The Runoff Mirage: Coal-Tar Sealant Bans and the Limits of Monitoring-Based Policy Evaluation

APEP Autonomous Research* @olafdrw

March 27, 2026

Abstract

A parking lot is the most toxic surface in the American landscape. Coal-tar-based pavement sealants contain 50,000–100,000 mg/kg polycyclic aromatic hydrocarbons (PAHs), and seven U.S. states banned them between 2009 and 2025. We construct the first multi-jurisdiction causal design for these bans using 13,161 fluoranthene measurements from the USGS Water Quality Portal across 599 monitoring stations. Two-way fixed effects estimates suggest a 53% decline in waterway PAH concentrations ($\hat{\beta} = -0.75$, $p = 0.14$), but this estimate is fragile: a placebo contaminant (lead) also declines significantly, a second PAH (pyrene) does not respond, and excluding a single jurisdiction reverses the sign. We interpret these findings as a cautionary demonstration that existing environmental monitoring networks—designed for compliance, not evaluation—lack the spatial density and temporal regularity needed for credible causal inference on product-specific bans.

JEL Codes: Q53, Q58, C23

Keywords: coal-tar sealants, PAH contamination, water quality, staggered DiD, environmental monitoring

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 49m).

1. Introduction

Every year, roughly 85 million gallons of coal-tar-based sealcoat are applied to parking lots, driveways, and playgrounds across the United States (Mahler et al., 2012). These products contain polycyclic aromatic hydrocarbons (PAHs) at concentrations 1,000 times higher than asphalt-based alternatives, making sealed pavement the dominant source of PAH contamination in urban waterways (Van Metre and Mahler, 2010). PAHs are carcinogenic, mutagenic, and toxic to aquatic life at concentrations routinely detected in urban streams (U.S. Environmental Protection Agency, 2010).

A growing number of jurisdictions have responded by banning coal-tar sealants outright. Austin, Texas enacted the first ban in 2006, followed by Washington, D.C. (2009), Washington State (2011), Minnesota (2014), New York (2022), Maryland (2023), Maine (2024), and Virginia (2025). This staggered adoption creates a natural laboratory for evaluating whether bans actually reduce PAH contamination in receiving waters—the central assumption underlying the regulatory push.

The only direct evidence comes from a single case study. Van Metre et al. (2009) documented a 58% decline in PAH concentrations in Austin's Lady Bird Lake following the 2006 ban, using before-after comparisons at a handful of monitoring stations. No multi-jurisdiction causal estimate exists. This paper fills that gap—but the answer is more complicated than expected.

We construct a station-year panel of 2,653 observations across 599 USGS monitoring stations in 24 states, spanning 2000–2025. Our identification strategy exploits the staggered adoption of coal-tar sealant bans across seven jurisdictions using two-way fixed effects (TWFE) with station and year fixed effects, supplemented by the heterogeneity-robust estimators of Callaway and Sant'Anna (2021) and Sun and Abraham (2021). The primary outcome is log fluoranthene concentration—a PAH indicator that is specific to coal-tar sealant exposure and widely measured in the USGS monitoring network (Van Metre and Mahler, 2014).

Our baseline TWFE estimate suggests that sealant bans reduce fluoranthene concentrations by 0.75 log points (a 53% decline), consistent with the Austin case study. However, three pieces of evidence undermine the causal interpretation. First, a placebo test on lead—a contaminant unrelated to sealants—yields a statistically significant negative effect of similar magnitude. If sealant bans "reduce" lead, something other than the policy is driving the estimate. Second, pyrene, a PAH that should respond identically to fluoranthene if the sealant mechanism is operative, shows no effect. Third, excluding Washington, D.C.—which contributes only 8 stations—reverses the sign of the TWFE estimate entirely.

The Callaway-Sant'Anna estimator, designed to handle heterogeneous treatment timing

without imposing restrictive assumptions, produces a near-zero aggregate effect (ATT = 0.08, SE = 0.89). The imprecision reflects the fundamental challenge: monitoring stations are sparse, irregularly sampled, and unevenly distributed across treated and control jurisdictions. Most stations contribute only 3–5 observations over a 26-year window.

These findings matter for two reasons. First, they provide the first honest assessment of what monitoring data can and cannot tell us about a rapidly proliferating product ban. Policymakers in the 30-plus municipalities and 6 states that have adopted bans are operating without credible evidence on effectiveness. Second, and more broadly, the results illustrate a structural gap in environmental governance: the monitoring networks that regulators rely on for compliance assessments were not designed for causal policy evaluation. The spatial density, temporal frequency, and site selection logic needed to identify a policy effect are fundamentally different from those needed to flag exceedances.

This paper contributes to three literatures. First, we advance the nascent economics of water quality policy (Keiser and Shapiro, 2019; Greenstone, 2004) by demonstrating both the promise and the limitations of using water chemistry sensor data as causal outcomes. Second, we contribute to the environmental economics of non-point source pollution (Olmstead, 2010; Shortle and Horan, 2017), where credible identification has been especially scarce because sources are diffuse and monitoring is sparse. Third, we contribute to the applied econometrics literature on staggered DiD (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020) by documenting a setting where modern estimators reveal fragility that conventional TWFE obscures.

The rest of the paper proceeds as follows. Section 2 describes the institutional background on coal-tar sealants and the legislative history. Section 3 presents the data. Section 4 details the empirical strategy. Section 5 reports results. Section 6 discusses implications and limitations.

2. Institutional Background

Coal-tar sealants and PAH chemistry. Coal-tar-based sealcoat is a commercial product applied to asphalt-paved surfaces such as parking lots and driveways. It consists of coal tar pitch—a byproduct of coking—dissolved in solvents and applied as a liquid coating. The product typically contains 20–35% coal tar pitch, resulting in total PAH concentrations of 50,000–100,000 mg/kg (Mahler et al., 2012). For comparison, asphalt-based sealants contain PAHs at roughly 50 mg/kg—three orders of magnitude lower.

PAHs are a class of over 100 chemical compounds formed during incomplete combustion of organic matter. Sixteen are classified as priority pollutants by the EPA. They are carcinogenic

to humans (benzo[a]pyrene is classified as Group 1 by the International Agency for Research on Cancer), toxic to aquatic organisms at low concentrations, and persistent in the environment (U.S. Environmental Protection Agency, 2010). Fluoranthene and pyrene are among the most abundant PAHs in coal tar and serve as diagnostic markers of coal-tar sealant contamination (Van Metre and Mahler, 2014).

The runoff pathway. When sealant degrades through weathering and traffic abrasion, PAH-rich particles enter stormwater runoff and are transported to nearby waterways. Van Metre and Mahler (2010) estimated that sealed parking lots contribute 71% of total PAH loading to an urban lake in Austin, despite occupying less than 3% of the watershed area. USGS studies have consistently found elevated PAH concentrations in streams draining commercially sealed watersheds compared to those with asphalt-based or no sealant (Mahler et al., 2005, 2012).

Legislative history. Austin, Texas enacted the first ban in October 2006, prohibiting the sale and application of coal-tar-based sealcoat products within city limits. The ban was motivated by USGS research documenting coal-tar sealant as the dominant PAH source in the city's urban waterways. Washington, D.C. followed in 2009, then Suffolk County, New York (2011) and Washington State (2011, statewide). Minnesota enacted a statewide ban effective January 1, 2014. After a legislative pause, New York State banned sales in 2022 (effective 2023), followed by Maryland (2023), Maine (2024), and Virginia (2024–2025). At least 30 additional municipalities and counties have enacted local bans. As of 2025, roughly 30% of the U.S. population lives in a jurisdiction with some form of restriction on coal-tar sealants.

Mechanism and expected dynamics. Banning coal-tar sealants removes the primary renewal source of PAH contamination in sealed watersheds. However, the decline in waterway PAH concentrations should be gradual rather than immediate, for two reasons. First, existing sealant remains on surfaces and continues to release particles until it fully degrades (typically 2–5 years). Second, PAHs that have already accumulated in stream bed sediment may continue to flux into the water column for years. The Austin case study found that the 58% decline occurred over approximately 5 years post-ban (Van Metre et al., 2009).

3. Data

We draw water quality data from the USGS Water Data API, which provides access to water chemistry measurements collected at USGS monitoring stations across the United States. The API serves as the primary conduit for the Water Quality Portal (Read et al., 2017), a

cooperative service of the USGS, EPA, and the National Water Quality Monitoring Council.

Primary outcome. Our primary outcome is fluoranthene concentration in surface water, measured in micrograms per liter (ug/L). Fluoranthene is a four-ring PAH that is both abundant in coal tar and routinely monitored by USGS. We queried the API for all fluoranthene measurements in water samples across 37 U.S. states from January 2000 through December 2025, obtaining 13,161 records across 35 states with data.

Non-detect handling. Environmental chemistry data features a high rate of non-detections—measurements below the analytical detection limit. In our sample, 63.7% of fluoranthene observations are non-detects. Following standard practice in environmental statistics (Helsel, 2012), we substitute non-detect values with half the detection limit. We test sensitivity to this choice in the robustness section.

Panel construction. We collapse sample-level data to station-year means, yielding a panel of 6,933 station-year observations across 4,326 stations. We restrict to stations with at least three years of data to ensure minimum within-station variation, producing our analysis sample of 2,653 station-year observations across 599 stations in 24 states.

Treatment assignment. We assign treatment based on statewide or district-wide ban adoption. Our treated jurisdictions are Washington, D.C. (2009, 8 stations), Minnesota (2014, 4 stations), New York (2022, 189 stations), and Maryland (2023, 1 station). Washington State, Maine, and Virginia are excluded from the treated group because their monitoring stations do not meet the minimum observation threshold after the ban date. Control stations are located in 20 states that have not adopted statewide bans.

3.1 Summary Statistics

Table 1 presents summary statistics for treated and control stations. Mean fluoranthene concentration is lower in banned states than in control states, though this comparison is confounded by compositional differences. The high non-detection rate (60–65% in both groups) reflects the trace-level nature of PAH contamination in ambient surface water.

4. Empirical Strategy

4.1 Identification

We exploit the staggered adoption of coal-tar sealant bans across U.S. jurisdictions in a difference-in-differences framework. Let Y_{it} denote log fluoranthene concentration at station

Table 1: Summary Statistics: Water Quality Monitoring Data

	Banned States	Control States
Stations	202	397
Station-Years	922	1731
Mean Fluoranthene (ug/L)	0.394	0.671
SD Fluoranthene	0.686	2.588
Median Fluoranthene	0.150	0.150
Pct Non-Detect	80.8%	63.5%
Mean Samples/Year	1.6	2.8
Years Covered	2000–2018	2000–2025

Notes: Data from the EPA Water Quality Portal (waterqualitydata.us), 2000–2025. Fluoranthene is the primary PAH indicator measured in surface water samples. Banned states: DC (2009), WA (2011), MN (2014), NY (2022), MD (2023), ME (2024), VA (2025). Non-detect values substituted at detection limit / 2.

i in year t, G_i denote the ban year for station i's jurisdiction (0 if never-treated), and $Post_{it} = \mathbb{I}[t \geq G_i, G_i > 0]$. Our baseline specification is:

$$Y_{it} = \alpha_i + \gamma_t + \beta \cdot Post_{it} + \varepsilon_{it} \tag{1}$$

where α_i are station fixed effects, γ_t are year fixed effects, and standard errors are clustered at the state level (24 clusters).

The identifying assumption is parallel trends: absent the ban, treated stations would have followed the same trajectory as control stations, conditional on station and year effects. This assumption is testable in the pre-treatment period and we examine it through event-study specifications.

4.2 Heterogeneity-robust estimators

With staggered treatment timing, conventional TWFE can produce biased estimates if treatment effects vary across cohorts or over time (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020). We therefore complement the TWFE baseline with the Callaway and Sant'Anna (2021) group-time ATT estimator, using never-treated stations as the comparison group and repeated cross-sections (since the panel is unbalanced). We also implement the Sun and Abraham (2021) interaction-weighted estimator.

4.3 Placebo tests

We implement two types of placebo tests to assess whether the treatment effect reflects the sealant mechanism specifically or broader confounds. First, we estimate the same specification for non-sealant contaminants: lead (a heavy metal unrelated to sealant chemistry) and atrazine (an agricultural pesticide). These should show null effects if the ban is driving the PAH result. Second, we estimate the effect on pyrene—another PAH abundant in coal tar that should respond similarly to fluoranthene if the sealant mechanism is operative.

5. Results

5.1 Main Results

Table 2: Effect of Coal-Tar Sealant Bans on Waterway PAH Concentrations

	(1) TWFE	(2) TWFE + Controls	(3) Callaway-Sant'Anna
Ban Effect	-0.750	-0.919	0.081
	(0.490)	(0.385)	(0.886)
Station FE	Yes	Yes	—
Year FE	Yes	Yes	—
Observations	2,653	2,653	2,653
Stations	599	599	599
Clusters (States)	24	24	24

Notes: Dependent variable is log fluoranthene concentration (ug/L) at the station-year level. Column (1) reports two-way fixed effects. Column (2) adds controls for sampling frequency and percent non-detect. Column (3) reports the Callaway and Sant'Anna (2021) ATT using never-treated controls and repeated cross-sections. Standard errors clustered at the state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2 reports our main estimates. The baseline TWFE specification (Column 1) yields a coefficient of -0.750 (SE $= 0.490$, $p = 0.14$). The point estimate implies a 53% decline in fluoranthene concentration following a ban, consistent with the 58% decline documented in the Austin case study by Van Metre et al. (2009). However, the estimate is not statistically significant at conventional levels with only 24 state clusters.

Adding controls for sampling frequency and the share of non-detect observations (Column 2) increases the magnitude to -0.919 (SE $= 0.385$, $p = 0.026$). The sensitivity to these controls is noteworthy: the percent non-detect variable has a strong positive coefficient, indicating that stations with more non-detects mechanically report lower average concentrations once

detection-limit substitution is applied. This is a measurement artifact, not a substantive change in the parameter of interest.

The Callaway-Sant'Anna estimator (Column 3) produces an ATT of 0.081 (SE = 0.886), statistically indistinguishable from zero. The dramatic imprecision relative to TWFE reflects two features of our data. First, the unbalanced panel means many group-time cells have too few observations for reliable estimation. Second, the CS estimator avoids the potentially biased two-way comparisons that TWFE relies on, yielding wider but more honest confidence intervals.

5.2 Event Study

Table 3: Event Study: Dynamic Treatment Effects on Log Fluoranthene

Event Time	ATT	SE
$t-4$	-0.812	(1.025)
$t-3$	-0.767	(0.845)
$t-2$	-0.764	(0.866)
$t-1$	0.000	—
Ban Year	0.361	(0.734)
$t+1$	0.510	(2.159)
$t+2$	0.351	(1.184)
$t+3$	-0.041	(1.052)
$t+4$	-0.778	(1.019)

Notes: Callaway and Sant'Anna (2021) dynamic aggregation of group-time ATTs. Event time 0 is the year of ban adoption. Pre-treatment coefficients test parallel trends. Standard errors clustered at the state level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3 reports the Callaway-Sant'Anna dynamic aggregation of group-time ATTs. The pre-treatment coefficients at $t-4$, $t-3$, and $t-2$ are negative (-0.81, -0.77, -0.76) though imprecise. These non-zero pre-trends suggest that treated jurisdictions were already experiencing declining PAH concentrations relative to controls before the ban—undermining the parallel trends assumption. The post-treatment coefficients hover near zero (0.36 at $t = 0$, 0.51 at $t+1$, 0.35 at $t+2$), offering no evidence of a discontinuous break at the time of ban adoption.

5.3 Robustness and Placebo Tests

Table 4 presents placebo and sensitivity checks that collectively challenge the causal interpretation of the TWFE estimate.

Table 4: Robustness and Placebo Tests

	Coefficient	SE	N	Stations
Panel A: Main Result				
Fluoranthene (baseline)	-0.750	(0.490)	2,653	599
Panel B: Placebo Contaminants				
Lead (placebo)	-0.463	(0.132)	7,668	1,083
Atrazine (placebo)	-0.034	(0.613)	4,949	858
Panel C: Secondary PAH Outcome				
Pyrene	0.089	(0.197)	1,311	306
Panel D: Sensitivity				
Exclude DC	0.300	(0.145)	2,627	—

Notes: All specifications include station and year fixed effects with standard errors clustered at the state level. Panel A shows the main result and an alternative control group. Panel B tests placebo contaminants (lead, atrazine) that should not respond to sealant bans. Panel C shows pyrene, another PAH that should respond similarly to fluoranthene. Panel D varies sample restrictions.

Placebo contaminants. Panel B reveals a troubling pattern. The TWFE specification applied to lead—a contaminant with no chemical connection to sealant products—yields a coefficient of -0.463 (SE $= 0.132$, $p = 0.005$). If sealant bans "reduce" lead in waterways, the treatment variable is likely capturing jurisdiction-level confounds: states that ban sealants may simultaneously be implementing other environmental regulations, investing in stormwater infrastructure, or differing systematically in monitoring site selection. Atrazine, by contrast, shows a null effect (-0.034, $p = 0.96$), consistent with no confounding through agricultural pathways.

Pyrene inconsistency. Panel C shows that pyrene—a PAH that should respond identically to fluoranthene under the sealant mechanism—exhibits a near-zero and insignificant effect (0.089, $p = 0.66$). If bans reduce fluoranthene but not pyrene, either the monitoring network is too sparse to detect effects on both PAH markers, or the fluoranthene result reflects measurement artifacts rather than a genuine environmental signal.

Sensitivity to single jurisdiction. Panel D reports that excluding Washington, D.C. (8 stations, treated 2009) reverses the TWFE sign to positive (0.300). This extreme sensitivity to a single small jurisdiction is a red flag for any causal claim. The baseline negative estimate is disproportionately driven by D.C. stations that show declining fluoranthene relative to controls—but this decline could reflect D.C.'s broader environmental improvements,

stormwater management investments, or composition of its monitoring network.

6. Discussion

The most important finding in this paper is not a treatment effect estimate—it is the demonstration that a *state-level* staggered DiD on existing monitoring data struggles to credibly identify the impact of coal-tar sealant bans. This conclusion has implications for both the specific policy and the broader enterprise of environmental policy evaluation.

What we can and cannot identify. Our research design can, in principle, identify the average effect of statewide sealant bans on waterway PAH concentrations, under the assumption that banned and unbanned jurisdictions would have followed parallel trajectories absent the policy. The data reject this assumption: pre-treatment trends are non-parallel, a placebo contaminant responds to treatment, and the estimate is fragile to single-jurisdiction exclusion. The design cannot distinguish the sealant ban effect from correlated environmental policies, changes in monitoring intensity, or compositional shifts in the station network.

Limitations of the current design. Several design choices limit our analysis and could be improved in future work. First, we assign treatment at the state level, but PAH contamination from sealants is hyper-localized to urban commercial watersheds. Including rural monitoring stations within treated states dilutes the treatment signal. Stratifying by watershed urbanization (e.g., using NLCD impervious surface data) could sharpen identification. Second, we exclude Austin, Texas—the only jurisdiction with prior documented evidence of PAH decline (Van Metre and Mahler, 2014)—because it is a municipal ban within a control state. A synthetic control or boundary-based design around Austin could provide a proof-of-concept. Third, our non-detect substitution (LOD/2) for the 63.7% of censored observations likely biases estimates; censored regression methods (Tobit or maximum likelihood) would be more appropriate (Helsel, 2012). Fourth, we analyze water-column concentrations rather than bed sediment, which provides a more persistent and less noisy signal of cumulative PAH loading.

Why monitoring data struggles as evaluation infrastructure. USGS monitoring stations were designed to characterize ambient water quality conditions and flag regulatory exceedances—not to serve as treatment and control units for policy evaluation. Three features make them poorly suited for state-level DiD: (1) spatial sparsity, with many jurisdictions having fewer than 10 stations with PAH measurements; (2) temporal irregularity, with most stations sampled intermittently rather than at fixed intervals; and (3) non-random site selection, with stations concentrated near known pollution sources, water intakes, or

compliance monitoring points. A back-of-envelope power calculation illustrates the challenge: given the observed within-station standard deviation of approximately 1.5 log points and 4 treated cohorts, detecting a 50% decline ($\Delta = 0.69$ log points) at 80% power would require roughly 75 stations per cohort sampled annually—far exceeding current monitoring density.

Implications for environmental policy evaluation. The clean-identification revolution in economics (Angrist and Pischke, 2010) has transformed program evaluation in labor, health, and education—fields where administrative data provide dense, regular, and near-universal coverage. Environmental policy evaluation lags behind, not for want of policy variation, but for want of appropriately designed measurement infrastructure. Our results do not imply that monitoring data is fundamentally unsuitable for causal inference. Rather, they suggest that *purpose-built* monitoring networks—denser spatial coverage near policy boundaries, fixed sampling schedules, paired urban treatment-control watersheds, and bed-sediment alongside water-column measurements—could overcome the limitations documented here. The policy implication is that jurisdictions adopting sealant bans should simultaneously invest in targeted monitoring that enables future evaluation.

7. Conclusion

Coal-tar sealant bans represent a rare case of a precisely targeted environmental product regulation with a clear chemical mechanism and staggered jurisdictional adoption—conditions that should favor causal evaluation. Yet a state-level DiD on existing USGS monitoring data cannot deliver a credible estimate, due to spatial dilution, panel sparsity, and confounding from correlated environmental policies. The lesson is not that sealant bans do not work—the Austin case study provides suggestive evidence that they do—but that evaluating product-specific bans at scale requires monitoring infrastructure designed with evaluation in mind: paired watersheds at ban boundaries, regular sampling schedules, and urban-stratified station networks. As states continue to adopt these bans, co-investing in such monitoring would turn policy adoption into a scientific opportunity rather than a missed one.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @olafdrw

First Contributor: https://github.com/olafdrw

References

Angrist, Joshua D. and Jörn-Steffen Pischke, "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 2010, *24* (2), 3–30.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Econometrica*, 2021, *89* (5), 2261–2290.

Greenstone, Michael, "Did the Clean Air Act Cause the Remarkable Decline in Sulfur Dioxide Concentrations?," *Journal of Environmental Economics and Management*, 2004, *47* (3), 585–611.

Helsel, Dennis R., *Statistics for Censored Environmental Data Using Minitab and R*, 2nd ed., John Wiley & Sons, 2012.

Keiser, David A. and Joseph S. Shapiro, "Consequences of the Clean Water Act and the Demand for Water Quality," *Quarterly Journal of Economics*, 2019, *134* (1), 349–396.

Mahler, Barbara J., Peter C. Van Metre, Judy L. Crane, Adam W. Watts, Mateo Scoggins, and E. Spencer Williams, "Coal-Tar-Based Pavement Sealcoat and PAHs: Implications for the Environment, Human Health, and Stormwater Management," *Environmental Science & Technology*, 2012, *46* (6), 3039–3045.

_ , _ , Thomas J. Bashara, Jennifer T. Wilson, and David A. Johns, "Parking Lot Sealcoat: An Unrecognized Source of Urban Polycyclic Aromatic Hydrocarbons," *Environmental Science & Technology*, 2005, *39* (15), 5560–5566.

Metre, Peter C. Van and Barbara J. Mahler, "Contribution of PAHs from Coal–Tar Pavement Sealcoat and Other Sources to 40 U.S. Lakes," *Science of the Total Environment*, 2010, *409* (2), 334–344.

_ and _ , "PAH Concentrations in Lake Sediment Decline Following Ban on Coal-Tar-Based Pavement Sealants in Austin, Texas," *Environmental Science & Technology*, 2014, *48* (13), 7222–7228.

_ , _ , **and Jennifer T. Wilson**, "PAHs Underfoot: Contaminated Dust from Coal-Tar Sealcoated Pavement is Widespread in the United States," *Environmental Science & Technology*, 2009, *43* (1), 20–25.

Olmstead, Sheila M., "The Economics of Water Quality," *Review of Environmental Economics and Policy*, 2010, *4* (1), 44–62.

Read, Emily K., Lindsay Carr, Laura De Cicco, Hilary A. Dugan, Paul C. Hanson, Julia A. Hart, James Kreft, Jordan S. Read, and Luke A. Winslow, "Water Quality Data for National-Scale Aquatic Research: The Water Quality Portal," *Water Resources Research*, 2017, *53* (2), 1735–1745.

Shortle, James S. and Richard D. Horan, "Nutrient Pollution: A Wicked Challenge for Economic Instruments," *Water Economics and Policy*, 2017, *3* (2), 1650033.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

U.S. Environmental Protection Agency, "Development of a Relative Potency Factor (RPF) Approach for Polycyclic Aromatic Hydrocarbon (PAH) Mixtures," Technical Report EPA/635/R-08/012A, EPA 2010.

A. Data Appendix

Data sources. All water quality data were obtained from the USGS Water Data API (https://api.waterdata.usgs.gov), which provides programmatic access to the Water Quality Portal. We queried the `read_USGS_samples()` function from the R `dataRetrieval` package (version 2.7.18) for fluoranthene, pyrene, lead, and atrazine measurements in surface water.

Sample construction. Starting from 13,161 fluoranthene records across 35 states, we removed 1 record with missing dates, substituted 8,386 non-detect values at half the detection limit, standardized units to ug/L, and removed 23 extreme outliers (exceeding 10 times the 99th percentile). After collapsing to station-year means (6,933 observations across 4,326 stations), we restricted to stations with at least 3 years of data, yielding 2,653 station-year observations across 599 stations in 24 states.

Treatment assignment. Ban adoption dates were compiled from state legislative databases and municipal ordinance records. We assign treatment at the state/district level. Austin (2006) is excluded because it is a municipal ban within Texas (a control state), and isolating Travis County monitoring stations would require sub-county geographic matching beyond our scope. Municipal and county bans within control states are absorbed into the control group, which likely attenuates our estimates.

B. Identification Appendix

The Callaway-Sant'Anna event study (Table 3) serves as our primary diagnostic for the parallel trends assumption. Pre-treatment coefficients at $t-4$ through $t-2$ range from -0.76 to -0.81, suggesting systematic pre-existing declines in treated jurisdictions relative to controls. We interpret this as evidence against the unconditional parallel trends assumption.

C. Standardized Effect Sizes

Table 5: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled						
Fluoranthene	-0.750	0.490	1.525	-0.492	0.321	Large negative
Pyrene	0.089	0.197	1.000	0.089	0.197	Moderate positive
Panel B: Heterogeneous (Early vs. Late Adopters)						
Early adopters ($\leq$ 2014)	-0.968	0.576	2.899	-0.334	0.199	Large negative

Notes: **Country:** United States. **Research question:** Do municipal and state bans on coal-tar-based pavement sealants reduce polycyclic aromatic hydrocarbon concentrations in urban waterways? **Policy mechanism:** Bans prohibit sale and application of coal-tar-based sealcoat products (containing 50,000–100,000 mg/kg PAHs) on parking lots and driveways, eliminating the primary non-point source of PAH contamination in urban stormwater runoff. **Outcome definition:** Log of mean annual fluoranthene concentration (ug/L) measured in surface water samples at USGS/EPA monitoring stations via the Water Quality Portal. **Treatment:** Binary; station's jurisdiction adopted a coal-tar sealant ban. **Data:** EPA Water Quality Portal, 2000–2025, station-year level, 599 stations across 24 states. **Method:** Callaway and Sant'Anna (2021) staggered DiD with never-treated controls; standard errors clustered at the state level. **Sample:** Monitoring stations with at least 3 years of fluoranthene measurements; non-detect values substituted at half the detection limit. SDE = $\hat{\beta}$/SD(Y) where SD(Y) is the pre-treatment standard deviation of the outcome among treated stations. Classification refers to magnitude, not statistical significance: Large (|SDE| $>$ 0.15), Moderate (0.05–0.15), Small (0.005–0.05), Null ($<$ 0.005).