outlier for flow volume (10,955 gallons per minute, with the next highest value being only half this amount). The resulting sample has 264 systems that we classify as either being state or association constructed and maintained (Appendix A.1).

Appendix Table B1 presents descriptive statistics for the systems. The typical system was built between 2003 and 2004 and was funded with 1.4 unique funding sources (e.g. private foundation, federal set-aside, state grants, in-kind contributions). Larger and more complex discharges may be more challenging to treat. To measure complexity, we utilize the number of inflows treated by the system. The typical system has one inflow, and few systems collect and treat multiple inflows. To measure size, we use the flow of water in gallons per minute entering the system from all inflows. The average and median systems in our sample take in 283 and 56 gallons of water per minute.

To measure the severity of water quality problems, we use the influent readings. For each system in our sample, we have at least one influent reading that allows us to define severity with a continuous and a threshold-based measure. The continuous measure is the actual value of influent pH, which is a log scale, and concentrations of manganese (Mn), aluminum (Al), total iron (Fe), and total suspended solids (TSS), which are measured in miligrams per liter (mg/L). For pH, which measures acidity, 7 indicates neutral water and lower values indicate greater acidity. Average pH and concentrations of the four metals over all readings taken at the systems are shown in Appendix Table B1. For instance, the average influent pH is 4.34, which is calculated by allowing each of the 264 systems in our sample to contribute one pH measure that is its average influent pH taken over all its tests in Datashed. We also calculate an average initial influent pH of 4.47, by taking the average pH from the first reading at each system across the 223 systems that have more than one reading. For these systems, the average number of years between the first and the last reading is

12 years.

Our threshold-based measures are based on regulatory thresholds for pH and the four metal concentrations. The Pennsylvania DEP (2016, p.4) sets water quality targets in streams that receive abandoned mine drainage for pH (greater than 6.0), iron (less than 1.5 mg/l), and aluminum (less than 0.5 mg/l). It lacks targets for the other two measures, so we use targets from the U.S. EPA's coal mining effluent limitation regulations—manganese (less than 2.0 mg/l) and TSS (less than 35 mg/l) (Code of Federal Regulations, 2023). Appendix Table B1 shows that 99 percent of the systems treat inflows that fail at least one of the five standards on average over their observed life, with pH (77 percent), iron (78 percent) and aluminum (76 percent) being the most commonly failed standards.

4.2 Effectiveness, Cost-Effectiveness, Monitoring, and Maintenance

Appendix Table B2 presents descriptive statistics for the performance outcome variables, which include measures of initial and average effectiveness, cost-effectiveness, monitoring, and long-term maintenance of effectiveness at improving water quality.

As with measures of influent water quality, we use both continuous and threshold approaches to measure system *effectiveness*. With the continuous measures, we account for initial pollution in influent water by calculating the change in water quality from influent to effluent. For instance, to calculate the change in manganese (Mn) at each system i we calculate $\Delta Ln(Mn_i) = Ln(Mn_{effluent,i}) - Ln(Mn_{influent,i})$. Over the observed life of the average system, effluent had around 61 percent less manganese than influent. The pH scale is logarithmic, so we simply take the change in pH from inflow to outflow and find an average improvement in pH of 1.59 points. To put the change in perspective, it is the difference between the acidity of tomato juice (4.3) and the acidity of water where aquatic life can flourish.

For the threshold approach to measuring effectiveness, we examine whether the effluent water meets the targets set by the Pennsylvania DEP (2016) for pH, iron, and aluminum, and by the U.S. EPA for manganese and TSS. While 82 percent of systems release effluent that fails at least one of the five standards, the particular standard that they fail varies. Of the 264 systems in our sample, 40 percent fail the pH standard, between 40 and 50 percent fail the manganese, aluminum, and iron standards, and very few fail the TSS standard.

To measure *cost effectiveness*, we use the system's initial cost per 100,000 gallons treated. We estimate it by dividing the initial cost of the system by the estimated number of gallons that would run through the system over an assumed 15 year lifespan. We calculate the estimated number of gallons by multiplying the flow rate in gallons per minute by the number of minutes in a 15 year period. On average, systems in our sample cost $206 for each 100,000 gallons of water treated. To measure *monitoring*, we use the combined number of water quality readings of both influent and effluent taken over the observed life of the system. The average system has 53 readings taken over its observed lifespan.

To measure *maintenance*, we calculate the most recent measure of effectiveness (effluent less influent at the last reading, $t = T$) and compare it to the initial effectiveness (effluent less influent at the initial reading, $t = 1$). For instance, to calculate the change in manganese at each system i we calculate:

$$\begin{aligned} Change\ \Delta Ln(Mn_{it}) = {} & (Ln(Mn_{eff,i,t=T}) - Ln(Mn_{inf,i,t=T})) \\ & - (Ln(Mn_{eff,i,t=1}) - Ln(Mn_{inf,i,t=1})). \end{aligned} \tag{1}$$

The average system's most recently observed effect on manganese is 0.12 log points better than its earliest observed effect (Appendix Table B2). Thus, if the system initially reduced manganese by 50 percent, it is most recently reducing it by around 60 percent, indicating increasing average effectiveness over time. In contrast, the typical system experiences a 28 percent increase in aluminum from inflow to outflow relative to the first reading, indicating decreasing average effectiveness over time. Mine discharge water quality should generally improve over time as several studies have documented (Burrows et al., 2015; Merritt and Power, 2022). If the state targets discharges from mines abandoned long ago, it should result in its systems treating less acidic or polluted water. This underscores the value of looking at initial inflow water quality across system types and accounting for any differences when estimating treatment effectiveness.

5 Empirical Approach

5.1 Institutional Selection

Our first empirical question is whether the state selects into addressing the most polluted and challenging discharges. To answer it, we compare the characteristics of its systems to those built by associations. Specifically, we compare means of the variables in Appendix Table B1—those that define system characteristics, complexity, size, and influent severity—using our complete sample of 203 association systems and 61 state systems. We present the actual difference in means and the associated t-statistics for the hypothesis that the two samples come from the same population. We also present the normalized mean difference, which takes a difference in means and divides it by the square root of the average variance of the two groups (Imbens and Wooldridge, 2009). As Imbens and Wooldridge (2009) note, the normalized difference in means is a better measure of covariate differences across two groups when doing causal inference. The traditional t-statistic necessarily increases as the sample size grows, but the normalized difference does not, which is desirable because the challenge of estimating an average treatment effect does not inherently become more difficult as the sample size grows. They note that with a normalized difference in means greater than 0.25 standard deviations, adjusting for the difference through least squares regressions is likely sensitive to model specification.

5.2 Matching and System Performance

Our second empirical question is whether state system's outperform association systems in terms of effectiveness, cost-effectiveness, monitoring, and maintenance. Results that follow show that state systems treat slightly more acidic discharges. If we used the full sample for performance comparisons, a finding that state systems perform worse could reflect more acidic water being more challenging to treat. Therefore, when testing our second through fifth hypotheses, we address the selection by defining our continuous outcomes as the change in water quality from influent to effluent, which accounts for initial pollution in influent water. We also limit our performance comparison to state and matched association systems that address similar water quality problems.

To match systems, we use a combination of exact matching and Mahalanobis Distance Matching (MDM). Exact matching matches units across groups, in our case state and association systems, that have the same value for a given variable. MDM matches units across groups that are nearest in covariate space, with nearness defined by a Mahalanobis distance. The Mahalanobis distance is calculated after standardizing each matching variable (which accounts for matching variables being in different units of measurement) and after accounting for correlations between the multiple matching variables (which accounts for redundancy). We use MDM, rather than other popular matching algorithms like Propensity Score Matching (PSM), for two reasons. First, King and Nielsen (2019) show that MDM produces better balance on observed covariates for a given number of observations drawn from the full sample. Achieving better balance for a given sample size improves our statistical power to identify differences between state and association systems. Second, Ripollone et al. (2018) find that MDM outperforms PSM when there is a high degree of initial covariate balance by matching observations that are closer in a multidimensional covariate space. Because we have a large sample of association systems that closely overlap our smaller sample of state systems, MDM will yield better matches.

Starting with the 61 state systems, we find exact matches on a binary variable indicating whether the water entering the system failed the state standard of 6 pH. This reflects the generally different types of systems needed to treat net acid versus net alkaline mine water (Hedin et al., 2013). The exact match should result in samples of state and association systems with an identical proportion of acidic discharges. We also exact match on a binary variable equal to one if the system is one of the 223 that have more than one water quality reading, which allows us to compare long-term maintenance outcomes across the two types of systems. For MDM matching, we calculate the Mahalanobis distance for each system based on the eight MDM variables in Table B3. They include system characteristics (year built, the number of inflows, and the inflow volume) and the four continuous measures of influent water quality. Matching on these variables helps balance the size and complexity of the treated discharges and the severity of the influent water quality problem. Tables B4 and B5 present descriptive statistics for all of our covariates and outcome variables for the one-to-two matching without replacement sample.

Our preferred matching approach is to match each of the 61 state systems to two association systems "without replacement," meaning that the association sample will include 122 unique systems. For our sample, one-to-one matching does not yield a clear improvement in covariate balance as measure by the normalized difference in means. Thus, one-to-two matching allows for a larger sample size while not increasing bias. We nonetheless test the sensitivity of our results to alternative matching approaches. In the appendices, we present descriptive statistics and results using a one-to-one without replacement approach (Appendix C), a one-to-two with replacement approach (Appendix D), and a one-to-one with replacement approach that matches exclusively on being the association system that is the nearest geographic neighbor to the state system (Appendix E). We discuss the results of the three alternative matching methods in Section 6.4.

5.3 Matching and Regression-Based Bias Correction

Although we exact match on two variables, the MDM will yield inexact covariate matches for the eight MDM variables. For the matched sample, we control for remaining covariate differences using the following model estimated with ordinary least squares:

$$Outcome_i = \beta_1 StateSystem_i + \beta_2 ln(FlowVolume_i) + \beta_3 ln(Inflows_i) + \lambda_{t(i)} + \varepsilon_i, \qquad (2)$$

where i indexes the system, $StateSystem_i$ is a binary variable equal to one if the system is a state system, and $ln(FlowVolume_i)$ and $ln(Inflows_i)$ provide the natural logarithms of two of the MDM variables, influent flow volume in gallons per minute and the number of inflow points. We also control for construction year fixed effects, $\lambda_{t(i)}$. With the year fixed effects, our coefficient of interest β_1 estimates the average difference in $Outcome_i$ across state and association systems built in the same year.[5] We estimate the equation with Huber-White heteroscedasticity robust standard errors. For the monitoring and cost-effectiveness outcomes, we estimate equation 2 as written. For the effectiveness and binary effectiveness threshold outcomes, we estimate parameters in the equation by weighting each observations using analytical weights of the number of readings that contribute to its average pH and pollutant concentrations.[6] This gives greater weight to what should be more accurate measures of water quality. For the maintenance outcomes, we additionally control for $ln(MeasurmentDuration_i)$, which is the numbers of years between the first and the last reading taken at system i.

Note that we control for the first three of our eight MDM variables in equation 2. For the other five, we directly account for differences in influent water quality by defining our continuous outcomes as the change in water quality from influent to effluent. For the non-continuous, binary outcomes, we condition each regression on failure of the influent water quality standard. For instance, for the binary indicator of effluent manganese being greater than 2 mg/l, we only include observations in the regression if the influent water is greater than 2 mg/l. This avoids the inclusion of observations in regressions for a given outcome where the system cannot improve water quality because the inflow already meets the standard. We do not estimate the equation for the TSS outcome because too few systems have influent that fails the standard to provide a sample large enough for the estimation of the model parameters.

[5] We use the "reghdfe" command in Stata to estimate equation 2. It does not report a constant, β_0. We therefore do not report a constant in the proceeding tables, which would simply be the mean outcome of the omitted build year.

[6] The exception is with the one-to-two matching with replacement sample in Appendix D. With this sample, we use frequency weights to weight each state system by one, and the 65 association systems by the number of times they are matched to a state system.

6 Results

6.1 Institutional Selection and Matching

In the full sample of 203 association systems and 61 state systems, we find limited evidence in support of the first hypothesis that the state will build systems in more complex and severe settings (Table 1). On average, the two types of systems address a similar number of inflows with a similar total flow, but the state systems appear to treat more acidic inflows. Average influent pH of state systems is 4.09 compared to 4.41 for association systems, a normalized mean difference of nearly one quarter. Moreover, 85 percent of state systems have an average influent pH less than the state standard of 6 compared to 74 percent of association systems.

Association systems are also newer on average, but one-to-two matching without replacement creates a far more comparable sample of 122 association systems (60 percent of all the association systems in our full sample). Matched association systems are still newer on average, but the normalized difference is 0.26 in the matched sample compared to 0.50 in the full sample (Table 2). Most notably, exact matching on a binary for failing the influent pH standard along with MDM on influent pH yields samples with similar influent acidity. For example, the normalized difference in average influent acidity is now only 0.01.

6.2 System Performance

To examine the performance of state and association systems, we compare means across the state and matched association systems for each of our operational definitions of effectiveness, cost-effectiveness, monitoring, and maintenance. State and association systems have similar average effectiveness over their observed lifespans and on both the continuous and threshold-based measures of water quality (Table 3).

To treat the same quantity of water, state systems cost 39 percent more than association systems, although the difference is imprecise (p-value of 0.14). For monitoring activity, the state takes many more water quality readings, which is better quantified by the proceeding biased-corrected regressions that account for differences in the amount of time for readings to occur by conditioning on the construction year fixed effects.

State and association systems also differ in their long-term effectiveness at reducing pH. The means for the state systems indicate declining effectiveness over time. The negative sign on the change in pH effectiveness indicates declining ability to improve pH over time, and the positive signs on the change in the natural logarithm of manganese, aluminum, and iron indicate increasing concentrations of metals over time.[7] Association systems, on the other hand, appear to maintain their performance at increasing pH and removing manganese over time.

6.3 Regression-Based Bias Correction

While our matched samples are similar on most observed covariates, an exception is with the year built variable (Table 2). To address the concern that we are comparing newer association systems to older state systems, we estimate regressions including construction year fixed effects, effectively comparing state and association systems built in the same year. We also control for inexact covariate matches on the complexity and size controls, and condition the effectiveness threshold regressions on failing the influent water quality standard for a given effluent outcome.

For the effectiveness outcomes, the two types of systems show similar performance at meeting state standards, raising pH, and removing metals from influent (Table 4). Taken together, the results for effectiveness in Tables 3 and 4 support our second hypothesis, which states that the two types of systems will achieve similar initial and average effectiveness in terms of increasing pH and removing heavy metals.

State systems are more costly than association systems on average, but the difference remains imprecise (Table 5). The 36 percent coefficient is a logarithmic approximation of percentage change, with the actual percentage change given by $100 \times [exp(.36) - 1] = 43\%$ (Halvorsen et al., 1980). Because the magnitude is large, we cannot rule out our third hypothesis that the asso-

[7] Of the five maintenance outcome means for state systems, two are statistically different from zero at the 95 percent level using a two-tailed t-test: the means for the change in pH and the change in aluminum. The means for association systems are not statistically different from zero using a two-tailed test.

ciation systems will be more cost-effective than state systems. This may indicate that for some systems there are inefficiencies from state contracting requirements (or efficiencies from in-kind contributions of labor by local watershed members), but that the differences are not ubiquitous.

State systems receive 77 percent ($100 \times [exp(.57) - 1] = 77\%$) more water quality readings than association systems, which supports our fourth hypothesis that state systems would be monitored more frequently due to their explicit preference for using of set-aside funds to monitor systems.

For the maintenance outcomes, association systems continue to show stronger performance relative to state systems when accounting for the bias correction (Table 6). The negative coefficient on the "state" binary in the pH regression indicates that association systems are superior to state systems at maintaining acidity reductions by nearly 1 pH scale point. The positive signs on the "state" binary in the other four regressions present imprecise evidence that, compared to association systems, state systems decline in effectiveness at removing metals and suspended solids from effluent. Appendix Figures B1 through B3 display overlapping histograms across state and association systems for the three maintenance outcomes with the strongest effects in Table 6. The figures show clear differences in the distributions of state and association systems.

Together, the mean comparisons (Table 3) and the bias-corrected regressions (Table 6) contradict our fifth hypothesis that the state will better maintain its systems. This is despite the Department of Environmental Protection's prioritizing its federal funds to operate and maintain its existing systems (Pennsylvania DEP, 2016, p.8). Our findings provide indirect evidence for the primary advantage of decentralized institutions described in the collective action literature, namely the immediate presence of local individuals to informally monitor the quality of environmental resources and intervene when necessary. This in turn ensures system effectiveness and protects water quality downstream.

6.4 Robustness

In Appendices C and D we present descriptive statistics and results using a one-to-one without replacement approach and a one-to-two with replacement matching approach. These approaches are less preferred because they achieve very modest improvements in covariate balance, but with smaller sample sizes of 61 and 84 unique association systems, which reduces statistical precision. While the three MDM approaches achieve the best balance on *observed* covariates across state and association systems, they leave open the possibility of differences in *unobserved* covariates related to geographic and community characteristics. In Appendix E, we present results using a one-to-one with replacement approach that matches exclusively on being the association system that is the nearest geographic neighbor to the state system. The approach achieves relatively strong balance on the observed covariates, but it fails to create samples of state and association systems with a similar share of inflows that fail the state standard for influent pH (Table E3). Like the full sample, the nearest neighbor matched sample suggests that there is some localized selection, with the state opting to treat slightly more acidic discharges, and provides more support for our first hypothesis that the state will build systems in more complex and severe settings.

The results in Appendices C through E are substantively similar to the one-to-two without replacement approach. Where there are differences, the biased-corrected regressions provide more precise evidence that state systems are more effective at bringing effluent above federal manganese standards (Table E5), and association systems outperform state systems at maintaining their effectiveness at removing manganese and iron (Tables D7 and E7).

7 Discussion and Policy Implications

Our main finding is that associations better maintain their systems' effectiveness over time. For insight into why, we examine associations that best maintained effectiveness. The Babb Creek Watershed Association in Tioga County is responsible for four of the best fifteen systems with above 95th percentile improvements in pH, as well as aluminum and manganese concentrations. Babb Creek built its twelve systems in our dataset in the late 1990s and early 2000s. They treat highly acidic discharges, with eleven out of twelve below the state standard of 6 pH at the first influent reading. As of the most recent effluent readings, ten of twelve are bringing pH above the state standard. Clearfield Creek Watershed Association has also successfully maintained its systems. Its two systems in our dataset that treat discharges into Little Laurel Run in Cambria County had an average initial influent pH of 3.5, and the most recent effluent readings are well above the state standard of 6 pH. Similarly, Blacklick Creek Watershed Association manages seven systems in Cambria and Indiana counties, with an average initial influent pH of 4. As of the most recent readings, all seven were bringing effluent pH above the state standard.

Although the three associations are primarily run by volunteers, they bring in between $100,000 and $600,000 in revenues in the typical year through private and state contributions (ProPublica, 2022b,c,a). While some of the contributions are small, like the sale of raffle tickets and the proceeds from festivals, fishing tournaments, and community events, others can be quite large. The three associations' financial reports show occasional sharp increases in revenues and expenditures in years where they receive large private donations or grants to build or maintain systems. For instance, Babb Creek Watershed Association received a grant of $186,000 in 2019 through the state's Environmental Stewardship and Watershed Protection Act to maintain one of its systems built in 2004 (Pennsylvania DEP, 2023b). Successful fundraising and maintenance activities of the three associations point to the advantages of local institutions identified in the collective action literature (e.g., (Ostrom, 2005)), with local individuals volunteering their time to manage natural resources, build social capital and community around the natural resource, and informally monitor environmental quality and intervene when necessary.

Another advantage lies in the ability to fund-raise quickly or apply for emergency funds, which may be quicker for repairing failing systems relative to rigid state procurement rules. For example, through the Quick Response Emergency Repair Funding program (funded by the Department of Environmental Protection and a private energy company) the Western Pennsylvania Coalition for Abandoned Mine Reclamation provides associations with money to restore failing systems. The Coalition states that in the most urgent cases, the money can arrive in a few days (Western Pennsylvania Coalitions for Abandoned Mine Reclamation, 2023). All three of the associations discussed in this section have received Quick Response funds, which helps explain their resiliency.

Despite having many well-maintained systems, the state is responsible for some systems that significantly declined in effectiveness over time. Its four systems with the worst performance at maintaining pH were built between 1998 and 2000. Two of the four have no record of ongoing maintenance in Datashed, and one has not been significantly modified since 2010. The fourth, which was built and initially effective at raising pH in 1999, was only recently maintained and updated in 2022, well after the most recent water quality reading. The state's seemingly slow response may be because of red tape surrounding procurement, or because they have over 60 systems statewide compared to the smaller number of geographically concentrated systems managed by each association. It may also be due to declining money (over our study period) in the federal Abandoned Mine Land Fund, which comes from the federal excise tax on coal production. The U.S. Department of the Interior (2023) reports that coal excise tax revenue has fallen nearly every year since the start of the shale oil and gas boom that precipitated a decline in domestic coal production. Between 2008 and 2022, the decline in coal production reduced excise tax revenues by 73 percent (U.S. Energy Information Administration, 2023; U.S. Department of the Interior, 2023).

More generally, relying on collective action institutions to address environmental problems in the U.S. may be important in light of stagnant funding for state environmental protection agencies, which are increasingly tasked with doing more with less (Cusick, 2017). The agencies are asked to implement and enforce federal and state programs to protect air, water, and land that are constantly in flux due to changing political administrations. But available funds to run programs and pay staff, like administrators, inspectors, and technical experts, appear to be decreasing. For example, the Pennsylvania Department of Environmental Protection had a total allocation of $786 million from all funding sources in 2019 (Commonwealth of Pennsylvania, 2021, p.E18-6), which is about 30 percent less in real terms than the Department's historically high allocations at the turn of the century (Commonwealth of Pennsylvania, 2002, p.E16.7).

As states and tribes prepare to spend the historic $11.3 billion earmarked in the 2021 Infrastructure Investment and Jobs Act for abandoned mine reclamation, our results suggest that the state should continue to rely on and invest in its collective action institutions to scale up reclamation efforts. While in the past, states have assumed full responsibility for spending federal funds to address abandoned mine issues (Pennsylvania DEP, 2023a), we find that watershed associations are at least as effective at treating mine discharges. The many systems that will be built with Infrastructure Act money must be maintained over their lifespans, which may be 25 years or more. The state can leverage watershed associations by investing in their organizational capacity and devolving more responsibility to construct and maintain systems. Doing so could help the money go further. We present suggestive evidence that associations are more cost-effective than the state, and the actual difference might be larger than we estimate. While state costs include direct labor, such as time spent by contractors building the systems, they exclude overhead labor costs associated with administration. By comparison, watershed associations rely in part on local volunteers. To the extent that people enjoy volunteering to improve water quality in their community, some of their labor represents a local benefit rather than a cost.

8 Conclusion

Our findings illustrate that sustained state cooperation with collective action institutions can help address complex and enduring environmental problems left by legacy hazards. State systems to treat mine drainage address slightly more polluted water than do systems managed by watershed associations. Discharge water quality, however, is similar across the two types of systems. Comparing systems that treat similar discharges, we find association systems perform at least as well as state systems. They have similar performance at cleaning discharge water initially and on average over their observed lifespans. Over time, state systems decline on average in their ability to treat the discharge, while association systems maintain their effectiveness. Suggestive evidence also indicates that association systems are more cost-effective. As such, current federal appropriations for addressing abandoned mine drainage and other hazards might go further—and maybe materialize more quickly—through increased partnerships with watershed associations or similar nonprofit partners.

References

Amirkhanyan, A. A., H. J. Kim, and K. T. Lambright (2008). Does the public sector outperform the nonprofit and for-profit sectors? Evidence from a national panel study on nursing home quality and access. *Journal of Policy Analysis and Management 27*(2), 326–353.

Ballou, D. and M. Podgursky (1998). Teacher recruitment and retention in public and private schools. *Journal of Policy Analysis and Management 17*(3), 393–417.

Banana, A. Y., N. D. Vogt, J. Bahati, and W. Gombya-Ssembajjwe (2007). Decentralized governance and ecological health: why local institutions fail to moderate deforestation in Mpigi district of Uganda. *Scientific Research and Essay 2*(10), 434–445.

Bates, R. H. (1988). Contra Contractarianism: Some Reflections on the New Institutionalism. *Politics & Society 16*(2-3), 387–401.

Born, S. M. and K. D. Genskow (2001). Toward understanding new watershed initiatives. *University of Wisconsin-Madison (en internet, http://clean-water. uwex. edu/initiatives/watershed. pdf)*.

Broms, R., C. Dahlström, and M. Nistotskaya (2023, 03). Provider Ownership and Indicators of Service Quality: Evidence from Swedish Residential Care Homes. *Journal of Public Administration Research and Theory*, muad002.

Burrows, J. E., S. C. Peters, and C. A. Cravotta (2015). Temporal geochemical variations in above- and below-drainage coal mine discharge. *Applied Geochemistry 62*, 84–95. A Special Issue to Honor the Geochemical Contributions of D. Kirk Nordstrom.

Chubb, J. E. and T. M. Moe (1988). Politics, Markets, and the Organization of Schools. *American Political Science Review 82*(4), 1065–1087.

Cleaver, F. and A. Toner (2006). The evolution of community water governance in Uchira, Tanzania: The implications for equality of access, sustainability and effectiveness. *Natural Resources Forum 30*(3), 207–218.

Code of Federal Regulations (2023). 40 CFR 434.22: Effluent limitation guidelines representing the degree of effluent reduction attainable by the application of the best practicable control technology currently available (BPT). `https://www.ecfr.gov/current/title-40/chapter-I/subchapter-N/part-434`.

Cohen, S. (2001). A Strategic Framework for Devolving Responsibility and Functions from Government to the Private Sector. *Public Administration Review 61*(4), 432–440.

Coleman, E. A. (2009). Institutional factors affecting biophysical outcomes in forest management. *Journal of Policy Analysis and Management 28*(1), 122–146.

Commonwealth of Pennsylvania (2002). Governor's Executive Budget 2002-2003. `https://www.budget.pa.gov/Publications%20and%20Reports/Documents/2002_03_Budget.pdf`.

Commonwealth of Pennsylvania (2021). Governor's Executive Buget 2021-2022. `https://www.budget.pa.gov/Publications%20and%20Reports/CommonwealthBudget/Documents/2021-22%20Proposed%20Budget/2021-2022%20Executive%20Budget%20Book.Web%20Version.UPDATED.030421.pdf`.

CRS (2020). (Congressional Research Service) The Abandoned Mine Reclamation Fund: Reauthorization Issues in the 116th Congress. `https://crsreports.congress.gov/product/pdf/R/R46266`.

Cusick, M. (2017). Cash-Strapped State Environmental Agencies Brace For Budget Cuts. *National Public Radio*.

Datashed (2023a). About Datashed. `https://www.datashed.org/help/3478110`.

Datashed (2023b). Datashed Projects. `https://www.datashed.org/`.

Dietz, T., E. Ostrom, and P. C. Stern (2003/12/12/, 2003 Dec 12). The Struggle to Govern the Commons. *Science 302*(5652), 1907–12. Copyright - Copyright American Association for the Advancement of Science Dec 12, 2003; Last updated - 2024-04-18; CODEN - SCIEAS; SubjectsTermNotLitGenreText - Social Networks; Natural Resources; Environment; Emotional Response; Social Capital; Animal Husbandry.

D'Souza, R. and H. Nagendra (2011). Changes in public commons as a consequence of urbanization: The Agara lake in Bangalore, India. *Environmental management 47*, 840–850.

Foster, S. R. and C. Iaione (2016). The City as a Commons. *Yale Law and Policy Review 34*(2), 281–349.

Frantzeskaki, N. (2019). Seven lessons for planning nature-based solutions in cities. *Environmental Science & Policy 93*, 101–111.

Ghate, R. and H. Nagendra (2005). Role of monitoring in institutional performance: forest management in Maharashtra, India. *Conservation and society 3*(2), 509–532.

Gibson, C. C., J. T. Williams, and E. Ostrom (2005). Local Enforcement and Better Forests. *World Development 33*(2), 273–284. Institutional arrangements for rural poverty reduction and resource conservation.

Glatzel, M. and B. Gordon (2018). The West's Sleeping Giant: Abandoned Mines and the role of the Good Samaritan. `https://waterinthewest.stanford.edu/news-events/news-insights/wests-sleeping-giant-abandoned-mines-and-role-good-samaritan`.

Halvorsen, R., R. Palmquist, et al. (1980). The Interpretation of Dummy Variables in Semilogarithmic Equations. *American Economic Review 70*(3), 474–475.

Hansen, E., A. Collins, S. Zegre, and A. Hereford (2010). The benefits of acid mine drainage remediation on the North Branch Potomac River. *Morgantown (WV). Downstream Strategies*.

Hansmann, H. (1996). The changing roles of public, private, and nonprofit enterprise in education, health care, and other human services. In *Individual and social responsibility: Child care, education, medical care, and long-term care in America*, pp. 245–276. University of Chicago Press.

Hardy, S. D. and T. M. Koontz (2008). Reducing nonpoint source pollution through collaboration: policies and programs across the US States. *Environmental Management 41*, 301–310.

Hayes, T. M. (2006). Parks, People, and Forest Protection: An Institutional Assessment of the Effectiveness of Protected Areas. *World Development 34*(12), 2064–2075.

Heckathorn, D. D. (1989). Collective Action and the Second-Order Free-Rider Problem. *Rationality and Society 1*(1), 78–100.

Hedin, R., T. Weaver, N. Wolfe, and G. Watzlaf (2013). Effective passive treatment of coal mine drainage. In *Proceedings of the 35th National association of abandoned mine land programs conference, Daniels, WV, USA*.

Imbens, G. W. and J. M. Wooldridge (2009, March). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature 47*(1), 5–86.

Johansen, M. and L. Zhu (2013, 06). Market Competition, Political Constraint, and Managerial Practice in Public, Nonprofit, and Private American Hospitals. *Journal of Public Administration Research and Theory 24*(1), 159–184.

Kenney, D. S. (1999). Historical and Sociolpolitcal Context of the Western Watersheds Movement. *JAWRA Journal of the American Water Resources Association 35*(3), 493–503.

King, G. and R. Nielsen (2019). Why propensity scores should not be used for matching. *Political analysis 27*(4), 435–454.

Klassen, S. and D. Evans (2020). Top-down and bottom-up water management: A diachronic model of changing water management strategies at Angkor, Cambodia. *Journal of Anthropological Archaeology 58*, 101166.

Koliba, C. J., R. M. Mills, and A. Zia (2011). Accountability in Governance Networks: An Assessment of Public, Private, and Nonprofit Emergency Management Practices Following Hurricane Katrina. *Public Administration Review 71*(2), 210–220.

Kruse Daniels, N., J. A. LaBar, and L. M. McDonald (2021). Acid mine drainage in Appalachia: sources, legacy, and treatment. *Appalachia's Coal-Mined Landscapes: Resources and Communities in a New Energy Era*, 193–216.

Lam, W. F. et al. (1998). *Governing irrigation systems in Nepal: institutions, infrastructure, and collective action.* Institute for Contemporary Studies.

Lan, Z. and H. G. Rainey (1992, 01). Goals, Rules, and Effectiveness in Public, Private, and Hybrid Organizations: More Evidence on Frequent Assertions About Differences. *Journal of Public Administration Research and Theory 2*(1), 5–28.

Lee, Y.-j. and V. M. Wilkins (2011). More Similarities or More Differences? Comparing Public and Nonprofit Managers' Job Motivations. *Public Administration Review 71*(1), 45–56.

Legere, L. (2021). Infrastrcture bill promises 'unprecdented'funding to clean up Pa.'s abandoned mine lands. *Pittsburgh Post-Gazette. November 11*.

Lubell, M., M. Schneider, J. T. Scholz, and M. Mete (2002). Watershed Partnerships and the Emergence of Collective Action Institutions. *American Journal of Political Science 46*(1), 148–163.

Marshall, G. R. (2010). Governance for a surprising world. *Resilience and transformation: Preparing Australia for uncertain futures*, 49–57.

Marshall, G. R. (2015). Polycentricity, subsidiarity and adaptive efficiency. In *Workshop on Polycentricity*. Ostrom Workshop, Indiana University.

Mathenge, J., C. N. Luwesi, C. A. Shisanya, I. Mahiri, R. A. Akombo, and M. N. Mutiso (2014). Community participation in water sector governance in Kenya: A performance based appraisal of community water management systems in Ngaciuma-Kinyaritha catchment, Tana basin, Mount Kenya region. *International Journal of Innovative Research and Development 3*(5), 783–92.

McCord, P., J. Dell'Angelo, E. Baldwin, and T. Evans (2017). Polycentric Transformation in Kenyan Water Governance: A Dynamic Analysis of Institutional and Social-Ecological Change. *Policy Studies Journal 45*(4), 633–658.

Merritt, P. and C. Power (2022). Assessing the long-term evolution of mine water quality in abandoned underground mine workings using first-flush based models. *Science of The Total Environment 846*, 157390.

North, D. C. (1990). *Institutions, Institutional Change, and Economic Performance.* Cambridge University Press.

North, D. C. (1994). Economic Performance Through Time. *The American Economic Review 84*(3), 359–368.

Office of Surface Mining Reclamation and Enforcement (2023). Acid Mine Drainage Set-Aside Program. `https://www.osmre.gov/sites/default/files/pdfs/4-130.pdf`.

Ostrom, E. (1965). *Public entrepreneurship: a case study in ground water basin management*. University of California, Los Angeles.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge university press.

Ostrom, E. (1999). Institutional rational choice: an assessment of the institutional analysis and development framework theories of the policy process. *Theories of the Policy Process. Boulder, Colarado: Westview*, 35–72.

Ostrom, E. (2005). *Understanding Institutional Diversity*. Princeton University Press.

Ostrom, E. (2009). A general framework for analyzing sustainability of social-ecological systems. *Science 325*(5939), 419–422.

Ostrom, E. and H. Nagendra (2006). Insights on linking forests, trees, and people from the air, on the ground, and in the laboratory. *Proceedings of the National Academy of Sciences 103*(51), 19224–19231.

Ostrom, E., R. B. Parks, and G. P. Whitaker (1973). Do We Really Want to Consolidate Urban Police Forces? A Reappraisal of Some Old Assertions. *Public Administration Review 33*(5), 423–432.

Ostrom, E. and G. Whitaker (1973). Does Local Community Control of Police make a Difference? Some Preliminary Findings. *American Journal of Political Science 17*(1), 48–76.

Pahl-Wostl, C., L. Lebel, C. Knieper, and E. Nikitina (2012). From applying panaceas to mastering complexity: Toward adaptive water governance in river basins. *Environmental Science & Policy 23*, 24–34.

Pennsylvania DEP (2016). (Department of Environmental Protection) Acid Mine Drainage Set-Aside Program Program Implementation Guidelines. `http://www.depgreenport.state.pa.us/elibrary/GetDocument?docId=8133&DocName=ACID%20MINE%20DRAINAGE%20SET-ASIDE%20PROGRAM%20IMPLEMENTATION%20GUIDELINES.PDF%20`.

Pennsylvania DEP (2019). Passive Abandoned Mine Drainage (AMD) Treatment Systems Constructed by PA-DEP-Bureau of Abandoned Mine Reclamation.

Pennsylvania DEP (2022). (Department of Environmental Protection) Integrated Water Quality Report – 2022. https://www.dep.pa.gov/Business/Water/CleanWater/WaterQuality/IntegratedWatersReport/Pages/2022-Integrated-Water-Quality-Report.aspx.

Pennsylvania DEP (2023a). (Department of Environmental Protection) Acid Mine Drainage (AMD) Program.

Pennsylvania DEP (2023b). Investment Tracker. http://www.dced.state.pa.us/InvestmentTracker/DefaultDEP.aspx.

Pennsylvania General Assembly (1999). Environmental Stewardship and Watershed Protection Act. https://www.legis.state.pa.us/cfdocs/legis/LI/consCheck.cfm?txtType=HTM&ttl=27&div=0&chpt=61.

ProPublica (2022a). Babb Creek Watershed Association Inc. https://projects.propublica.org/nonprofits/organizations/232995774.

ProPublica (2022b). Blacklick Creek Watershed Association Inc. https://projects.propublica.org/nonprofits/organizations/251763928.

ProPublica (2022c). Clearfield Creek Watershed Association. https://projects.propublica.org/nonprofits/organizations/251883781.

Rainey, H. G. and B. Bozeman (2000, 04). Comparing Public and Private Organizations: Empirical Research and the Power of the A Priori. *Journal of Public Administration Research and Theory 10*(2), 447–470.

Rainey, H. G. and Y. H. Chun (2007, 06). 71 Public and Private Management Compared. In *The Oxford Handbook of Public Management*. Oxford University Press.

Ripollone, J. E., K. F. Huybrechts, K. J. Rothman, R. E. Ferguson, and J. M. Franklin (2018, 05). Implications of the Propensity Score Matching Paradox in Pharmacoepidemiology. *American Journal of Epidemiology 187*(9), 1951–1961.

Rushing, W. (1974). Differences in Profit and Nonprofit Organizations: A Study of Effectiveness and Efficiency in General Short-Stay Hospitals. *Administrative Science Quarterly 19*(4), 474–484.

Salamon, L. M. (1987). *Partners in Public service: The scope and theory of government nonprofit relations. In W. W. Powell (Ed.), The nonprofit sector: A research handbook*. New Haven, CT: Yale University Press.

Shivakoti, G., D. Vermillion, W.-F. Lam, E. Ostrom, U. Pradhan, and R. Yoder (2005). *Asian irrigation in transition: Responding to challenges*. SAGE Publishing India.

Sobel, R. S. and P. T. Leeson (2006). Government's response to Hurricane Katrina: A public choice analysis. *Public Choice 127*, 55–73.

Tarlock, D. (2000). Putting rivers back in the landscape: The revival of watershed management in the United States. *UC Law Environmental Journal 6*(2), 167.

U.S. Department of the Interior (2003). Technical Note 409: Passive Treatment Systems for Acid Mine Drainage. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1018&context=usblmpub#:~:text=PASSIVE%20TREATMENT%20SYSTEMS%20provide%20a,Department%20of%20Environmental%20Protection%201999).

U.S. Department of the Interior (2022). Interior Department Announces $122.5 Million for Abandoned Mine Land Economic Revitalization Grants. www.doi.gov/pressreleases/interior-department-announces-1225-million-abandoned-mine-land-economic-revitalization.

U.S. Department of the Interior (2023). Coal Excise Tax. https://revenuedata.doi.gov/how-revenue-works/coal-excise-tax/.

U.S. Energy Information Administration (2023). Coal Data Broswer: Aggregate Coal Mine Production. https://www.eia.gov/coal/data/browser/.

U.S. Environmental Protection Agency (2017). Surf Your Watershed. https://19january2017snapshot.epa.gov/waterdata/surf-your-watershed_.html.

Wade, R. (1988). *Village Republics: Economic Conditions for Collective Action in South India*. Cambridge: Cambridge University Press.

Western Pennsylvania Coalitions for Abandoned Mine Reclamation (2023). Quick Response for Funding Emergency Repairs. http://www.wpcamr.org/projects/QuickResponse/index.html.

Westmoreland Conservation District (2023). Watershed Associations. http://westmorelandconservation.org/watershed-associations/.

Note: Geographic data are from Datashed water quality reports and project documents.

Figure 1: Locations of State and Watershed Association Treatment Systems in Pennsylvania

Table 1: Comparison of State and Association Systems

	Association System (N=203)	State System (N=61)	Diff.	Normalized Diff.	P-Value
Characteristics	.	.	.	.	.
Year Built	2,004.53	2,001.84	2.70	0.50	0.00
Number of Funding Sources	1.32	1.52	-0.20	-0.31	0.05
Severity	.	.	.	.	.
Initial Influent pH	4.51	4.33	0.19	0.14	0.40
Avg Influent pH	4.41	4.09	0.31	0.24	0.10
Influent Manganese (mg/l)	8.13	8.88	-0.75	-0.04	0.79
Influent Aluminum (mg/l)	11.81	11.45	0.36	0.01	0.93
Influent Iron (mg/l)	22.04	29.72	-7.68	-0.20	0.20
Influent TSS (mg/l)	24.67	16.43	8.25	0.08	0.43
Severity Threshold	.	.	.	.	.
Any Influent Standard Failed	0.99	1.00	-0.01	-0.17	0.08
Influent pH ≤ 6	0.73	0.78	-0.05	-0.12	0.44
Avg Influent pH ≤ 6	0.74	0.85	-0.11	-0.27	0.05
Influent Mang. ≥ 2 (mg/l)	0.65	0.61	0.04	0.08	0.59
Influent Aluminum $\geq .5$ (mg/l)	0.74	0.82	-0.08	-0.19	0.17
Influent Iron ≥ 1.5 (mg/l)	0.77	0.80	-0.03	-0.08	0.56
Influent TSS ≥ 35 (mg/l)	0.06	0.08	-0.02	-0.09	0.56
Complexity	.	.	.	.	.
Number of Inflows	1.33	1.36	-0.03	-0.04	0.79
Size	.	.	.	.	.
Inflow Volume (gal/min)	290.64	259.66	30.98	0.04	0.74

Note: Normalized mean differences are calculated using the method of Imbens and Wooldridge (2009), which is the difference in means divided by the square root of the average variance. P-Values are two-tailed and from a t-test for the difference in means of unpaired data, assuming unequal variances. The Severity Threshold variables are all binary variables.

Table 2: Comparison of One-to-Two Without Replacement MDM Matched State and Association Systems

	Association System (N=122)	State System (N=61)	Diff.	Normalized Diff.	P-Value
Characteristics	.	.	.	.	.
Year Built	2,003.02	2,001.84	1.19	0.26	0.08
Number of Funding Sources	1.40	1.52	-0.12	-0.18	0.27
Severity	.	.	.	.	.
Initial Influent pH	4.23	4.33	-0.10	-0.08	0.66
Avg Influent pH	4.11	4.09	0.02	0.01	0.93
Influent Manganese (mg/l)	7.27	8.88	-1.62	-0.09	0.59
Influent Aluminum (mg/l)	10.27	11.45	-1.18	-0.07	0.66
Influent Iron (mg/l)	21.66	29.72	-8.06	-0.21	0.20
Influent TSS (mg/l)	10.65	16.43	-5.78	-0.29	0.10
Severity Threshold	.	.	.	.	.
Any Influent Standard Failed	0.99	1.00	-0.01	-0.13	0.32
Initial Influent pH ≤ 6	0.81	0.78	0.03	0.07	0.67
Avg Influent pH ≤ 6	0.85	0.85	0.00	0.00	1.00
Influent Mang. ≥ 2 (mg/l)	0.64	0.61	0.03	0.07	0.67
Influent Aluminum $\geq .5$ (mg/l)	0.80	0.82	-0.02	-0.04	0.79
Influent Iron ≥ 1.5 (mg/l)	0.76	0.80	-0.04	-0.10	0.52
Influent TSS ≥ 35 (mg/l)	0.03	0.08	-0.05	-0.21	0.21
Complexity	.	.	.	.	.
Number of Inflows	1.32	1.36	-0.04	-0.06	0.69
Size	.	.	.	.	.
Inflow Volume (gal/min)	253.94	259.66	-5.72	-0.01	0.95

Note: Normalized mean differences are calculated using the method of Imbens and Wooldridge (2009), which is the difference in means divided by the square root of the average variance. P-Values are two-tailed and from a t-test for the difference in means of unpaired data, assuming unequal variances. The Severity Threshold variables are all binary variables.

Table 3: Mean Differences: One-to-Two Without Replacement MDM Matched State and Association Systems

	Association System (N=122)	State System (N=61)	Diff.	P-Value
Effectiveness	.	.	.	.
Initial Change in pH	2.07	2.41	-0.34	0.21
Avg Change in pH	1.62	1.62	0.01	0.97
Δ Ln Manganese (mg/l)	-0.59	-0.59	0.01	0.96
Δ Ln Aluminum (mg/l)	-1.53	-1.30	-0.23	0.29
Δ Ln Iron (mg/l)	-1.65	-1.75	0.10	0.67
Δ Ln TSS (mg/l)	0.01	0.05	-0.04	0.78
Effectiveness Threshold	.	.	.	.
Any Effluent Standard Failed	0.82	0.85	-0.03	0.57
Initial Effluent pH ≤ 6	0.25	0.18	0.07	0.32
Avg Effluent pH ≤ 6	0.47	0.49	-0.02	0.76
Effluent Mang. ≥ 2 (mg/l)	0.45	0.49	-0.04	0.60
Effluent Aluminum $\geq .5$ (mg/l)	0.54	0.61	-0.07	0.40
Effluent Iron ≥ 1.5 (mg/l)	0.41	0.48	-0.07	0.40
Effluent TSS ≥ 35 (mg/l)	0.02	0.11	-0.10	0.02
Cost Effectiveness	.	.	.	.
Ln Cost 100k Gallons Treated	3.75	4.14	-0.39	0.14
Monitoring	.	.	.	.
Ln Number of Readings	3.26	4.14	-0.88	0.00
Maintenance	.	.	.	.
Change in pH Effectiveness	0.00	-0.70	0.71	0.02
Change in Δ Ln Manganese	-0.20	0.05	-0.25	0.34
Change in Δ Ln Aluminum	0.22	0.53	-0.31	0.32
Change in Δ Ln Iron	0.31	0.26	0.05	0.88
Change in Δ Ln TSS	-0.19	-0.04	-0.16	0.57

Note: P-Values are two-tailed and from a t-test for the difference in means of unpaired data, assuming unequal variances. The Effectiveness Threshold variables are all binary variables.

Table 4: Biased Corrected Effects: Effectiveness, One-to-Two Without Replacement

	Any	Initial pH		pH		Mn		Al		Fe	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	p(Fail)	Δ	p(Fail)	Δ	p(Fail)	Δ Ln	p(Fail)	Δ Ln	p(Fail)	Δ Ln	p(Fail)
State System	0.02	0.27	-0.07	0.01	0.03	-0.05	-0.01	0.23	0.06	-0.14	0.13
	(0.06)	(0.27)	(0.10)	(0.24)	(0.10)	(0.13)	(0.11)	(0.25)	(0.09)	(0.30)	(0.10)
Ln Influent Flow (gal/min)	0.01	-0.04	0.00	-0.09	0.02	0.06	-0.01	0.13*	0.03	0.20**	0.04
	(0.02)	(0.09)	(0.02)	(0.07)	(0.03)	(0.04)	(0.04)	(0.06)	(0.03)	(0.07)	(0.03)
Ln Number of Inflows	0.03	-0.27	0.13	0.04	0.07	-0.13	-0.00	-0.21	-0.03	-0.41	-0.08
	(0.07)	(0.33)	(0.12)	(0.33)	(0.13)	(0.19)	(0.16)	(0.33)	(0.12)	(0.28)	(0.12)
R-squared	0.10	0.13	0.12	0.09	0.07	0.10	0.09	0.13	0.10	0.10	0.16
N	168	148	116	169	141	168	105	168	134	168	129
Cond. on Failed Influent Std.	Y	N	Y	N	Y	N	Y	N	Y	N	Y

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction. Each system in the regressions is weighted using analytical weights of the number of readings that contribute its average pH and pollutant concentrations.

Table 5: Biased Corrected Effects: Cost-Effectiveness and Monitoring, One-to-Two Without Replacement

	Ln(Cost Per 100k Gallons)	Ln(Number of Readings)
	(1)	(2)
State System	0.36	0.57**
	(0.28)	(0.18)
Ln Number of Inflows	0.36	0.21
	(0.34)	(0.19)
Ln Influent Flow (gal/min)		0.08
		(0.06)
R-squared	0.19	0.34
N	169	169

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction. We do not control for influent flow in gallons per minute in column 1, because it is directly used to estimate the demoninator of the cost per 100,000 gallons outcome.

Table 6: Biased Corrected Effects: Maintenance, One-to-Two Without Replacement

	Change in Δ pH	Change in Δ Ln			
	(1)	(2)	(3)	(4)	(5)
	pH	Mn	Al	Fe	TSS
State System	-0.80**	0.31	0.29	0.09	0.20
	(0.31)	(0.33)	(0.33)	(0.38)	(0.34)
Ln Influent Flow (gal/min)	-0.08	-0.03	0.03	0.14	0.14
	(0.09)	(0.09)	(0.11)	(0.09)	(0.09)
Ln Number of Inflows	0.08	0.61	-0.44	0.67	-0.10
	(0.38)	(0.35)	(0.44)	(0.39)	(0.32)
Ln Measurement Duration	-0.31	-0.63	-0.11	-0.67	-0.29
	(0.21)	(0.35)	(0.29)	(0.37)	(0.31)
R-squared	0.16	0.18	0.11	0.16	0.12
N	148	132	122	132	122

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction.

Appendices For Online Publication

A Data Notes

A.1 Classifying State Systems

We begin with water quality readings at 264 passive systems in Datashed. To classify each system as constructed and maintained by the state or an association, we utilize three variable categories in Datashed. First, each system has a "Project Name" field. We compare the project names to a list of state-constructed passive systems published by Pennsylvania DEP (2019), and classify each system on the list as a state system. Second, each system has fields for its "Responsible Organization" and "Contact Organization." We classify any system where the Department is the responsible organization or contact organization as a state system. Third, each system contains data on the funding sources used to construct the system. We classify any system constructed with federal set-aside money as a state system. We use all three approaches, rather than just one, because some state systems are missing from the Pennsylvania DEP (2019) list, and because there is missingness in the responsible organization, contact organization, and funding fields. We manually checked the individual system web-pages on Datashed, which contain project documents, to verify that each system is properly classified. Altogether, we classify 61 systems as state constructed and maintained and the other 203 as association systems.

B Additional Tables and Figures

Table B1: Treatment System Descriptive Statistics

	Mean	SD	Min.	p25	p50	p75	Max	N
Characteristics	.	.	.	.	.	.	.	.
State System	0.23	0.42	0.00	0.00	0.00	0.00	1.00	264.00
Year Built	2,003.91	6.11	1,988.00	2,000.00	2,004.00	2,008.00	2,021.00	264.00
Number of Funding Sources	1.37	0.61	1.00	1.00	1.00	2.00	4.00	264.00
Severity	.	.	.	.	.	.	.	.
Initial Influent pH	4.47	1.38	2.50	3.30	3.90	6.05	7.65	223.00
Avg Influent pH	4.34	1.34	1.96	3.27	3.82	5.70	7.79	264.00
Influent Manganese (mg/l)	8.30	16.21	0.00	1.46	2.87	7.28	154.02	264.00
Influent Aluminum (mg/l)	11.73	46.00	0.00	0.52	3.70	11.36	724.52	264.00
Influent Iron (mg/l)	23.82	34.55	0.00	2.10	8.88	28.92	217.09	264.00
Influent TSS (mg/l)	22.77	123.14	0.00	5.33	7.91	14.76	1,944.50	264.00
Severity Threshold	.	.	.	.	.	.	.	.
Any Influent Standard Failed	0.99	0.11	0.00	1.00	1.00	1.00	1.00	264.00
Initial Influent pH ≤ 6	0.74	0.44	0.00	0.00	1.00	1.00	1.00	223.00
Avg Influent pH ≤ 6	0.77	0.42	0.00	1.00	1.00	1.00	1.00	264.00
Influent Mang. ≥ 2 (mg/l)	0.64	0.48	0.00	0.00	1.00	1.00	1.00	264.00
Influent Aluminum $\geq .5$ (mg/l)	0.76	0.43	0.00	1.00	1.00	1.00	1.00	264.00
Influent Iron ≥ 1.5 (mg/l)	0.78	0.42	0.00	1.00	1.00	1.00	1.00	264.00
Influent TSS ≥ 35 (mg/l)	0.06	0.25	0.00	0.00	0.00	0.00	1.00	264.00
Complexity	.	.	.	.	.	.	.	.
Number of Inflows	1.34	0.70	1.00	1.00	1.00	1.00	5.00	264.00
Size	.	.	.	.	.	.	.	.
Inflow Volume (gal/min)	283.48	767.40	0.00	17.81	55.57	177.50	7,239.38	264.00

Note: Data are from Datashed water quality reports and project documents.

Table B2: Treatment System Descriptive Statistics, Outcome Variables

	Mean	SD	Min.	p25	p50	p75	Max	N
Effectiveness	.	.	.	.	.	.	.	.
Initial Change in pH	2.08	1.60	-2.85	0.65	2.30	3.46	4.93	223.00
Avg Change in pH	1.59	1.35	-0.87	0.49	1.29	2.70	4.54	264.00
Δ Ln Manganese (mg/l)	-0.61	0.92	-6.71	-0.88	-0.38	-0.06	1.39	262.00
Δ Ln Aluminum (mg/l)	-1.45	1.55	-6.28	-2.36	-1.09	-0.17	1.37	263.00
Δ Ln Iron (mg/l)	-1.65	1.58	-6.72	-2.60	-1.40	-0.60	3.43	263.00
Δ Ln TSS (mg/l)	-0.05	0.98	-4.45	-0.48	0.00	0.42	4.36	262.00
Effectiveness Threshold	.	.	.	.	.	.	.	.
Any Effluent Standard Failed	0.82	0.38	0.00	1.00	1.00	1.00	1.00	264.00
Initial Effluent pH $\leq$ 6	0.17	0.38	0.00	0.00	0.00	0.00	1.00	223.00
Avg Effluent pH $\leq$ 6	0.40	0.49	0.00	0.00	0.00	1.00	1.00	264.00
Effluent Mang. $\geq$ 2 (mg/l)	0.48	0.50	0.00	0.00	0.00	1.00	1.00	264.00
Effluent Aluminum $\geq$ .5 (mg/l)	0.50	0.50	0.00	0.00	0.00	1.00	1.00	264.00
Effluent Iron $\geq$ 1.5 (mg/l)	0.44	0.50	0.00	0.00	0.00	1.00	1.00	264.00
Effluent TSS $\geq$ 35 (mg/l)	0.05	0.22	0.00	0.00	0.00	0.00	1.00	264.00
Cost Effectiveness	.	.	.	.	.	.	.	.
Cost Per 100k Gallons Treated	206.37	857.56	0.00	20.26	54.65	146.19	10,477.85	255.00
Monitoring	.	.	.	.	.	.	.	.
Number of Readings	53.30	54.42	2.00	11.50	36.00	79.00	361.00	264.00
Maintenance	.	.	.	.	.	.	.	.
Change in pH Effectiveness	-0.21	1.41	-5.60	-0.72	-0.11	0.40	4.45	223.00
Change in Δ Ln Manganese	-0.12	1.45	-5.96	-0.75	-0.08	0.45	5.95	201.00
Change in Δ Ln Aluminum	0.28	1.45	-3.98	-0.51	0.09	0.99	5.58	191.00
Change in Δ Ln Iron	0.24	1.62	-4.40	-0.84	0.23	1.18	5.76	199.00
Change in Δ Ln TSS	-0.09	1.40	-4.71	-0.75	0.01	0.64	4.41	180.00

Note: Data are from Datashed water quality reports and project documents.

Table B3: Matching Variables

Variable	Description
Exact Matching	
Influent pH ≤ 6	Mean influent pH is less than state standard of 6
Missingness of time varying variables	Binary variable indicating one of the 223 systems with more than one sample
Mahalanobis Distance Matching (MDM)	
Year Built	Year the system was built
Number of Inflows	Number of inflows into the system
Inflow Volume	Total inflow into the system in gallons per minute
Influent pH	Mean influent pH on the pH scale
Influent Manganese (mg/l)	Mean influent Manganese in mg/l
Influent Aluminum (mg/l)	Mean influent Aluminum in mg/l
Influent Iron (mg/l)	Mean influent Iron in mg/l
Influent TSS (mg/l)	Mean influent Suspended Solids in mg/l

Table B4: Descriptive Statistics: One-to-Two Without Replacement MDM Sample

	Mean	SD	Min.	p25	p50	p75	Max	N
Characteristics	.	.	.	.	.	.	.	.
State System	0.33	0.47	0.00	0.00	0.00	1.00	1.00	183.00
Year Built	2,002.63	4.82	1,988.00	1,999.00	2,003.00	2,005.00	2,019.00	183.00
Number of Funding Sources	1.44	0.67	1.00	1.00	1.00	2.00	4.00	183.00
Severity	.	.	.	.	.	.	.	.
Initial Influent pH	4.26	1.32	2.50	3.20	3.72	5.36	7.21	153.00
Avg Influent pH	4.11	1.23	1.96	3.20	3.65	4.73	6.94	183.00
Influent Manganese (mg/l)	7.80	16.73	0.00	1.41	2.87	7.25	154.02	183.00
Influent Aluminum (mg/l)	10.66	14.84	0.00	0.75	5.78	13.03	103.13	183.00
Influent Iron (mg/l)	24.35	37.06	0.00	2.19	8.25	28.37	217.09	183.00
Influent TSS (mg/l)	12.57	17.34	0.00	5.57	7.60	13.67	171.49	183.00
Severity Threshold	.	.	.	.	.	.	.	.
Any Influent Standard Failed	0.99	0.07	0.00	1.00	1.00	1.00	1.00	183.00
Initial Influent pH ≤ 6	0.80	0.40	0.00	1.00	1.00	1.00	1.00	153.00
Avg Influent pH ≤ 6	0.85	0.36	0.00	1.00	1.00	1.00	1.00	183.00
Influent Mang. ≥ 2 (mg/l)	0.63	0.48	0.00	0.00	1.00	1.00	1.00	183.00
Influent Aluminum $\geq .5$ (mg/l)	0.81	0.39	0.00	1.00	1.00	1.00	1.00	183.00
Influent Iron ≥ 1.5 (mg/l)	0.78	0.42	0.00	1.00	1.00	1.00	1.00	183.00
Influent TSS ≥ 35 (mg/l)	0.05	0.22	0.00	0.00	0.00	0.00	1.00	183.00
Complexity	.	.	.	.	.	.	.	.
Number of Inflows	1.33	0.67	1.00	1.00	1.00	1.00	4.00	183.00
Size	.	.	.	.	.	.	.	.
Inflow Volume (gal/min)	255.85	641.64	0.00	17.00	54.47	180.00	5,637.25	183.00

Note: Data are from Datashed water quality reports and project documents.

Table B5: Descriptive Statistics: One-to-Two Without Replacement MDM Sample, Outcome Variables

	Mean	SD	Min.	p25	p50	p75	Max	N
Effectiveness	.	.	.	.	.	.	.	.
Initial Change in pH	2.18	1.59	-2.85	0.82	2.40	3.46	4.93	153.00
Avg Change in pH	1.62	1.32	-0.87	0.58	1.33	2.70	4.53	183.00
Δ Ln Manganese (mg/l)	-0.59	0.91	-6.71	-0.88	-0.34	-0.06	1.14	181.00
Δ Ln Aluminum (mg/l)	-1.46	1.46	-6.23	-2.29	-1.16	-0.27	1.06	182.00
Δ Ln Iron (mg/l)	-1.68	1.60	-6.72	-2.64	-1.45	-0.63	3.43	182.00
Δ Ln TSS (mg/l)	0.02	0.90	-2.68	-0.42	0.00	0.43	4.36	182.00
Effectiveness Threshold	.	.	.	.	.	.	.	.
Any Effluent Standard Failed	0.83	0.38	0.00	1.00	1.00	1.00	1.00	183.00
Initial Effluent pH ≤ 6	0.22	0.42	0.00	0.00	0.00	0.00	1.00	153.00
Avg Effluent pH ≤ 6	0.48	0.50	0.00	0.00	0.00	1.00	1.00	183.00
Effluent Mang. ≥ 2 (mg/l)	0.46	0.50	0.00	0.00	0.00	1.00	1.00	183.00
Effluent Aluminum $\geq .5$ (mg/l)	0.56	0.50	0.00	0.00	1.00	1.00	1.00	183.00
Effluent Iron ≥ 1.5 (mg/l)	0.43	0.50	0.00	0.00	0.00	1.00	1.00	183.00
Effluent TSS ≥ 35 (mg/l)	0.05	0.22	0.00	0.00	0.00	0.00	1.00	183.00
Cost Effectiveness	.	.	.	.	.	.	.	.
Cost Per 100k Gallons Treated	196.90	670.80	0.00	21.90	65.76	149.26	7,268.57	176.00
Monitoring	.	.	.	.	.	.	.	.
Number of Readings	59.46	53.53	2.00	14.00	52.00	87.00	361.00	183.00
Maintenance	.	.	.	.	.	.	.	.
Change in pH Effectiveness	-0.23	1.59	-5.60	-0.78	-0.10	0.44	4.45	153.00
Change in Δ Ln Manganese	-0.12	1.56	-5.96	-0.82	-0.09	0.61	5.95	137.00
Change in Δ Ln Aluminum	0.32	1.52	-3.98	-0.52	0.09	1.20	4.77	128.00
Change in Δ Ln Iron	0.29	1.69	-4.40	-0.81	0.35	1.26	5.76	137.00
Change in Δ Ln TSS	-0.15	1.40	-4.71	-0.76	-0.00	0.63	4.41	127.00

Note: Data are from Datashed water quality reports and project documents.

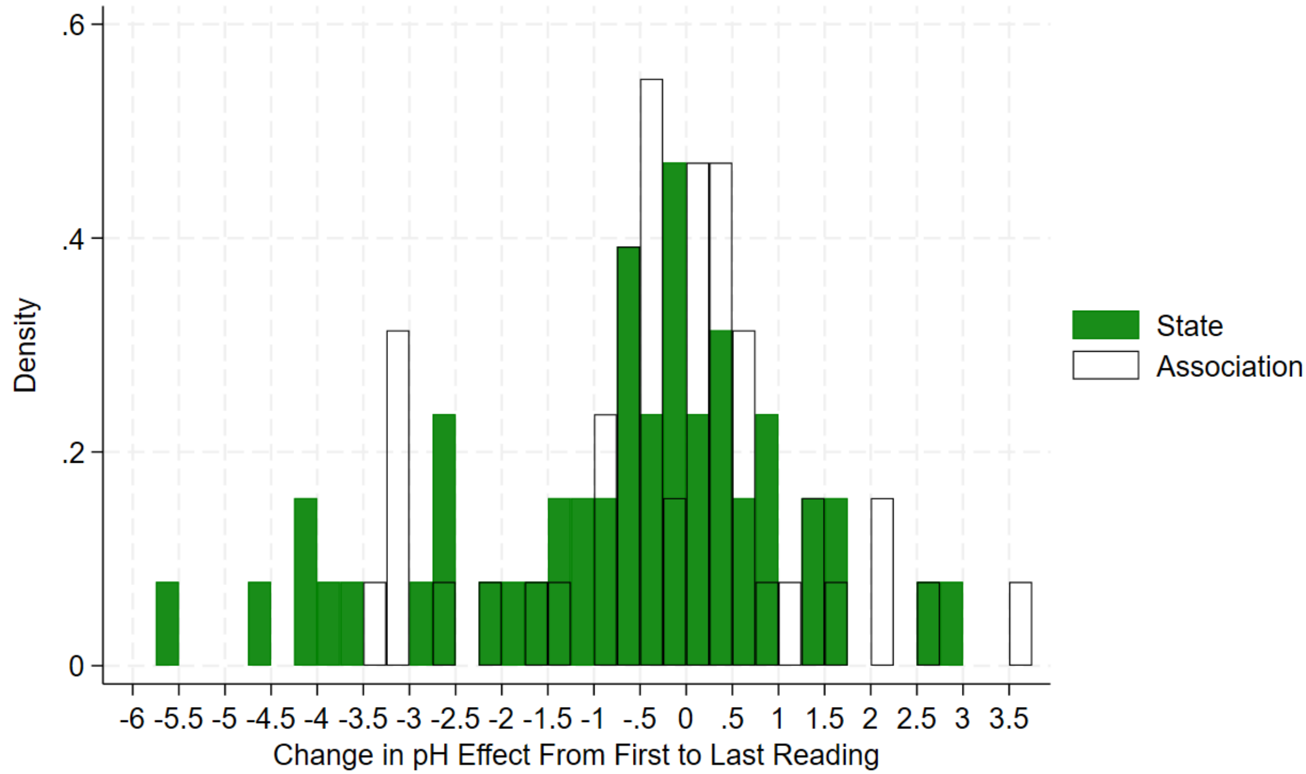

Figure B1: Histograms of State and Non-State Systems Effect on pH Over Time

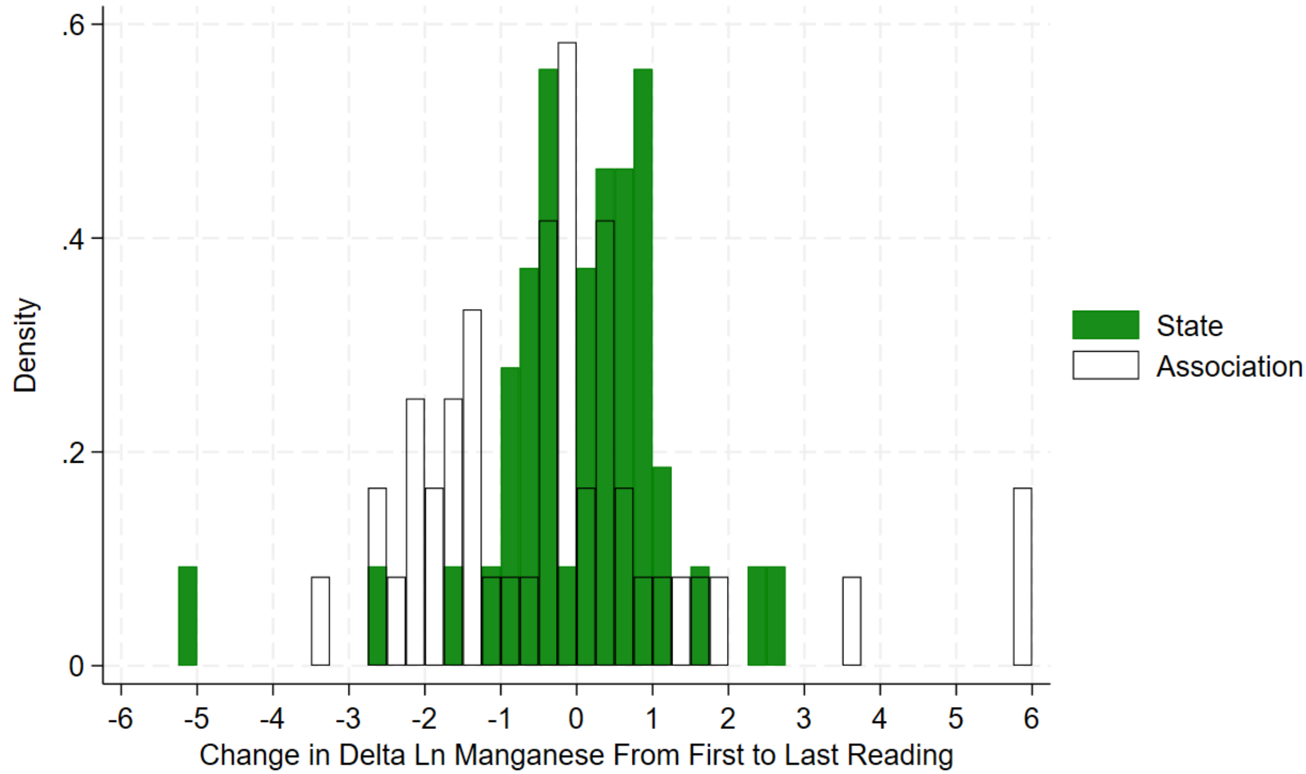

Figure B2: Histograms of State and Non-State Systems Effect on Manganese Over Time

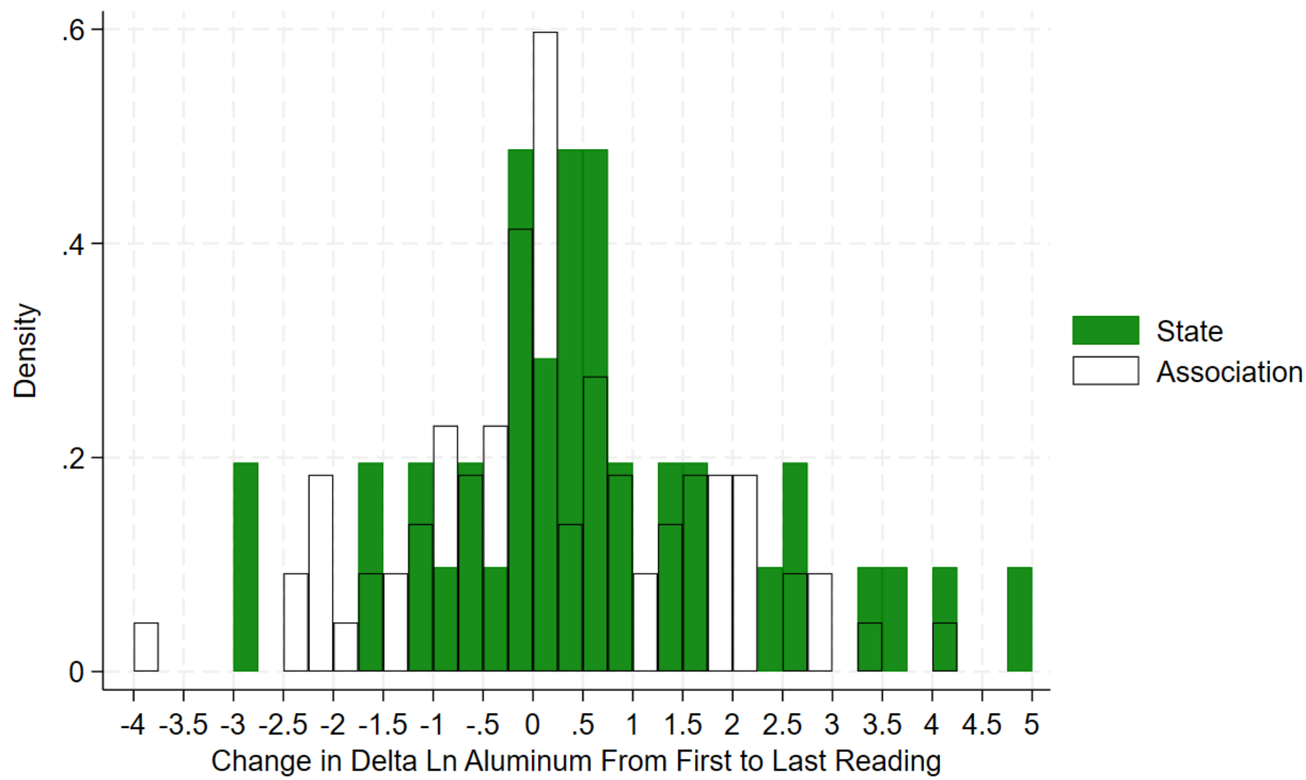

Figure B3: Histograms of State and Non-State Systems Effect on Aluminum Over Time

C Additional Tables: One-to-One Matching

Table C1: Descriptive Statistics: One-to-One Without Replacement MDM Sample

	Mean	SD	Min.	p25	p50	p75	Max	N
Characteristics	.	.	.	.	.	.	.	.
State System	0.50	0.50	0.00	0.00	0.50	1.00	1.00	122.00
Year Built	2,002.41	4.25	1,990.00	1,999.00	2,002.00	2,005.00	2,019.00	122.00
Number of Funding Sources	1.48	0.71	1.00	1.00	1.00	2.00	4.00	122.00
Severity	.	.	.	.	.	.	.	.
Initial Influent pH	4.24	1.33	2.50	3.20	3.72	5.21	7.00	102.00
Avg Influent pH	4.10	1.24	1.96	3.20	3.69	4.61	6.94	122.00
Influent Manganese (mg/l)	7.82	17.75	0.00	1.50	2.92	7.23	154.02	122.00
Influent Aluminum (mg/l)	10.96	15.79	0.00	0.73	5.41	12.95	103.13	122.00
Influent Iron (mg/l)	27.15	37.82	0.00	2.84	13.31	35.86	217.09	122.00
Influent TSS (mg/l)	14.17	19.83	0.00	5.72	8.25	14.56	171.49	122.00
Severity Threshold	.	.	.	.	.	.	.	.
Any Influent Standard Failed	0.99	0.09	0.00	1.00	1.00	1.00	1.00	122.00
Initial Influent pH ≤ 6	0.79	0.41	0.00	1.00	1.00	1.00	1.00	102.00
Avg Influent pH ≤ 6	0.85	0.36	0.00	1.00	1.00	1.00	1.00	122.00
Influent Mang. ≥ 2 (mg/l)	0.63	0.48	0.00	0.00	1.00	1.00	1.00	122.00
Influent Aluminum $\geq .5$ (mg/l)	0.81	0.39	0.00	1.00	1.00	1.00	1.00	122.00
Influent Iron ≥ 1.5 (mg/l)	0.80	0.40	0.00	1.00	1.00	1.00	1.00	122.00
Influent TSS ≥ 35 (mg/l)	0.06	0.23	0.00	0.00	0.00	0.00	1.00	122.00
Complexity	.	.	.	.	.	.	.	.
Number of Inflows	1.34	0.66	1.00	1.00	1.00	1.00	4.00	122.00
Size	.	.	.	.	.	.	.	.
Inflow Volume (gal/min)	249.62	572.36	0.00	17.80	61.79	180.00	3,832.22	122.00

Note: Data are from Datashed water quality reports and project documents.

Table C2: Descriptive Statistics: One-to-One Without Replacement MDM Sample, Outcome Variables

	Mean	SD	Min.	p25	p50	p75	Max	N
Effectiveness	.	.	.	.	.	.	.	.
Initial Change in pH	2.41	1.50	-0.97	1.06	2.80	3.60	4.93	102.00
Avg Change in pH	1.66	1.30	-0.64	0.61	1.35	2.70	4.53	122.00
Δ Ln Manganese (mg/l)	-0.55	0.74	-2.90	-0.94	-0.37	-0.05	1.14	120.00
Δ Ln Aluminum (mg/l)	-1.49	1.52	-6.23	-2.36	-1.19	-0.27	1.06	121.00
Δ Ln Iron (mg/l)	-1.76	1.50	-5.72	-2.81	-1.59	-0.75	3.17	121.00
Δ Ln TSS (mg/l)	0.02	0.99	-2.68	-0.48	0.00	0.62	4.36	121.00
Effectiveness Threshold	.	.	.	.	.	.	.	.
Any Effluent Standard Failed	0.80	0.40	0.00	1.00	1.00	1.00	1.00	122.00
Initial Effluent pH ≤ 6	0.15	0.36	0.00	0.00	0.00	0.00	1.00	102.00
Avg Effluent pH ≤ 6	0.46	0.50	0.00	0.00	0.00	1.00	1.00	122.00
Effluent Mang. ≥ 2 (mg/l)	0.47	0.50	0.00	0.00	0.00	1.00	1.00	122.00
Effluent Aluminum $\geq .5$ (mg/l)	0.55	0.50	0.00	0.00	1.00	1.00	1.00	122.00
Effluent Iron ≥ 1.5 (mg/l)	0.48	0.50	0.00	0.00	0.00	1.00	1.00	122.00
Effluent TSS ≥ 35 (mg/l)	0.07	0.25	0.00	0.00	0.00	0.00	1.00	122.00
Cost Effectiveness	.	.	.	.	.	.	.	.
Cost Per 100k Gallons Treated	121.51	182.91	0.00	22.53	61.01	146.19	1,152.23	117.00
Monitoring	.	.	.	.	.	.	.	.
Number of Readings	66.60	56.52	2.00	21.00	63.00	88.00	361.00	122.00
Maintenance	.	.	.	.	.	.	.	.
Change in pH Effectiveness	-0.45	1.68	-5.60	-1.05	-0.25	0.45	3.70	102.00
Change in Δ Ln Manganese	-0.10	1.60	-5.07	-0.94	-0.13	0.74	5.95	91.00
Change in Δ Ln Aluminum	0.51	1.56	-2.93	-0.39	0.36	1.50	4.77	86.00
Change in Δ Ln Iron	0.26	1.82	-4.40	-0.93	0.19	1.41	5.76	91.00
Change in Δ Ln TSS	-0.22	1.43	-4.71	-0.94	0.02	0.63	4.41	83.00

Note: Data are from Datashed water quality reports and project documents.

Table C3: Comparison of One-to-One Without Replacement MDM Matched State and Association Systems

	Association System (N=61)	State System (N=61)	Diff.	Normalized Diff.	P-Value
Characteristics	.	.	.	.	.
Year Built	2,002.98	2,001.84	1.15	0.27	0.14
Number of Funding Sources	1.44	1.52	-0.08	-0.12	0.52
Severity	.	.	.	.	.
Initial Influent pH	4.16	4.33	-0.16	-0.12	0.54
Avg Influent pH	4.10	4.09	0.00	0.00	0.99
Influent Manganese (mg/l)	6.75	8.88	-2.13	-0.12	0.51
Influent Aluminum (mg/l)	10.47	11.45	-0.98	-0.06	0.73
Influent Iron (mg/l)	24.59	29.72	-5.13	-0.14	0.46
Influent TSS (mg/l)	11.92	16.43	-4.51	-0.23	0.21
Severity Threshold	.	.	.	.	.
Any Influent Standard Failed	0.98	1.00	-0.02	-0.18	0.32
Initial Influent pH ≤ 6	0.80	0.78	0.02	0.05	0.81
Avg Influent pH ≤ 6	0.85	0.85	0.00	0.00	1.00
Influent Mang. ≥ 2 (mg/l)	0.66	0.61	0.05	0.10	0.58
Influent Aluminum $\geq .5$ (mg/l)	0.80	0.82	-0.02	-0.04	0.82
Influent Iron ≥ 1.5 (mg/l)	0.80	0.80	0.00	0.00	1.00
Influent TSS ≥ 35 (mg/l)	0.03	0.08	-0.05	-0.21	0.25
Complexity	.	.	.	.	.
Number of Inflows	1.31	1.36	-0.05	-0.07	0.68
Size	.	.	.	.	.
Inflow Volume (gal/min)	239.59	259.66	-20.07	-0.03	0.85

Note: Normalized Mean Differences are calculated using the method of Imbens and Wooldridge (2009), which is the difference in means divided by the square root of the average variance. P-Values are two-tailed and from a t-test for the difference in means of unpaired data, assuming unequal variances. The Severity Threshold variables are all binary variables.

Table C4: Mean Differences: One-to-One Without Replacement MDM Matched State and Association Systems

	Association System (N=61)	State System (N=61)	Diff.	P-Value
Effectiveness	.	.	.	.
Initial Change in pH	2.40	2.41	-0.00	0.99
Avg Change in pH	1.70	1.62	0.08	0.74
Δ Ln Manganese (mg/l)	-0.51	-0.59	0.08	0.53
Δ Ln Aluminum (mg/l)	-1.66	-1.30	-0.36	0.19
Δ Ln Iron (mg/l)	-1.76	-1.75	-0.01	0.97
Δ Ln TSS (mg/l)	-0.01	0.05	-0.07	0.71
Effectiveness Threshold	.	.	.	.
Any Effluent Standard Failed	0.75	0.85	-0.10	0.17
Initial Effluent pH ≤ 6	0.12	0.18	-0.06	0.41
Avg Effluent pH ≤ 6	0.43	0.49	-0.07	0.47
Effluent Mang. ≥ 2 (mg/l)	0.44	0.49	-0.05	0.59
Effluent Aluminum $\geq .5$ (mg/l)	0.49	0.61	-0.11	0.21
Effluent Iron ≥ 1.5 (mg/l)	0.48	0.48	0.00	1.00
Effluent TSS ≥ 35 (mg/l)	0.02	0.11	-0.10	0.03
Cost Effectiveness	.	.	.	.
Ln Cost 100k Gallons Treated	3.68	4.14	-0.46	0.13
Monitoring	.	.	.	.
Ln Number of Readings	3.31	4.14	-0.83	0.00
Maintenance	.	.	.	.
Change in pH Effectiveness	-0.20	-0.70	0.50	0.13
Change in Δ Ln Manganese	-0.23	0.05	-0.27	0.41
Change in Δ Ln Aluminum	0.50	0.53	-0.02	0.95
Change in Δ Ln Iron	0.25	0.26	-0.00	0.99
Change in Δ Ln TSS	-0.37	-0.04	-0.34	0.29

Note: P-Values are two-tailed and from a t-test for the difference in means of unpaired data, assuming unequal variances. The Effectiveness Threshold variables are all binary variables.

Table C5: Biased Corrected Effects: Effectiveness, One-to-One Without Replacement

	Any	Initial pH		pH		Mn		Al		Fe	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	p(Fail)	Δ	p(Fail)	Δ	p(Fail)	Δ Ln	p(Fail)	Δ Ln	p(Fail)	Δ Ln	p(Fail)
State System	0.10	0.24	0.09	0.09	0.13	-0.08	0.07	0.04	0.14	0.14	0.06
	(0.08)	(0.33)	(0.10)	(0.28)	(0.12)	(0.14)	(0.14)	(0.28)	(0.12)	(0.32)	(0.12)
Ln Influent Flow (gal/min)	0.02	-0.14	0.02	-0.13	0.05	0.05	0.01	0.15*	0.03	0.25**	0.03
	(0.03)	(0.10)	(0.02)	(0.08)	(0.04)	(0.05)	(0.05)	(0.07)	(0.03)	(0.08)	(0.04)
Ln Number of Inflows	0.07	-0.97*	0.21	-0.55	0.05	0.00	0.05	0.18	0.07	-0.26	0.07
	(0.09)	(0.40)	(0.12)	(0.41)	(0.16)	(0.22)	(0.19)	(0.40)	(0.15)	(0.34)	(0.14)
R-squared	0.11	0.20	0.25	0.13	0.11	0.14	0.11	0.17	0.09	0.12	0.25
N	110	99	78	111	94	110	71	110	89	110	91
Cond. on Failed Influent Std.	Y	N	Y	N	Y	N	Y	N	Y	N	Y

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction. Each system in the regressions is weighted using analytical weights of the number of readings that contribute its average pH and pollutant concentrations.

Table C6: Biased Corrected Effects: Cost-Effectiveness and Monitoring, One-to-One Without Replacement

	Ln(Cost Per 100k Gallons)	Ln(Number of Readings)
	(1)	(2)
State System	0.33	0.59**
	(0.32)	(0.20)
Ln Number of Inflows	0.06	0.45*
	(0.37)	(0.22)
Ln Influent Flow (gal/min)		0.09
		(0.06)
R-squared	0.05	0.37
N	111	111

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction.

Table C7: Biased Corrected Effects: Maintenance, One-to-One Without Replacement

	Change in Δ pH	Change in Δ Ln			
	(1)	(2)	(3)	(4)	(5)
	pH	Mn	Al	Fe	TSS
State System	-0.86*	0.22	0.23	0.44	0.53
	(0.37)	(0.40)	(0.40)	(0.47)	(0.43)
Ln Influent Flow (gal/min)	0.00	-0.00	0.02	0.15	0.22
	(0.11)	(0.12)	(0.14)	(0.13)	(0.12)
Ln Number of Inflows	0.69	0.53	-0.65	0.66	0.21
	(0.46)	(0.44)	(0.58)	(0.51)	(0.38)
Ln Measurement Duration	-0.46	-0.47	0.04	-0.35	0.01
	(0.29)	(0.49)	(0.34)	(0.42)	(0.41)
R-squared	0.21	0.19	0.11	0.28	0.24
N	99	88	82	88	79

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction.

D Additional Tables: One-to-Two Matching With Replacement

Table D1: Descriptive Statistics: One-to-Two With Replacement MDM Sample

	Mean	SD	Min.	p25	p50	p75	Max	N
Characteristics	.	.	.	.	.	.	.	.
State System	0.33	0.47	0.00	0.00	0.00	1.00	1.00	183.00
Year Built	2,002.48	4.21	1,990.00	2,000.00	2,003.00	2,005.00	2,019.00	183.00
Number of Funding Sources	1.48	0.69	1.00	1.00	1.00	2.00	4.00	183.00
Severity	.	.	.	.	.	.	.	.
Initial Influent pH	4.21	1.32	2.50	3.18	3.72	5.21	7.21	153.00
Avg Influent pH	4.06	1.22	1.96	3.15	3.61	4.49	6.94	183.00
Influent Manganese (mg/l)	7.42	16.48	0.00	1.63	2.97	7.19	154.02	183.00
Influent Aluminum (mg/l)	11.07	14.53	0.00	0.74	6.36	13.21	103.13	183.00
Influent Iron (mg/l)	27.00	37.25	0.00	2.83	13.96	37.96	217.09	183.00
Influent TSS (mg/l)	12.81	17.43	0.00	5.36	7.67	13.67	171.49	183.00
Severity Threshold	.	.	.	.	.	.	.	.
Any Influent Standard Failed	0.99	0.07	0.00	1.00	1.00	1.00	1.00	183.00
Initial Influent pH ≤ 6	0.80	0.40	0.00	1.00	1.00	1.00	1.00	153.00
Avg Influent pH ≤ 6	0.85	0.36	0.00	1.00	1.00	1.00	1.00	183.00
Influent Mang. ≥ 2 (mg/l)	0.62	0.49	0.00	0.00	1.00	1.00	1.00	183.00
Influent Aluminum $\geq .5$ (mg/l)	0.82	0.39	0.00	1.00	1.00	1.00	1.00	183.00
Influent Iron ≥ 1.5 (mg/l)	0.79	0.41	0.00	1.00	1.00	1.00	1.00	183.00
Influent TSS ≥ 35 (mg/l)	0.05	0.22	0.00	0.00	0.00	0.00	1.00	183.00
Complexity	.	.	.	.	.	.	.	.
Number of Inflows	1.32	0.63	1.00	1.00	1.00	1.00	4.00	183.00
Size	.	.	.	.	.	.	.	.
Inflow Volume (gal/min)	227.68	511.48	0.00	16.75	61.40	169.43	3,832.22	183.00

Note: Data are from Datashed water quality reports and project documents.

Table D2: Descriptive Statistics: One-to-Two With Replacement MDM Sample, Outcome Variables

	Mean	SD	Min.	p25	p50	p75	Max	N
Effectiveness	.	.	.	.	.	.	.	.
Initial Change in pH	2.31	1.55	-2.85	1.00	2.80	3.55	4.93	153.00
Avg Change in pH	1.69	1.34	-0.78	0.60	1.33	2.77	4.53	183.00
Δ Ln Manganese (mg/l)	-0.61	0.87	-6.71	-1.02	-0.34	-0.06	1.14	181.00
Δ Ln Aluminum (mg/l)	-1.61	1.68	-6.23	-2.71	-1.19	-0.18	1.06	182.00
Δ Ln Iron (mg/l)	-1.78	1.67	-6.72	-2.87	-1.49	-0.63	3.17	182.00
Δ Ln TSS (mg/l)	0.03	0.91	-2.68	-0.44	0.00	0.62	4.36	182.00
Effectiveness Threshold	.	.	.	.	.	.	.	.
Any Effluent Standard Failed	0.78	0.41	0.00	1.00	1.00	1.00	1.00	183.00
Initial Effluent pH ≤ 6	0.18	0.38	0.00	0.00	0.00	0.00	1.00	153.00
Avg Effluent pH ≤ 6	0.47	0.50	0.00	0.00	0.00	1.00	1.00	183.00
Effluent Mang. ≥ 2 (mg/l)	0.43	0.50	0.00	0.00	0.00	1.00	1.00	183.00
Effluent Aluminum $\geq .5$ (mg/l)	0.54	0.50	0.00	0.00	1.00	1.00	1.00	183.00
Effluent Iron ≥ 1.5 (mg/l)	0.45	0.50	0.00	0.00	0.00	1.00	1.00	183.00
Effluent TSS ≥ 35 (mg/l)	0.04	0.21	0.00	0.00	0.00	0.00	1.00	183.00
Cost Effectiveness	.	.	.	.	.	.	.	.
Cost Per 100k Gallons Treated	142.09	237.90	0.00	22.53	60.52	148.36	1,231.16	174.00
Monitoring	.	.	.	.	.	.	.	.
Number of Readings	61.99	54.72	2.00	14.00	58.00	87.00	361.00	183.00
Maintenance	.	.	.	.	.	.	.	.
Change in pH Effectiveness	-0.34	1.67	-5.60	-0.90	-0.19	0.42	4.45	153.00
Change in Δ Ln Manganese	-0.28	1.58	-5.96	-1.10	-0.19	0.48	5.95	137.00
Change in Δ Ln Aluminum	0.49	1.56	-2.93	-0.39	0.37	1.50	4.77	129.00
Change in Δ Ln Iron	0.32	1.69	-4.40	-0.71	0.35	1.26	5.76	137.00
Change in Δ Ln TSS	-0.19	1.42	-4.71	-0.76	-0.01	0.64	4.41	126.00

Note: Data are from Datashed water quality reports and project documents.

Table D3: Comparison of One-to-Two With Replacement MDM Matched State and Association Constructed Systems

	Association System (N=122)	State System (N=61)	Diff.	Normalized Diff.	P-Value
Characteristics	.	.	.	.	.
Year Built	2,002.80	2,001.84	0.96	0.23	0.13
Number of Funding Sources	1.46	1.52	-0.07	-0.09	0.56
Severity	.	.	.	.	.
Initial Influent pH	4.16	4.33	-0.17	-0.13	0.46
Avg Influent pH	4.04	4.09	-0.06	-0.04	0.78
Influent Manganese (mg/l)	6.68	8.88	-2.20	-0.12	0.46
Influent Aluminum (mg/l)	10.88	11.45	-0.57	-0.04	0.83
Influent Iron (mg/l)	25.63	29.72	-4.09	-0.11	0.52
Influent TSS (mg/l)	11.00	16.43	-5.43	-0.27	0.12
Severity Threshold	.	.	.	.	.
Any Influent Standard Failed	0.99	1.00	-0.01	-0.13	0.32
Initial Influent pH ≤ 6	0.81	0.78	0.03	0.07	0.67
Avg Influent pH ≤ 6	0.85	0.85	0.00	0.00	1.00
Influent Mang. ≥ 2 (mg/l)	0.62	0.61	0.02	0.03	0.83
Influent Aluminum $\geq .5$ (mg/l)	0.82	0.82	0.00	0.00	1.00
Influent Iron ≥ 1.5 (mg/l)	0.79	0.80	-0.02	-0.04	0.80
Influent TSS ≥ 35 (mg/l)	0.03	0.08	-0.05	-0.21	0.21
Complexity	.	.	.	.	.
Number of Inflows	1.30	1.36	-0.07	-0.10	0.52
Size	.	.	.	.	.
Inflow Volume (gal/min)	211.70	259.66	-47.96	-0.09	0.57

Note: Normalized mean differences are calculated using the method of Imbens and Wooldridge (2009), which is the difference in means divided by the square root of the average variance. P-Values are two-tailed and from a t-test for the difference in means of unpaired data, assuming unequal variances. The Severity Threshold variables are all binary variables.

Table D4: Mean Differences: One-to-Two With Replacement MDM Matched State and Association Constructed Systems

	Association System (N=122)	State System (N=61)	Diff.	P-Value
Effectiveness	.	.	.	.
Initial Change in pH	2.26	2.41	-0.15	0.58
Avg Change in pH	1.73	1.62	0.12	0.57
Δ Ln Manganese (mg/l)	-0.62	-0.59	-0.03	0.83
Δ Ln Aluminum (mg/l)	-1.76	-1.30	-0.46	0.06
Δ Ln Iron (mg/l)	-1.80	-1.75	-0.05	0.85
Δ Ln TSS (mg/l)	0.02	0.05	-0.03	0.84
Effectiveness Threshold	.	.	.	.
Any Effluent Standard Failed	0.75	0.85	-0.11	0.08
Initial Effluent pH ≤ 6	0.18	0.18	0.00	1.00
Avg Effluent pH ≤ 6	0.46	0.49	-0.03	0.68
Effluent Mang. ≥ 2 (mg/l)	0.39	0.49	-0.10	0.21
Effluent Aluminum $\geq .5$ (mg/l)	0.51	0.61	-0.10	0.21
Effluent Iron ≥ 1.5 (mg/l)	0.44	0.48	-0.03	0.68
Effluent TSS ≥ 35 (mg/l)	0.01	0.11	-0.11	0.01
Cost Effectiveness	.	.	.	.
Ln Cost 100 Gallons Treated	3.85	4.14	-0.30	0.24
Monitoring	.	.	.	.
Ln Number of Readings	3.32	4.14	-0.82	0.00
Maintenance	.	.	.	.
Change in pH Effectiveness	-0.16	-0.70	0.55	0.07
Change in Δ Ln Manganese	-0.43	0.05	-0.48	0.07
Change in Δ Ln Aluminum	0.47	0.53	-0.05	0.87
Change in Δ Ln Iron	0.35	0.26	0.09	0.79
Change in Δ Ln TSS	-0.26	-0.04	-0.23	0.42

Note: P-Values are two-tailed and from a t-test for the difference in means of unpaired data, assuming unequal variances. The Effectiveness Threshold variables are all binary variables.

Table D5: Biased Corrected Effects: Effectiveness, One-to-Two Matching With Replacement

	Any	Initial pH		pH		Mn		Al		Fe	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	p(Fail)	Δ	p(Fail)	Δ	p(Fail)	Δ Ln	p(Fail)	Δ Ln	p(Fail)	Δ Ln	p(Fail)
State System	0.11	0.27	-0.04	-0.17	0.10	0.05	0.10	0.48	0.16	0.01	0.09
	(0.06)	(0.26)	(0.09)	(0.24)	(0.10)	(0.12)	(0.11)	(0.26)	(0.10)	(0.28)	(0.10)
Ln Influent Flow (gal/min)	0.04	-0.10	0.01	-0.17*	0.06	0.11**	0.00	0.25***	0.05	0.21**	0.05
	(0.02)	(0.09)	(0.02)	(0.07)	(0.03)	(0.04)	(0.04)	(0.07)	(0.03)	(0.08)	(0.03)
Ln Number of Inflows	0.06	-0.82*	0.31**	0.14	-0.13	-0.07	-0.01	-0.28	-0.10	-0.15	0.04
	(0.08)	(0.32)	(0.11)	(0.34)	(0.13)	(0.19)	(0.16)	(0.36)	(0.13)	(0.28)	(0.12)
R-squared	0.22	0.15	0.18	0.16	0.13	0.19	0.15	0.24	0.17	0.13	0.22
N	169	151	117	170	141	169	105	169	136	169	136
Cond. on Failed Influent Std.	Y	N	Y	N	Y	N	Y	N	Y	N	Y

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction. Using frequency weights, we weight each state system in the regressions by one, and the 65 association systems by the number of times they are matched to a state system.

Table D6: Biased Corrected Effects: Cost-Effectiveness and Monitoring, One-to-Two Matching With Replacement

	Ln(Cost Per 100k Gallons)	Ln(Number of Readings)
	(1)	(2)
State System	0.17	0.71^{***}
	(0.27)	(0.18)
Ln Number of Inflows	0.94^{*}	0.22
	(0.36)	(0.21)
Ln Influent Flow (gal/min)		0.15^{**}
		(0.05)
R-squared	0.26	0.35
N	170	170

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction. Using frequency weights, we weight each state system in the regressions by one, and the 65 association systems by the number of times they are matched to a state system.

Table D7: Biased Corrected Effects: Maintenance, One-to-Two Matching With Replacement

	Change in Δ pH	Change in Δ Ln			
	(1)	(2)	(3)	(4)	(5)
	pH	Mn	Al	Fe	TSS
State System	-0.77**	0.63*	0.29	0.23	0.53
	(0.30)	(0.30)	(0.34)	(0.40)	(0.34)
Ln Influent Flow (gal/min)	-0.01	-0.05	-0.04	0.11	0.13
	(0.09)	(0.09)	(0.12)	(0.10)	(0.10)
Ln Number of Inflows	0.41	0.52	-0.78	0.51	-0.13
	(0.37)	(0.34)	(0.46)	(0.40)	(0.33)
Ln Measurement Duration	-0.14	-0.48	-0.18	-0.52	-0.26
	(0.23)	(0.32)	(0.30)	(0.33)	(0.35)
R-squared	0.22	0.19	0.18	0.18	0.17
N	151	135	124	135	123

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction. Using frequency weights, we weight each state system in the regressions by one, and the 65 association systems by the number of times they are matched to a state system.

E Additional Tables: One-to-One Matching with Nearest Geographic Neighbor

Table E1: Descriptive Statistics: One-to-One Matching with Nearest Geographic Neighbor

	Mean	SD	Min.	p25	p50	p75	Max	N
Characteristics	.	.	.	.	.	.	.	.
State System	0.50	0.50	0.00	0.00	0.50	1.00	1.00	122.00
Year Built	2,002.28	4.94	1,988.00	1,999.00	2,002.00	2,005.00	2,018.00	122.00
Number of Funding Sources	1.26	0.84	0.00	1.00	1.00	2.00	4.00	122.00
Severity	.	.	.	.	.	.	.	.
Initial Influent pH	4.43	1.38	2.52	3.23	3.81	6.10	7.00	99.00
Avg Influent pH	4.26	1.35	1.96	3.21	3.64	5.62	6.96	122.00
Influent Manganese (mg/l)	9.59	20.43	0.00	1.56	2.89	7.36	154.02	122.00
Influent Aluminum (mg/l)	10.58	16.40	0.00	0.61	4.10	11.92	103.13	122.00
Influent Iron (mg/l)	26.53	37.28	0.00	2.83	11.09	37.96	217.09	122.00
Influent TSS (mg/l)	15.90	23.96	0.00	5.65	8.18	15.14	171.49	122.00
Severity Threshold	.	.	.	.	.	.	.	.
Any Influent Standard Failed	1.00	0.00	1.00	1.00	1.00	1.00	1.00	122.00
Initial Influent pH ≤ 6	0.73	0.45	0.00	0.00	1.00	1.00	1.00	99.00
Avg Influent pH ≤ 6	0.78	0.42	0.00	1.00	1.00	1.00	1.00	122.00
Influent Mang. ≥ 2 (mg/l)	0.62	0.49	0.00	0.00	1.00	1.00	1.00	122.00
Influent Aluminum $\geq .5$ (mg/l)	0.79	0.41	0.00	1.00	1.00	1.00	1.00	122.00
Influent Iron ≥ 1.5 (mg/l)	0.82	0.39	0.00	1.00	1.00	1.00	1.00	122.00
Influent TSS ≥ 35 (mg/l)	0.09	0.29	0.00	0.00	0.00	0.00	1.00	122.00
Complexity	.	.	.	.	.	.	.	.
Number of Inflows	1.50	0.92	1.00	1.00	1.00	2.00	5.00	122.00
Size	.	.	.	.	.	.	.	.
Inflow Volume (gal/min)	225.04	462.07	0.00	19.87	60.12	175.00	3,832.22	122.00

Note: Data are from Datashed water quality reports and project documents.

Table E2: Descriptive Statistics: One-to-One Matching with Nearest Geographic Neighbor, Outcome Variables

	Mean	SD	Min.	p25	p50	p75	Max	N
Effectiveness	.	.	.	.	.	.	.	.
Initial Change in pH	2.27	1.57	-0.97	0.78	2.70	3.68	4.80	99.00
Avg Change in pH	1.49	1.29	-0.64	0.45	1.04	2.49	4.54	122.00
Δ Ln Manganese (mg/l)	-0.49	0.80	-3.42	-0.72	-0.27	-0.02	1.04	121.00
Δ Ln Aluminum (mg/l)	-1.25	1.43	-6.23	-2.13	-0.91	-0.11	1.37	121.00
Δ Ln Iron (mg/l)	-1.65	1.52	-6.72	-2.48	-1.36	-0.58	0.99	121.00
Δ Ln TSS (mg/l)	0.05	1.01	-2.68	-0.39	0.04	0.44	4.36	121.00
Effectiveness Threshold	.	.	.	.	.	.	.	.
Any Effluent Standard Failed	0.92	0.28	0.00	1.00	1.00	1.00	1.00	122.00
Initial Effluent pH ≤ 6	0.15	0.36	0.00	0.00	0.00	0.00	1.00	99.00
Avg Effluent pH ≤ 6	0.49	0.50	0.00	0.00	0.00	1.00	1.00	122.00
Effluent Mang. ≥ 2 (mg/l)	0.53	0.50	0.00	0.00	1.00	1.00	1.00	122.00
Effluent Aluminum $\geq .5$ (mg/l)	0.62	0.49	0.00	0.00	1.00	1.00	1.00	122.00
Effluent Iron ≥ 1.5 (mg/l)	0.53	0.50	0.00	0.00	1.00	1.00	1.00	122.00
Effluent TSS ≥ 35 (mg/l)	0.09	0.29	0.00	0.00	0.00	0.00	1.00	122.00
Cost Effectiveness	.	.	.	.	.	.	.	.
Cost Per 100k Gallons Treated	108.01	147.96	0.65	22.04	63.54	133.12	1,152.23	110.00
Monitoring	.	.	.	.	.	.	.	.
Number of Readings	68.58	54.29	3.00	29.00	67.00	89.00	361.00	122.00
Maintenance	.	.	.	.	.	.	.	.
Change in pH Effectiveness	-0.50	1.53	-5.60	-1.03	-0.11	0.35	2.98	99.00
Change in Δ Ln Manganese	-0.44	1.46	-5.96	-0.88	-0.27	0.41	2.72	89.00
Change in Δ Ln Aluminum	0.34	1.45	-2.93	-0.37	0.16	0.88	4.77	84.00
Change in Δ Ln Iron	-0.10	1.86	-4.40	-1.29	0.16	1.14	4.21	86.00
Change in Δ Ln TSS	-0.18	1.48	-3.64	-1.17	-0.14	0.63	4.41	78.00

Note: Data are from Datashed water quality reports and project documents.

Table E3: Comparison of Samples: One-to-One Matching with Nearest Geographic Neighbor

	Association System (N=61)	State System (N=61)	Diff.	Normalized Diff.	P-Value
Characteristics	.	.	.	.	.
Year Built	2,002.72	2,001.84	0.89	0.18	0.32
Number of Funding Sources	1.11	1.41	-0.30	-0.35	0.05
Severity	.	.	.	.	.
Initial Influent pH	4.54	4.33	0.21	0.15	0.45
Avg Influent pH	4.42	4.09	0.33	0.24	0.18
Influent Manganese (mg/l)	10.31	8.88	1.42	0.07	0.70
Influent Aluminum (mg/l)	9.71	11.45	-1.74	-0.11	0.56
Influent Iron (mg/l)	23.33	29.72	-6.39	-0.17	0.35
Influent TSS (mg/l)	15.36	16.43	-1.06	-0.04	0.81
Severity Threshold	.	.	.	.	.
Any Influent Standard Failed	1.00	1.00	0.00	.	.
Initial Influent pH ≤ 6	0.67	0.78	-0.12	-0.26	0.19
Avg Influent pH ≤ 6	0.70	0.85	-0.15	-0.36	0.05
Influent Mang. ≥ 2 (mg/l)	0.64	0.61	0.03	0.07	0.71
Influent Aluminum $\geq .5$ (mg/l)	0.75	0.82	-0.07	-0.16	0.38
Influent Iron ≥ 1.5 (mg/l)	0.84	0.80	0.03	0.08	0.64
Influent TSS ≥ 35 (mg/l)	0.10	0.08	0.02	0.06	0.75
Complexity	.	.	.	.	.
Number of Inflows	1.64	1.36	0.28	0.31	0.10
Size	.	.	.	.	.
Inflow Volume (gal/min)	190.42	259.66	-69.23	-0.15	0.41

Note: Normalized Mean Differences are calculated using the method of Imbens and Wooldridge (2009), which is the difference in means divided by the square root of the average variance. P-Values are two-tailed and from a t-test for the difference in means of unpaired data, assuming unequal variances. The Severity Threshold variables are all binary variables.

Table E4: Mean Differences: One-to-One Matching with Nearest Geographic Neighbor

	Association System (N=61)	State System (N=61)	Diff.	P-Value
Effectiveness	.	.	.	.
Initial Change in pH	2.12	2.41	-0.29	0.37
Avg Change in pH	1.36	1.62	-0.26	0.27
Δ Ln Manganese (mg/l)	-0.38	-0.59	0.21	0.15
Δ Ln Aluminum (mg/l)	-1.20	-1.30	0.10	0.69
Δ Ln Iron (mg/l)	-1.56	-1.75	0.19	0.49
Δ Ln TSS (mg/l)	0.05	0.05	-0.00	0.99
Effectiveness Threshold	.	.	.	.
Any Effluent Standard Failed	0.98	0.85	0.13	0.01
Initial Effluent pH ≤ 6	0.12	0.18	-0.05	0.48
Avg Effluent pH ≤ 6	0.49	0.49	0.00	1.00
Effluent Mang. ≥ 2 (mg/l)	0.57	0.49	0.08	0.37
Effluent Aluminum $\geq .5$ (mg/l)	0.64	0.61	0.03	0.71
Effluent Iron ≥ 1.5 (mg/l)	0.59	0.48	0.11	0.21
Effluent TSS ≥ 35 (mg/l)	0.07	0.11	-0.05	0.35
Cost Effectiveness	.	.	.	.
Ln Cost 100k Gallons Treated	3.66	4.14	-0.48	0.08
Monitoring	.	.	.	.
Ln Number of Readings	3.53	4.14	-0.61	0.00
Maintenance	.	.	.	.
Change in pH Effectiveness	-0.28	-0.70	0.43	0.16
Change in Δ Ln Manganese	-0.89	0.05	-0.94	0.00
Change in Δ Ln Aluminum	0.17	0.53	-0.36	0.27
Change in Δ Ln Iron	-0.47	0.26	-0.73	0.07
Change in Δ Ln TSS	-0.33	-0.04	-0.29	0.39

Note: P-Values are two-tailed and from a t-test for the difference in means of unpaired data, assuming unequal variances. The Effectiveness Threshold variables are all binary variables.

Table E5: Biased Corrected Effects: Effectiveness, One-to-One Matching with Nearest Geographic Neighbor

	Any	Initial pH		pH		Mn		Al		Fe	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	p(Fail)	Δ	p(Fail)	Δ	p(Fail)	Δ Ln	p(Fail)	Δ Ln	p(Fail)	Δ Ln	p(Fail)
State System	-0.09*	0.09	0.08	0.10	0.04	-0.10	-0.27**	0.21	0.02	-0.28	0.00
	(0.04)	(0.34)	(0.12)	(0.30)	(0.11)	(0.20)	(0.08)	(0.32)	(0.11)	(0.41)	(0.12)
Ln Influent Flow (gal/min)	0.00	-0.09	0.06	-0.14	0.02	0.06	0.01	0.17	0.05	0.33**	0.00
	(0.03)	(0.12)	(0.03)	(0.10)	(0.04)	(0.06)	(0.05)	(0.09)	(0.04)	(0.10)	(0.05)
Ln Number of Inflows	-0.03	-0.46	0.17	-0.68	0.33**	0.30	0.09	0.73	0.17	0.21	-0.03
	(0.07)	(0.39)	(0.10)	(0.37)	(0.12)	(0.25)	(0.18)	(0.43)	(0.11)	(0.40)	(0.12)
R-squared	0.22	0.26	0.31	0.19	0.28	0.15	0.26	0.20	0.17	0.20	0.26
N	106	95	69	106	84	105	64	105	82	105	85
Cond. on Failed Influent Std.	Y	N	Y	N	Y	N	Y	N	Y	N	Y

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction. Each system in the regressions is weighted using analytical weights of the number of readings that contribute its average pH and pollutant concentrations.

Table E6: Biased Corrected Effects: Cost-Effectiveness and Monitoring, One-to-One Matching with Nearest Geographic Neighbor

	Ln(Cost Per 100k Gallons)	Ln(Number of Readings)
	(1)	(2)
State System	0.33	0.40*
	(0.28)	(0.18)
Ln Number of Inflows	0.45	0.45*
	(0.33)	(0.20)
Ln Influent Flow (gal/min)		0.07
		(0.08)
R-squared	0.29	0.42
N	106	106

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction.

Table E7: Biased Corrected Effects: Maintenance, One-to-One Matching with Nearest Geographic Neighbor

	Change in Δ pH	Change in Δ Ln			
	(1)	(2)	(3)	(4)	(5)
	pH	Mn	Al	Fe	TSS
State System	-0.60	1.22**	0.32	0.92*	0.32
	(0.35)	(0.39)	(0.40)	(0.45)	(0.48)
Ln Influent Flow (gal/min)	-0.07	0.08	0.11	0.28	0.07
	(0.14)	(0.11)	(0.17)	(0.18)	(0.16)
Ln Number of Inflows	0.30	0.31	0.54	1.06*	0.06
	(0.44)	(0.47)	(0.56)	(0.51)	(0.39)
Ln Measurement Duration	-0.86*	0.46	0.33	0.03	0.57
	(0.41)	(0.59)	(0.60)	(0.60)	(0.69)
R-squared	0.26	0.33	0.20	0.36	0.16
N	95	85	78	82	74

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. All models include fixed effects for the year of system construction.