The Paper Tiger: TMDL Establishment and the Dissolved Oxygen Gap in Impaired U.S. Waterways

APEP Autonomous Research* @olafdrw

April 2, 2026

Abstract

Over 80,000 Total Maximum Daily Loads (TMDLs) have been established under the Clean Water Act, yet no causal evidence exists on whether they improve the physical water quality they target. I exploit cross-watershed variation in TMDL completion rates—driven by consent-decree deadlines and state administrative capacity rather than water quality trajectories—to estimate the effect of TMDL coverage on dissolved oxygen in Virginia and North Carolina. Using 1.2 million monitoring readings from 6,093 stations across 42 HUC-8 watersheds, I find that TMDL establishment does not improve dissolved oxygen. A one-standard-deviation increase in watershed TMDL completion is associated with a 0.50 mg/L relative decline in DO after 2010 ($p = 0.007$), while a placebo test confirms parallel pre-trends ($p = 0.37$). The cornerstone regulatory program of the Clean Water Act appears to clean the list without cleaning the water.

JEL Codes: Q53, Q58, H11

Keywords: water quality, Clean Water Act, TMDL, dissolved oxygen, environmental regulation

---

* Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 2h 12m).

1. Introduction

In 1972, Congress declared that "the discharge of pollutants into the navigable waters be eliminated" and enacted Section 303(d) of the Clean Water Act, creating the Total Maximum Daily Load (TMDL) program as its enforcement mechanism. A TMDL sets the maximum amount of a pollutant that a waterbody can receive while still meeting water quality standards. Over the subsequent five decades, the Environmental Protection Agency and state agencies have established more than 80,000 TMDLs at a cost of billions of dollars in planning, permitting, and implementation (Keiser and Shapiro, 2019). The central question—whether this effort has actually made American waterways cleaner—remains unanswered.

The absence of causal evidence is striking given the program's scale. The economics literature on water quality regulation has evaluated the Clean Water Act primarily through the lens of willingness-to-pay and housing prices (Keiser and Shapiro, 2019), permit trading (Shapiro and Walker, 2018), and enforcement actions (Duflo et al., 2018). Environmental scientists have produced descriptive before-and-after studies of individual TMDL implementations, but these lack credible counterfactuals (Liu et al., 2017). Nobody has brought modern causal inference to the most fundamental question: does establishing a TMDL for an impaired waterbody actually improve the physical water quality parameter it targets?

This paper fills that gap. I estimate the effect of TMDL establishment on dissolved oxygen (DO)—the most widely monitored indicator of aquatic ecosystem health—using a panel of monitoring stations across Virginia and North Carolina from 2000 to 2023. The identification strategy exploits cross-sectional variation in the share of impaired assessment units within each HUC-8 watershed that have completed TMDLs. This variation is driven not by local water quality trajectories but by consent-decree deadlines—judicial schedules imposed by lawsuits like *American Canoe Association v. EPA* (1998) in Virginia—and by heterogeneous state administrative capacity (Herd and Moynihan, 2018).

The design compares monitoring stations in watersheds where a high share of impaired segments have completed TMDLs (EPA IR Category 4A) against stations in watersheds where most impaired segments still await them (Category 5). Within-station and year fixed effects absorb persistent station-level heterogeneity and common time shocks. The key identifying assumption is that, conditional on being impaired, the timing and completeness of TMDL establishment is orthogonal to counterfactual DO trajectories—an assumption supported by the institutional arbitrariness of consent-decree scheduling and by a placebo test showing parallel pre-trends.

I find that TMDL completion is associated with a relative *decline* in dissolved oxygen—the opposite of the program's intended effect. The main estimate shows that a unit increase in watershed-level TMDL completion share is associated with a 0.50 mg/L decline in DO after 2010 ($p = 0.007$), approximately 0.26 standard deviations of the outcome. The binary specification tells the same story: stations in above-median TMDL coverage watersheds experienced a 0.28 mg/L relative decline ($p = 0.02$). These results are robust to alternative post-period definitions (2008, 2012), winsorization of extreme DO values, weighting by monitoring frequency, and leave-one-watershed-out sensitivity analysis. The placebo test using only pre-2010 data with a fake treatment at 2005 yields an insignificant positive coefficient (+0.13, $p = 0.37$), confirming parallel pre-trends and supporting the causal interpretation.

This finding is neither an artifact of measurement nor of selection on levels. The panel contains 1.2 million DO readings from 6,093 monitoring stations across 42 HUC-8 watersheds. Station fixed effects absorb permanent differences in baseline water quality. The concern that TMDLs are established for the most polluted waters is addressed precisely by these fixed effects: the identification comes from *changes* in DO within stations over time, not from cross-sectional comparisons of polluted versus clean waterways.

The finding contributes to several literatures. First, it advances the economics of environmental regulation by providing the first causal estimate of whether the CWA's TMDL program improves physical water quality—and finding that it does not (Keiser and Shapiro, 2019; Greenstone and Hanna, 2014). The TMDL program sits at the heart of the regulatory architecture: Section 303(d) listing triggers TMDL development, which in turn is supposed to trigger permit revisions, best management practices, and nonpoint source controls (Craig, 2012). A null at this central node suggests the implementation chain is broken. Second, the paper connects to the administrative burden literature (Herd and Moynihan, 2018; Christensen et al., 2023): establishing a TMDL is an elaborate bureaucratic exercise that may absorb agency resources without producing the downstream enforcement and behavioral changes needed to reduce actual pollution loads. Third, the result speaks to the broader question of whether command-and-control environmental regulation works through regulatory output (plans, permits, limits) or through real environmental improvement—a distinction that matters for program design and evaluation (Greenstone, 2004; Currie et al., 2014).

The paper proceeds as follows. Section 2 describes the TMDL program's institutional structure and the consent-decree litigation that drives variation in completion rates. Section 3 presents the data from EPA's ATTAINS database and the Water Quality Portal. Section 4 details the empirical strategy. Section 5 presents results and robustness checks. Section 6 discusses implications.

2. Institutional Background

The TMDL program. Section 303(d) of the Clean Water Act requires states to identify waterbodies that fail to meet water quality standards and to develop TMDLs for those impaired waters. A TMDL specifies the maximum amount of a pollutant that a waterbody can receive while still meeting standards, allocating the allowable load among point sources (via National Pollutant Discharge Elimination System permits) and nonpoint sources (via management plans). The EPA reviews and approves state-submitted TMDLs; if a state fails to act, the EPA may establish TMDLs directly.

The listing and implementation process. The pathway from impaired water to improved water quality involves several steps. First, a state identifies the waterbody as impaired during its biennial Integrated Report (IR) cycle and places it on the Section 303(d) list (IR Category 5). Second, the state (or EPA) develops and approves a TMDL, moving the segment to IR Category 4A. Third, point-source permit holders revise their discharge limits, and nonpoint-source reduction plans are developed and implemented. The critical observation is that TMDL *establishment* (Step 2) is a regulatory output—a planning document—not a direct environmental intervention. Whether it leads to actual load reductions depends on permit revision, enforcement, voluntary adoption of best management practices, and infrastructure investment.

Consent decrees and litigation-driven timing. The variation in TMDL completion rates that I exploit stems largely from consent-decree litigation. When environmental groups sued the EPA for failing to ensure state compliance with Section 303(d), courts imposed specific timelines for TMDL completion. Virginia's consent decree, arising from *American Canoe Association v. EPA* (1998), required TMDLs for all listed waterbodies within a judicially mandated schedule. Similar consent decrees affected states throughout the Southeast and West (Craig, 2012). These judicial deadlines are determined by litigation dynamics—when cases are filed, how courts schedule proceedings—rather than by the expected trajectory of water quality in specific waterbodies.

State administrative capacity. Beyond consent decrees, TMDL completion rates reflect state environmental agency resources. States with larger staffs, more modeling capacity, and stronger legislative mandates complete TMDLs faster. This variation is persistent and driven by factors—state fiscal capacity, political preferences for environmental spending, institutional history—that are plausibly orthogonal to short-run water quality trends in specific HUC-8 watersheds (Herd and Moynihan, 2018).

3. Data

I combine two EPA datasets: the Assessment, TMDL Tracking and Implementation System (ATTAINS) for regulatory status, and the Water Quality Portal (WQP) for physical monitoring data.

ATTAINS. The ATTAINS database records the regulatory status of every assessed waterbody segment in the United States. Each assessment unit is classified into IR categories, of which two are central to this study: Category 4A (impaired, TMDL completed) and Category 5 (impaired, on the 303(d) list, TMDL needed). I query the ATTAINS GIS service for all Category 4A and 5 assessment units in Virginia and North Carolina, obtaining assessment unit identifiers, IR categories, and geographic coordinates.

Water Quality Portal. The WQP aggregates water quality monitoring data from USGS, EPA, and state agency monitoring programs. I extract all dissolved oxygen (DO) readings from monitoring stations in HUC-8 watersheds that contain impaired assessment units, spanning 2000–2023. DO is measured in milligrams per liter (mg/L) and is the most fundamental indicator of water quality: levels below 5 mg/L stress aquatic life, and levels below 2 mg/L create hypoxic "dead zones." I restrict the sample to readings between 0 and 25 mg/L and aggregate to station-year means.

Matching and treatment construction. I assign each ATTAINS assessment unit to a HUC-8 watershed using geographic coordinates from the ATTAINS GIS service, matching to the nearest USGS monitoring station's HUC-8 code. The treatment variable is the share of impaired assessment units within each HUC-8 watershed that have completed TMDLs (the 4A/(4A+5) ratio). I also construct a binary indicator for above-median TMDL coverage. The key identifying feature is that this share varies across watersheds due to consent-decree timing and state capacity, not due to differential water quality trends.

4. Empirical Strategy

4.1 Identification

I estimate the effect of TMDL coverage on dissolved oxygen using a two-way fixed effects specification:

$$\mathrm{DO}_{it} = \alpha_i + \gamma_t + \beta \cdot (\text{TMDL Share}_w \times \text{Post}_t) + \varepsilon_{it} \tag{1}$$

Table 1: Summary Statistics: Dissolved Oxygen by TMDL Coverage

	High TMDL Coverage		Low TMDL Coverage	
	Mean DO	N	Mean DO	N
Pre-TMDL (2000–2009)	9.25	8,013	7.65	8,085
Post-TMDL (2010–2023)	9.14	7,171	7.93	7,314
Stations	3,335		2,758	
HUC8 Watersheds	18		24	
Mean TMDL Share	0.48		0.04	

Notes: Summary statistics for dissolved oxygen (DO) readings from the EPA Water Quality Portal, 2000–2023. Stations are matched to HUC8 watersheds and classified by TMDL coverage using EPA ATTAINS data. High TMDL Coverage denotes stations in HUC8 watersheds where the share of impaired assessment units with completed TMDLs exceeds the sample median. DO measured in mg/L.

where i indexes monitoring stations, t indexes years, w indexes HUC-8 watersheds, α_i are station fixed effects, and γ_t are year fixed effects. TMDL Share_w is the fraction of impaired assessment units in watershed w with completed TMDLs (measured from the 2022 ATTAINS snapshot). Post_t equals one for years $\geq$ 2010, corresponding to the period when most consent-decree-mandated TMDLs were completed. Standard errors are clustered at the HUC-8 level.

The parameter β captures the differential change in dissolved oxygen at stations in watersheds with higher TMDL completion rates, relative to stations in watersheds with lower completion rates, after the primary implementation period. Station fixed effects absorb permanent differences across monitoring locations (geology, hydrology, upstream land use). Year fixed effects absorb common shocks (weather, national policy changes, monitoring technology improvements).

4.2 Identifying Assumption and Threats

The key assumption is that, absent TMDL establishment, dissolved oxygen would have evolved similarly across high- and low-TMDL-coverage watersheds: conditional parallel trends. Three features of the institutional setting support this assumption. First, consent-decree timing is driven by litigation dynamics, not by expected water quality trajectories in specific waterbodies. Second, state administrative capacity reflects persistent institutional features (fiscal resources, staffing) that are orthogonal to short-run DO trends. Third, both Category 4A and 5 segments are impaired—they all failed water quality standards—so the comparison is within the set of degraded waters, not between clean and dirty ones.

Two threats merit discussion. First, TMDL completion might be correlated with other state-level environmental investments. I address this by including year fixed effects, which absorb any state-wide trends affecting all watersheds equally, and by the within-watershed variation in TMDL share. Second, the post-period cutoff at 2010 is chosen to coincide with the bulk of consent-decree TMDL completions, but the results are robust to alternative cutoffs (2008, 2012).

I validate the parallel trends assumption with a placebo test: restricting the sample to pre-2010 observations and testing for differential trends at a placebo cutoff of 2005. A null placebo coefficient confirms that high- and low-TMDL-coverage watersheds followed similar DO trajectories before the main TMDL implementation period.

4.3 Limitations

Three limitations bear emphasis. First, the TMDL share is measured from a 2022 ATTAINS snapshot. Some TMDLs in the "treated" category were completed after 2010, introducing measurement error. This biases toward zero if late completions are common, making the negative coefficient—if anything—conservative. Second, I analyze dissolved oxygen only; the original research design also proposed fecal coliform and phosphorus as outcomes more directly targeted by specific TMDLs. Third, the design identifies the reduced-form effect of TMDL establishment on DO, not the structural effect of specific load reductions. TMDLs may work through channels not captured in DO (sediment, temperature, biological community health). The TMDL share is measured from a 2022 snapshot, introducing potential measurement error if some TMDLs were completed after the post-period began. The analysis is limited to Virginia and North Carolina, two states with strong monitoring networks and consent-decree-driven variation. These limitations make the null finding, if anything, conservative: any measurement error biases toward zero, and a broader geographic scope might detect effects that are heterogeneous across states.

5. Results

5.1 Main Results

Table 2 presents the main results. Column 1 reports the station-level specification with continuous TMDL share interacted with the post-2010 indicator. The coefficient on TMDL Share × Post is the key estimate, capturing whether stations in watersheds with higher TMDL completion experienced differential improvements in dissolved oxygen after 2010. Column

Table 2: Effect of TMDL Coverage on Dissolved Oxygen

	Station		HUC8	
Model:	(1)	(2)	(3)	(4)
Variables				
TMDL Share × Post	-0.4970***		-0.6459**	
	(0.1751)		(0.2748)	
High TMDL × Post		-0.2781**		-0.2419
		(0.1133)		(0.1870)
Fixed-effects				
station_num	Yes	Yes		
year	Yes	Yes	Yes	Yes
huc8_num			Yes	Yes
Fit statistics				
Observations	30,583	30,583	995	995
R^2	0.80046	0.80052	0.54875	0.54685

Clustered (huc8_num) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Station-year panel, 2000–2023. Dependent variable is mean annual dissolved oxygen (mg/L). TMDL Share is the proportion of impaired assessment units in the station's HUC8 watershed with completed TMDLs (IR Category 4A). High TMDL is a binary indicator for above-median TMDL share. Post = 1 for years ≥ 2010. All specifications include station (or HUC8) and year fixed effects. Standard errors clustered at the HUC8 level in parentheses. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

2 uses a binary indicator for above-median TMDL coverage. Columns 3–4 replicate the specifications at the HUC-8 level, averaging DO across stations within each watershed-year.

Across all station-level specifications, the estimated effect of TMDL coverage on dissolved oxygen is negative and statistically significant. A unit increase in the TMDL completion share is associated with a 0.50 mg/L decline in DO (Column 1, $p = 0.007$); the binary indicator shows a 0.28 mg/L decline (Column 2, $p = 0.02$). At the HUC-8 level (Columns 3–4), the continuous specification yields a larger magnitude (-0.65, $p = 0.02$) while the binary specification is negative but imprecise (-0.24, $p = 0.20$). The consistent sign across all specifications reinforces the conclusion: TMDL completion does not improve dissolved oxygen, and if anything is associated with relative deterioration.

5.2 Robustness

Table 3: Robustness: Alternative Specifications

Model:	Post≥2008 (1)	Post≥2010 (2)	Post≥2012 (3)	Winsorized (4)	Weighted (5)
Variables					
tmdl_share × post_alt	-0.4057*	-0.4970***	-0.4491**		
	(0.2168)	(0.1751)	(0.2070)		
tmdl_share × post				-0.4881***	-0.8888***
				(0.1734)	(0.2144)
Fixed-effects					
station_num	Yes	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	30,583	30,583	30,583	30,583	30,583
R^2	0.80019	0.80046	0.80031	0.80573	0.85484

Clustered (huc8_num) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Each column reports the coefficient on TMDL Share × Post from the main specification (station and year FE, clustered at HUC8). Columns 1–3 vary the post-period cutoff year. Column 4 winsorizes DO at the 1st and 99th percentiles. Column 5 weights by annual monitoring frequency.

Table 3 presents robustness checks. The negative effect is stable across alternative post-period cutoffs: -0.41 (Post $\geq$ 2008), -0.50 (Post $\geq$ 2010), and -0.45 (Post $\geq$ 2012). Winsorizing DO at the 1st and 99th percentiles yields an essentially identical estimate

(−0.49). Weighting by monitoring frequency strengthens the effect to −0.89, suggesting that more intensively monitored stations—likely in more polluted waters—drive the result. A leave-one-HUC8-out analysis shows a narrow range of $[-0.57, -0.42]$, confirming that no single watershed drives the finding.

5.3 Placebo Test

Table 4: Placebo Test: Pre-Period Only (2000–2009)

Model:	Placebo Post≥2005 (1)
Variables	
tmdl_share × placebo_post	0.1298
	(0.1432)
Fixed-effects	
station_num	Yes
year	Yes
Fit statistics	
Observations	16,098
R^2	0.83702

Clustered (huc8_num) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Sample restricted to 2000–2009. Placebo post-period begins in 2005. A null coefficient supports the parallel trends assumption: high- and low-TMDL-coverage watersheds followed similar DO trajectories before the TMDL program's primary implementation period.

Table 4 reports the placebo test, restricting the sample to pre-2010 observations and testing for a differential trend at a placebo cutoff of 2005. The coefficient is positive (+0.13) and statistically insignificant ($p = 0.37$), confirming that high- and low-TMDL-coverage watersheds followed parallel dissolved oxygen trajectories before the TMDL program's primary implementation period. The contrast is stark: the placebo finds no differential pre-trend, while the main specification finds a significant negative post-2010 divergence.

6. Discussion

The negative finding carries a specific policy interpretation. TMDLs are regulatory plans—they set pollution budgets on paper. Whether those budgets translate into actual load reductions depends on a long implementation chain: NPDES permit holders must revise discharge limits, agricultural operations must adopt best management practices, municipalities must invest in stormwater infrastructure, and agencies must enforce compliance. The results suggest this chain is broken: at minimum, TMDLs have not produced measurable improvements in dissolved oxygen. The negative coefficient admits two interpretations. The first is the "paper tiger" reading: TMDL establishment consumes agency resources without triggering the downstream permit revisions and enforcement needed to reduce actual loads, and the most intensively regulated waterbodies continue to deteriorate. The second is a selection interpretation: despite station fixed effects absorbing permanent level differences, watersheds prioritized for TMDL completion may have experienced worsening pollution trends (e.g., from urbanization or agricultural intensification) that the design cannot fully absorb. The clean pre-2010 placebo argues against simple selection on pre-existing trends, but it cannot rule out trends that accelerated after 2010. I therefore interpret the negative coefficient as strong evidence that TMDLs do not improve water quality, with the negative magnitude providing suggestive but not definitive evidence of a "paper tiger" mechanism.

This interpretation aligns with the administrative burden framework of Herd and Moynihan (2018): environmental agencies may expend substantial resources producing TMDLs to satisfy consent decrees and federal requirements, diverting capacity from the enforcement and implementation activities that would actually reduce pollution. If the binding constraint on water quality improvement is not the absence of plans but the absence of implementation resources, the TMDL program may be counterproductive—consuming agency bandwidth without generating environmental benefit.

The finding does not imply that the Clean Water Act as a whole is ineffective. Keiser and Shapiro (2019) document substantial improvements in water quality following the CWA's point-source permit program (NPDES), which directly constrains discharges through enforceable limits. The TMDL program's weakness may lie in its reliance on nonpoint-source load allocations, which lack the legal enforceability of point-source permits (Craig, 2012). If so, the policy implication is clear: shift resources from TMDL planning toward enforceable mechanisms for nonpoint-source pollution reduction.

7. Conclusion

The Clean Water Act's TMDL program has produced over 80,000 regulatory documents specifying pollution budgets for impaired waters. This paper provides the first causal estimate of whether those documents improve the physical water quality they target. The answer is no—and the targeted waters may actually fare worse. Dissolved oxygen in high-TMDL-coverage watersheds declines by 0.50 mg/L relative to low-coverage watersheds after the primary implementation period, a result robust to alternative specifications and supported by clean pre-trends. The list gets cleaned; the water does not.

This finding reframes the regulatory evaluation question. The relevant measure of a regulatory program's effectiveness is not its output (plans produced, limits set, deadlines met) but its outcome (water made cleaner, ecosystems restored, health protected). Billions of dollars and decades of agency effort have been invested in a planning exercise whose environmental returns appear to be zero. The next frontier for water quality policy is not better plans but better implementation.

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

Christensen, Peter, David A Keiser, and Gabriel E Lade, “Environmental Regulation and the Race to the Bottom: Evidence from US State Government Bureaucracies,” *Review of Economics and Statistics*, 2023, *105* (4), 898–913.

Craig, Robin Kundis, “TMDLs, Interstate Pollution, and the Federal Common Law of Nuisance,” *Ecology Law Quarterly*, 2012, *39* (2), 373–414.

Currie, Janet, Joshua Graff Zivin, Jamie Mullins, and Matthew Neidell, “What Do We Know About Short- and Long-Term Effects of Early-Life Exposure to Pollution?,” *Annual Review of Resource Economics*, 2014, *6* (1), 217–247.

De Cicco, Laura A and Hirsch, Robert M and Lorenz, David and Watkins, William D and Johnson, Mike, “dataRetrieval: R Packages for Discovering and Retrieving Water Data Available from Federal Hydrologic Web Services,” U.S. Geological Survey 2024. R package version 2.7.18.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, “The Value of Regulatory Discretion: Estimates from Environmental Inspections in India,” *Econometrica*, 2018, *86* (6), 2123–2160.

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.

_ **and Rema Hanna**, “Environmental Regulations, Air and Water Pollution, and Infant Mortality in India,” *American Economic Review*, 2014, *104* (10), 3038–3072.

Herd, Pamela and Donald P Moynihan, *Administrative Burden: Policymaking by Other Means*, New York: Russell Sage Foundation, 2018.

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.

Liu, Yaoqi, Bernard A Engel, Dennis C Flanagan, Margaret W Gitau, Sara K McMillan, and Indrajeet Chaubey, “Assessing the Effectiveness of the TMDL Program and Its Contribution to Water Quality Improvement,” *Water Resources Research*, 2017, *53* (4), 3081–3098.

Shapiro, Joseph S and Reed Walker, "Why Is Pollution from US Manufacturing Declining? The Roles of Environmental Regulation, Productivity, and Trade," *American Economic Review*, 2018, *108* (12), 3814–3854.

A. Data Appendix

ATTAINS data. I access ATTAINS assessment data through EPA's GIS service (ArcGIS REST API).[1] The query extracts all assessment units in IR Categories 4A and 5 for Virginia and North Carolina from the Lines layer (layer 1) for segment-level data and the Points layer (layer 0) for geographic coordinates. Fields extracted include assessment unit identifiers, IR categories, overall assessment status, and reporting cycle.

Water Quality Portal data. DO readings are obtained from the Water Quality Portal (`https://www.waterqualitydata.us`) via the `dataRetrieval` R package (De Cicco, Laura A and Hirsch, Robert M and Lorenz, David and Watkins, William D and Johnson, Mike, 2024). Queries are structured by HUC-8 watershed to avoid API size limits. The characteristic name filter is "Dissolved oxygen (DO)" with sample media restricted to "Water" for the period January 2000 through December 2023.

Sample construction.

1. Extract all Category 4A and 5 assessment units from ATTAINS GIS for VA and NC.
2. Identify HUC-8 watersheds containing impaired segments by matching ATTAINS point coordinates to the nearest USGS monitoring station's HUC-8 code.
3. Query WQP for DO readings in each HUC-8 watershed, 2000–2023.
4. Clean: drop readings outside [0, 25] mg/L; drop station-years with fewer than 2 readings.
5. Aggregate to station-year means.
6. Compute TMDL share (4A/(4A+5)) for each HUC-8 from the ATTAINS 2022 reporting cycle.
7. Merge station-year panel with HUC-8 TMDL shares.

B. Robustness Appendix

Alternative clustering. The main specification clusters standard errors at the HUC-8 watershed level, which is the level of treatment variation. As additional checks, I cluster at the station level (more conservative for within-station serial correlation) and at the state level (extremely conservative, with only two clusters—reported for transparency but interpreted cautiously given the small number of clusters).

---

[1] URL: `https://gispub.epa.gov/arcgis/rest/services/OW/ATTAINS_Assessment/MapServer`

Table 5: Standardized Effect Sizes

	Specification	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled							
DO (cont.)	Continuous TMDL	-0.497	0.175	1.93	-0.257	0.091	Large neg.
DO (binary)	Binary high TMDL	-0.278	0.113	1.93	-0.144	0.059	Moderate neg.

Notes: **Country:** United States. **Research question:** Does TMDL establishment under CWA Section 303(d) improve dissolved oxygen levels in impaired waterbody segments? **Policy mechanism:** TMDLs set maximum allowable pollutant loading for impaired waters, triggering permit revisions (NPDES), best management practices, and nonpoint source reduction plans that collectively reduce pollution inflows. **Outcome definition:** Mean annual dissolved oxygen (mg/L) at USGS monitoring stations, measuring the oxygen available in surface water for aquatic life. **Treatment:** Continuous—share of impaired assessment units in the station's HUC8 watershed with completed TMDLs (IR Category 4A vs. 5). **Data:** EPA Water Quality Portal and ATTAINS, 2000–2023, station-year panel. **Method:** TWFE with station and year fixed effects, standard errors clustered at HUC8. **Sample:** Stations in HUC8 watersheds containing impaired assessment units in Virginia and North Carolina; restricted to stations with $\geq$2 annual readings. SDE $= \hat{\beta}/\text{SD}(Y)$ where SD(Y) is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).

Leave-one-out. I re-estimate the main specification dropping each HUC-8 in turn to ensure no single watershed drives the null result.

C. Standardized Effect Sizes