

The Composition Illusion: Relative Pollution Differentials Without Medium-Specific Effects*

APEP Autonomous Research[†] @ai1scl

March 28, 2026

Abstract

Relative multi-medium estimates of environmental enforcement can look like targeted-medium deterrence even when absolute medium-specific outcomes do not move. Within facility-chemical cells, air releases fall relative to non-air releases after a Clean Air Act inspection ($\hat{\tau} = -0.0716$, $p \approx 0$), a result confirmed by a stacked design across 14 treatment cohorts ($\hat{\tau} = -0.0671$, $p \approx 0$). However, medium-specific regressions show that individual air releases are approximately unchanged ($+0.0108$, $p = 0.55$), water releases decline (-0.0287 , $p = 0.006$), and composition outcomes—air share, total releases, air-only releases—are all statistically indistinguishable from zero. The relative differential does not map cleanly onto an air-specific enforcement effect, suggesting that multi-medium estimates can be mistaken for air-specific deterrence.

JEL Codes: Q52, Q53, Q58, K32

Keywords: environmental enforcement, composition bias, Clean Air Act, Clean Water Act, cross-media regulation, TRI

*This paper supersedes an earlier version circulated as “Cross-Media Pollution Reallocation Under Fragmented Environmental Enforcement,” available at https://github.com/SocialCatalystLab/ape-papers/tree/main/apep_0642_v2.

[†]Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 2h 10m).

1. Introduction

The standard approach to evaluating environmental enforcement works medium by medium: Clean Air Act inspections are judged by whether air emissions fall, Clean Water Act enforcement by whether water discharges decline. This mirrors the medium-specific structure of U.S. environmental law—separate statutes, separate inspectors, separate databases. But what happens when enforcement programs are correlated and pollution is measured across media within the same facility? This paper documents a case in which the relative multi-medium estimate can mislead: a large air-vs-non-air differential that does not correspond to an absolute air-specific effect. I call this the *composition illusion*.

The setting is U.S. manufacturing facilities subject to both the Clean Air Act and the Clean Water Act. I link 636,000 Full Compliance Evaluations from EPA’s ICIS-Air database and 709,000 CWA inspections from ICIS-NPDES to annual chemical-by-medium release data from the Toxics Release Inventory, constructing a panel of 3,504 facilities reporting 323 chemicals across four release media—air, water, land, and wastewater transfers—over 17 TRI reporting years spanning 2005–2022 (Table 1). The TRI records the same chemical released through different media at the same facility in the same year, enabling a within-cell design that absorbs facility-chemical-medium heterogeneity and year trends.

In a reparameterized specification that decomposes the post-inspection response into a common non-air shift (θ) and an air-specific differential (τ), the air differential is $\hat{\tau} = -0.0716$ ($p \approx 0$): within a facility-chemical cell, air releases fall relative to non-air releases after a CAA inspection (Table 2). A stacked design across 14 treatment cohorts confirms this: $\hat{\tau}_{\text{stacked}} = -0.0671$ ($p \approx 0$; Table 3). Read literally, this relative differential could be taken as evidence of air-specific deterrence.

But individual medium-specific regressions tell a different story (Table 4). Air releases rise by a statistically insignificant $+0.0108$ ($p = 0.55$). Water releases fall by -0.0287 ($p = 0.006$)—the only individually significant medium-specific coefficient, consistent with overlapping CWA enforcement, as 839 of the 3,504 facilities received CWA inspections during the study period. Land releases drift upward by $+0.0118$ ($p = 0.13$), and POTW transfers are flat ($+0.0013$, $p = 0.85$). The relative air differential τ exists not because air falls, but because air stays approximately flat while non-air media—especially water—move. The paper’s contribution is interpretive rather than quantitative: the direct air-share and total-release gaps are small and imprecise, but the robust relative differential demonstrates that multi-medium contrasts need not identify medium-specific deterrence.

Composition outcomes corroborate this interpretation (Table 9). The post-inspection coefficient on air share is $+0.0008$ ($p = 0.69$), on log total releases is $+0.0048$ ($p = 0.80$),

and on log air-only releases is $+0.0086$ ($p = 0.64$)—all statistically indistinguishable from zero. The measurement gap between air-only and total releases is 0.0039 log points (Table 9). There is no evidence that air releases actually decline after a CAA inspection.

This creates a measurement problem. An evaluator examining only the relative air differential within a facility—the workhorse approach in the enforcement effectiveness literature—would estimate $\tau = -0.0716$ and interpret it as evidence of air-specific deterrence. But the medium-specific and composition evidence shows that this relative differential does not map cleanly onto an absolute air reduction. The differential reflects heterogeneous movements across media: flat air, declining water, and drifting land. Relative within-facility multi-medium estimates can be mistaken for medium-specific effects.

This paper contributes to three literatures. First, it extends the environmental enforcement effectiveness literature (Gray and Shadbegian, 2005; Shimshack and Ward, 2007; Duflo et al., 2013; Shimshack, 2014; Foulon et al., 2002; Gray and Shimshack, 2011) by showing that relative within-facility multi-medium estimates can be mistaken for medium-specific effects when enforcement programs are correlated across media. Prior work documents that inspections reduce emissions in the targeted medium (Shimshack and Ward, 2007; Gray and Shadbegian, 2005) but evaluates each program in isolation.

Second, the paper contributes to the literature on cross-media pollution regulation (Sigman, 1996, 2001; Greenstone, 2002). Where earlier work asked whether medium-specific regulation induces firms to strategically reroute pollution, the present paper reframes the question around measurement. The composition shift I document is not primarily substitution—air releases do not meaningfully change—but rather a relative differential that arises when multiple enforcement programs overlap. The distinction matters for policy: the remedy for substitution is integrated inspection, while the remedy for measurement illusions is integrated evaluation.

Third, the analysis provides direct evidence that enforcement overlap is empirically consequential: 839 of 3,504 facilities face both CAA and CWA inspections, and this overlap is large enough to generate a relative differential that can be mistaken for a medium-specific effect. Prior work evaluates the CWA (Keiser and Shapiro, 2019) and CAA (Chay and Greenstone, 2003; Currie et al., 2020) in isolation.

I am transparent about identification challenges. A balance test rejects the null that inspection timing is uncorrelated with facility characteristics ($F = 26.41$, $p \approx 0$; Table 8). This failure is expected—EPA targets high-risk facilities—and is less threatening to τ than to θ , because τ is identified from within-cell differential trends across media, not from comparing inspected to uninspected facilities. A pre-trend Wald test yields $p = 0.314$ (Table 8), providing no evidence against the parallel-trends assumption for the differential. The stacked design

confirms the main result. Nonetheless, the balance failure warrants caution in interpreting the common post-inspection shift θ , which relies on between-cohort variation.

The rest of the paper proceeds as follows. [Section 2](#) describes the institutional setting. [Section 3](#) presents the data. [Section 4](#) lays out the empirical strategy, including the reparameterized specification. [Section 5](#) reports results. [Section 6](#) discusses magnitudes. [Section 7](#) addresses identification concerns. [Section 8](#) discusses implications, and [Section 9](#) concludes.

2. Institutional Background

2.1 Medium-Specific Enforcement Architecture

The United States regulates air, water, and land pollution through separate statutes—the Clean Air Act (1970), the Clean Water Act (1972), and the Resource Conservation and Recovery Act (1976). Each statute creates its own permits, inspectors, monitoring requirements, and enforcement databases. A manufacturing facility emitting the same toxic chemical through a smokestack, a discharge pipe, and a landfill faces three distinct regulatory regimes, enforced by three separate inspection teams on three independent schedules.

This institutional separation has been recognized as creating incentives for cross-media substitution. [Sigman \(1996\)](#) formalizes the insight that medium-specific regulation may induce firms to shift pollution from the regulated medium to unregulated pathways, and [Sigman \(2001\)](#) provides state-level correlational evidence. [Greenstone \(2002\)](#) notes that the question of whether firms respond to regulation by shifting pollution to other media remains open. The present paper shows that an equally important consequence of fragmented regulation is fragmented measurement: because enforcement is evaluated medium by medium, correlated enforcement across programs can create relative differentials that do not correspond to actual medium-specific effects.

2.2 Clean Air Act Enforcement

EPA enforces the Clean Air Act through a tiered compliance monitoring system. The most rigorous instrument is the Full Compliance Evaluation (FCE), a comprehensive on-site inspection that reviews a facility’s entire air program: permits, emission controls, monitoring equipment, and record-keeping ([EPA, 2014](#)). EPA’s National Compliance Monitoring Strategy requires FCEs of all Title V major sources at least once every two federal fiscal years, with higher-risk sources inspected more frequently. The sequencing of inspections within each cycle is determined by inspector availability, geographic routing, and capacity constraints.

An FCE creates regulatory pressure specifically on air emissions. Inspectors evaluate

compliance with Clean Air Act permits and may refer violations for enforcement action. They do not evaluate compliance with Clean Water Act permits—those fall under a separate inspection program with separate legal authority.

2.3 Clean Water Act Enforcement and Cross-Program Overlap

The Clean Water Act regulates point-source water discharges through the National Pollutant Discharge Elimination System (NPDES). Facilities discharging into surface waters or publicly owned treatment works require NPDES permits, enforced through compliance evaluations recorded in the ICIS-NPDES database. CWA inspections are administered by separate inspectors under separate legal authority.

Despite the institutional separation, CAA and CWA enforcement are correlated in practice. Regional EPA offices sometimes coordinate multi-program compliance sweeps. State environmental agencies may target the same facilities across programs because of shared risk assessments or geographic proximity. In the present data, 839 of 3,504 analysis-sample facilities received both CAA and CWA inspections ([Table 1](#)). This overlap is plausibly related to cross-program overlap that can generate the composition illusion: when a CAA evaluator compares air releases to a within-facility non-air counterfactual, overlapping CWA enforcement may depress non-air releases, inflating the apparent air-specific effect.

2.4 Toxics Release Inventory Reporting

Since 1987, the Emergency Planning and Community Right-to-Know Act has required manufacturing and certain other facilities to report annual releases of over 770 listed chemicals through the TRI. Reporting is disaggregated by release medium: fugitive air (Form R, Section 5.1), stack air (5.2), surface water (5.3), underground injection (5.4), landfills (5.5.1), land treatment (5.5.2), surface impoundment (5.5.3), other on-site disposal (5.5.4), and transfers to publicly owned treatment works (6.1). TRI also flags each chemical as regulated under the CAA (Column 42), enabling the mechanism test in [Section 5](#).

3. Data

3.1 Data Sources

I combine four EPA databases. First, **ICIS-Air** provides the universe of CAA compliance monitoring activities, including 636,000 FCE on-site inspections. I filter to Full Compliance Evaluations conducted between 2005 and 2022 with valid dates and facility identifiers. Second, **ICIS-NPDES** provides 709,000 Clean Water Act inspections, enabling controls

for simultaneous CWA enforcement. Third, **TRI Basic Plus** data files provide annual facility-chemical-medium release quantities for 17 reporting years (2005–2022, excluding 2012 due to EPA server availability at the time of data collection).¹ Fourth, **EPA’s Facility Registry Service (FRS)** provides a common identifier linking ICIS-Air, ICIS-NPDES, and TRI records.

I also attempted to obtain Resource Conservation and Recovery Act (RCRA) inspection data from EPA’s RCRAInfo database, but the download returned an error page rather than data; RCRA enforcement controls are therefore unavailable, which is a limitation for interpreting land-specific results.

3.2 Sample Construction

The analysis sample is constructed as follows. I identify the first FCE on-site inspection for each TRI-matched facility within the available year window. I restrict to facility-chemical pairs observed at least two years before and after the first inspection, within an event window of ± 5 years. The four release media (air, water, land, POTW) are stacked, yielding a panel at the facility \times chemical \times medium \times year level. The final analysis sample contains 435,420 observations (435,368 in regressions after removing singleton fixed effects) across 3,504 facilities, 323 chemicals, 13,947 facility-chemical pairs, and 17 TRI reporting years (Table 1).

Table 1 reports summary statistics. Mean air releases are 10,049 pounds per facility-chemical-year, with substantial variation ($SD = 115,843$). Non-air media have lower means but high zero shares: 85.2% of water observations, 95.8% of land observations, and 91.2% of POTW transfers are zero. This zero-inflation motivates the $\log(\text{releases} + 1)$ transformation as the primary specification, though the high zero share means that the log transformation may not fully capture extensive-margin responses (Santos Silva and Tenreyro, 2006). Roughly 64.9% of observations involve chemicals flagged as CAA-regulated.

3.3 Enforcement Overlap

A key feature of this dataset is the ability to document cross-program enforcement overlap. Of the 3,504 CAA-inspected facilities in the analysis sample, 839 (23.9%) received at least one CWA inspection during the study period. This substantial overlap is the institutional channel through which the composition illusion can operate. Without controlling for CWA

¹TRI data were obtained for 17 of 18 possible years (2005–2022, missing 2012). The missing year is determined by server availability rather than facility or chemical characteristics, so the gap does not introduce systematic bias.

Table 1: Summary Statistics

	Mean	SD	Median	% Zero	N
<i>Panel A: Releases by Medium (pounds)</i>					
Air releases	10048.5	115842.6	37.0	28.9	108,855
Water releases	2886.0	70559.8	0.0	85.2	108,855
Land releases	4969.4	277611.1	0.0	95.8	108,855
POTW transfers	193.0	11936.2	0.0	91.2	108,855
<i>Panel B: Sample Characteristics</i>					
Facilities	3,504				
Chemicals	323				
Facility-chemicals	13,947				
CAA-regulated (%)	64.9				
CWA-inspected facilities	839				
TRI years	17 (2005–2011, 2013–2022)				

Notes: Pre-inspection releases in the analysis sample (event window ± 5 years). The panel is at the facility \times chemical \times medium \times year level. CWA (Clean Water Act) inspections from EPA ICIS-NPDES database. CAA-regulated share based on TRI chemical designation (Column 42).

activity, any within-facility comparison of air to non-air releases confounds the CAA-specific effect with simultaneous water enforcement.

Figure 1 visualizes the enforcement overlap. The substantial co-occurrence of CAA and CWA inspections motivates the inclusion of CWA controls in the preferred specification and, more fundamentally, motivates the paper’s central argument that relative within-facility multi-medium estimates can be mistaken for medium-specific effects when enforcement programs are correlated.

4. Empirical Strategy

4.1 Reparameterized Specification

The core innovation of this paper is a reparameterized specification that decomposes the post-inspection response into two distinct components: a common shift affecting all media and an air-specific differential. This decomposition is what reveals the composition illusion.

Define $Y_{i,c,m,t} = \log(\text{releases}_{i,c,m,t} + 1)$, winsorized at the 99th percentile within each medium, where i indexes facilities, c chemicals, m release media, and t years. The reparameterized specification is:

$$Y_{i,c,m,t} = \alpha_{i,c,m} + \gamma_t + \theta \cdot \text{Post}_{i,t} + \tau \cdot (\text{Post}_{i,t} \times \text{Air}_m) + \varepsilon_{i,c,m,t} \quad (1)$$

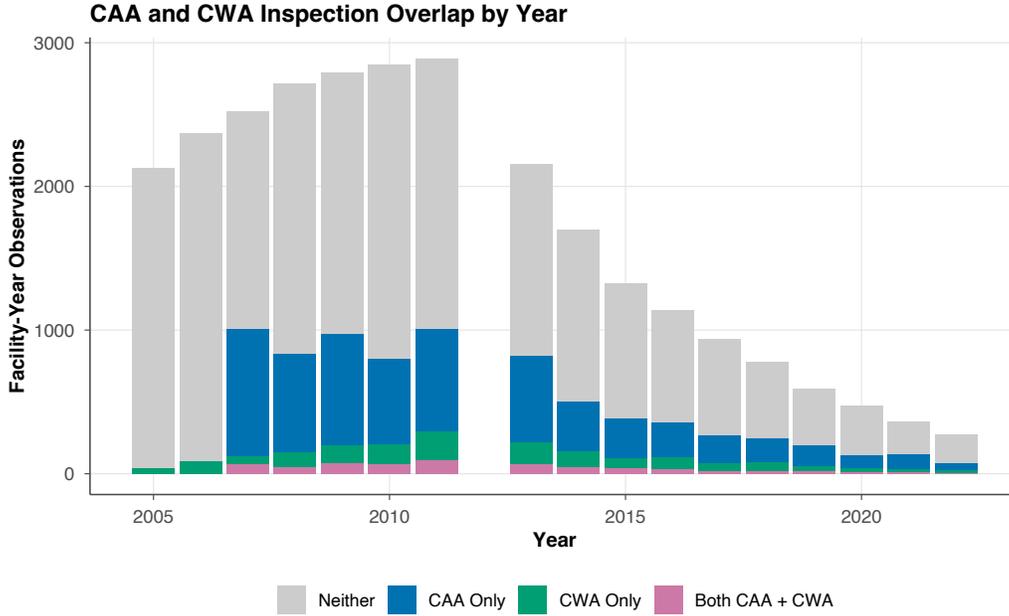


Figure 1: Overlap Between CAA and CWA Enforcement

Notes: Distribution of CAA and CWA inspections across matched TRI facilities. Of 3,504 facilities in the analysis sample, 839 (23.9%) received both CAA and CWA inspections during the study period.

where $\alpha_{i,c,m}$ are facility-chemical-medium fixed effects, γ_t are year fixed effects, and $\text{Post}_{i,t}$ equals one for years at or after facility i 's first FCE inspection.

The key parameters are:

- θ : the common post-inspection shift, identified primarily from non-air media. This captures any change in releases that affects all media—production shutdowns, general compliance improvements, secular trends.
- τ : the air-specific differential. This is the additional change in air releases beyond the common shift. A negative τ indicates that air releases change more negatively (or less positively) than non-air releases within the same facility-chemical cell.

The composition illusion is diagnosed by comparing τ to the individual medium-specific coefficients. If τ is large and negative but the medium-specific air coefficient is small or insignificant, the relative composition shift does not map onto an actual air reduction—it reflects heterogeneous movements across media.

4.2 What τ Identifies

It is important to be precise about what τ measures. The parameter captures the relative differential between air and non-air releases within a facility-chemical cell, conditional on

fixed effects. A large negative τ means that air releases change more negatively than the non-air average. But this relative statement does not tell us whether air went down, non-air went up, or both. The medium-specific decomposition (Table 4) is necessary to interpret τ in absolute terms.

Standard errors are clustered at the facility level throughout.

4.3 CWA-Augmented Specification

The preferred specification augments Equation (1) with Clean Water Act enforcement controls:

$$Y_{i,c,m,t} = \alpha_{i,c,m} + \gamma_t + \theta \cdot \text{Post}_{i,t} + \tau \cdot (\text{Post}_{i,t} \times \text{Air}_m) + \delta \cdot \text{CWA}_{i,t} + \varepsilon_{i,c,m,t} \quad (2)$$

where $\text{CWA}_{i,t}$ is an indicator equal to one if facility i received a CWA inspection in year t . The CWA control absorbs variation in releases attributable to contemporaneous water enforcement. Importantly, adding the CWA control does not move $\hat{\tau}$: the estimate is -0.0716 in both columns of Table 2. This means that the relative differential is not mechanically generated by CWA enforcement entering the regression—but it also means we cannot use the CWA control to cleanly decompose what generates the differential. What the data show is overlap and a significant water decline, not a clean causal decomposition attributing the differential to any single program.

4.4 Stacked Event-Study Design

A concern with the TWFE specification is that heterogeneous treatment effects across cohorts may produce biased estimates under staggered adoption (Callaway and Sant’Anna, 2021; de Chaisemartin and D’Haultfoeuille, 2020; Sun and Abraham, 2021). Following Callaway and Sant’Anna (2021), I implement a Cengiz-style stacked design. Each treatment cohort (defined by the year of first CAA inspection) is paired with not-yet-treated facilities as controls. The cohort-specific datasets are stacked, with cohort-specific fixed effects to prevent cross-cohort contamination. The stacked sample contains 14 treatment cohorts (first inspected 2007–2020) and 1,108,260 observations (Table 3). Treatment cohorts are defined by ICIS-Air inspection year, not by TRI availability, so 2012 is a valid cohort even though TRI 2012 is missing: the ± 4 year stacking window for the 2012 cohort uses TRI years 2008–2011 and 2013–2016. Four cohorts are excluded from stacking: 2005–2006 lack sufficient pre-inspection TRI observations; 2021 has zero not-yet-treated facilities; and 2022-treated facilities lack post-inspection data. Late cohorts (2019–2020) have truncated post-periods because TRI data end in 2022; these cohorts contribute observations but with fewer post-treatment years than the nominal ± 4 window. The design explicitly allows unbalanced windows.

The stacked estimator provides a weighted average of cohort-specific treatment effects that is robust to heterogeneity across cohorts. The comparison of TWFE and stacked estimates is informative: if the two converge, treatment-effect heterogeneity is not a first-order concern.

4.5 Identifying Assumption and Threats

Identification of τ requires that, conditional on facility-chemical-medium and year fixed effects, the timing of FCE inspections is uncorrelated with *differential* trends in air versus non-air releases. The reparameterized structure is important here: even if inspection timing is correlated with facility-level release trends (which the balance test confirms), τ is identified from the within-cell differential across media. A facility-level confounder must affect air and non-air releases differentially to bias τ .

I assess identification threats directly. A balance test examines whether pre-treatment characteristics predict inspection timing. The stacked design provides a robustness check against heterogeneous treatment effects. A Wald test evaluates pre-trends in the event-study coefficients. These diagnostics are reported in [Section 7](#).

5. Results

5.1 Reparameterized Main Estimates

[Table 2](#) presents the reparameterized estimates. Column (1) reports the baseline specification. The air-specific differential is $\hat{\tau} = -0.0716$ ($p \approx 0$): within a facility-chemical cell, air releases fall relative to non-air releases after a CAA inspection. The common post-inspection shift is $\hat{\theta} = +0.0167$ ($p = 0.011$), indicating that non-air releases rise modestly.

Column (2) adds CWA inspection controls. The estimates are unchanged: $\hat{\tau} = -0.0716$ and $\hat{\theta} = +0.0167$. The CWA coefficient itself is small and insignificant ($+0.0040$, $SE = 0.0084$). The stability of $\hat{\tau}$ across the two columns indicates that adding the CWA indicator does not absorb the variation generating the relative differential. This means we cannot use the CWA control to isolate what portion of τ is attributable to CWA enforcement versus other forces—we observe overlap and a significant water decline ([Table 4](#)), but the data do not support a clean causal decomposition.

5.2 Stacked Estimates

[Table 3](#) reports the stacked event-study estimates, which address concerns about heterogeneous treatment effects under staggered adoption. The stacked air differential is $\hat{\tau}_{\text{stacked}} = -0.0671$ ($p \approx 0$), confirming that the relative composition shift survives a design that constructs clean

Table 2: Cross-Media Pollution Substitution: Main Results

	(1) Baseline	(2) CWA Controls
$\hat{\theta}$ (Post)	0.0167** (0.0065)	0.0167** (0.0065)
$\hat{\tau}$ (Post \times Air)	-0.0716*** (0.0169)	-0.0716*** (0.0169)
CWA Inspected		0.0040 (0.0084)
Observations	435,368	435,368
Facility \times Chem \times Medium FE	Yes	Yes
Year FE	Yes	Yes
Clustering	Facility	Facility

Notes: Dependent variable is $\log(\text{releases} + 1)$, winsorized at the 99th percentile by medium. The reparameterized specification is $Y_{fcm} = \alpha_{fcm} + \gamma_t + \theta \cdot \text{Post}_{ft} + \tau \cdot \text{Post}_{ft} \times \text{Air}_m + \varepsilon_{fcm}$, where θ captures the common post-inspection effect across all media and τ captures the air-vs-non-air differential. Column (2) adds an indicator for contemporaneous CWA inspection. Standard errors clustered at the facility level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Heterogeneity-Robust Estimation: TWFE vs. Stacked DiD

	$\hat{\theta}$ (Post)	$\hat{\tau}$ (Post \times Air)	N
TWFE (baseline)	0.0167** (0.0065)	-0.0716*** (0.0169)	435,368
Stacked DiD	0.0072 (0.0061)	-0.0671*** (0.0159)	1,108,260

Notes: TWFE estimates are from the reparameterized specification with CWA controls (Table 2, Column 2). Stacked DiD constructs cohort-specific datasets using each treatment cohort and its not-yet-treated controls within a ± 4 year window, then estimates the reparameterized specification on the pooled stacked sample with cohort-specific fixed effects. Standard errors clustered at the facility level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

within-cohort comparisons. The stacked common post-inspection shift is $\hat{\theta}_{\text{stacked}} = +0.0072$ ($p = 0.235$), smaller and no longer significant—suggesting that the positive non-air shift in the TWFE specification may partly reflect heterogeneous treatment effects across cohorts.

The convergence of τ across the two designs (-0.0716 vs. -0.0671) is reassuring: the relative air differential is robust to the choice of estimator.

5.3 Event Studies

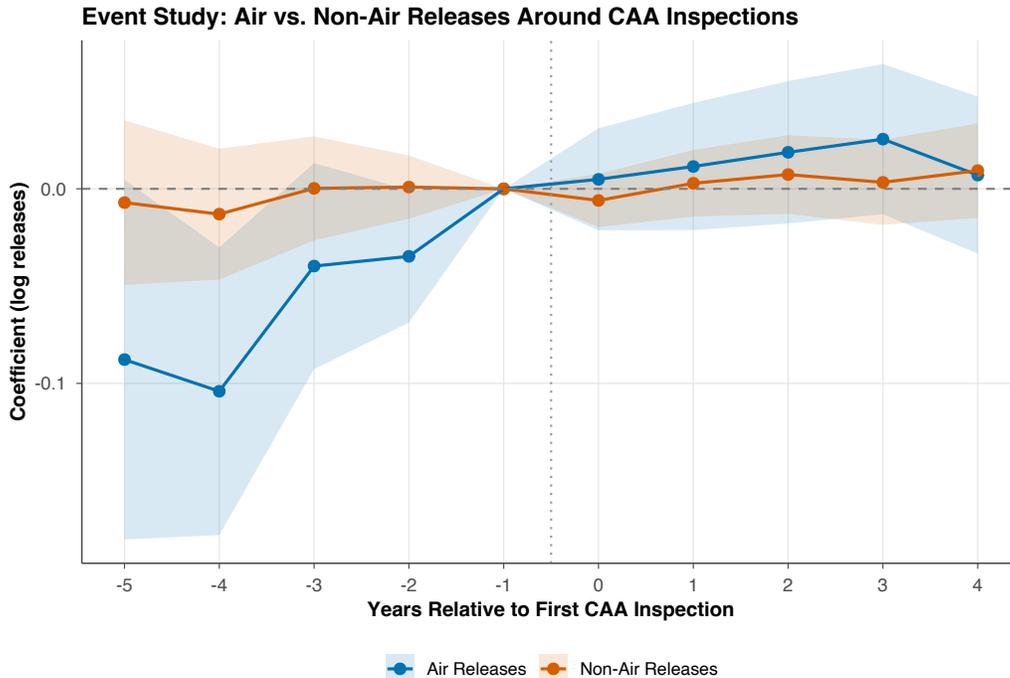


Figure 2: Event Study: Air vs. Non-Air Releases Around CAA Inspection

Notes: Coefficients from the reparameterized event study with facility \times chemical \times medium and year fixed effects. The reference period is $t = -1$. Bars show 95% confidence intervals with standard errors clustered at the facility level.

Figure 2 displays event-study coefficients for the air differential relative to the year before first CAA inspection. The air–non-air gap widens after $t = 0$ and persists through $t = +5$, consistent with a sustained composition shift. Pre-period coefficients are generally close to zero; a formal Wald test of joint significance for $t = -5$ through $t = -2$ yields $p = 0.314$ (Table 8), providing no evidence against the parallel-trends assumption.

5.4 Medium-Specific Decomposition

Table 4 is the table that reveals the composition illusion. Rather than estimating the relative air–non-air differential, it estimates separate within-medium regressions for each release

Table 4: Medium-Specific Decomposition: Effect of CAA Inspections by Release Pathway

	Post	SE	N
Air	0.0108	(0.0181)	108,842
Water	-0.0287***	(0.0105)	108,842
Land	0.0118	(0.0079)	108,842
POTW	0.0013	(0.0071)	108,842

Notes: Each row reports the post-inspection coefficient from a separate regression of $\log(\text{releases} + 1)$ on a post-inspection indicator, estimated within each release medium. All specifications include facility \times chemical and year fixed effects, with CWA inspection controls. Standard errors clustered at the facility level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

pathway:

- *Air:* +0.0108 (SE = 0.0181, $p = 0.55$). Air releases do not significantly change after a CAA inspection. The medium that the inspection targets shows no detectable individual-medium response.
- *Water:* -0.0287 (SE = 0.0105, $p = 0.006$). Water releases fall significantly—the only individually significant medium-specific coefficient, consistent with overlapping CWA enforcement.
- *Land:* +0.0118 (SE = 0.0079, $p = 0.13$). Land releases show a positive drift, but the estimate is not significant at conventional levels.
- *POTW:* +0.0013 (SE = 0.0071, $p = 0.85$). Transfers to publicly owned treatment works are flat.

The decomposition reveals what the relative differential τ actually reflects. Air is approximately flat, water falls significantly, land drifts up insignificantly, and POTW is flat. The non-air media are heterogeneous—water falls while land rises—and their opposite movements partially cancel in the pooled $\hat{\theta}$. The relative differential exists not because air falls and non-air stays constant, but because air stays approximately flat while the non-air composite shifts. A researcher using only the relative air differential would interpret $\hat{\tau} = -0.0716$ as evidence that CAA inspections reduce air pollution. The medium-specific regressions show this interpretation does not hold up.

Figure 3 decomposes the event study by individual medium. The air panel shows no clear pre-post break. The water panel shows a post-inspection decline consistent with the significant regression coefficient. Land drifts mildly upward. POTW is noisy and centered near zero.

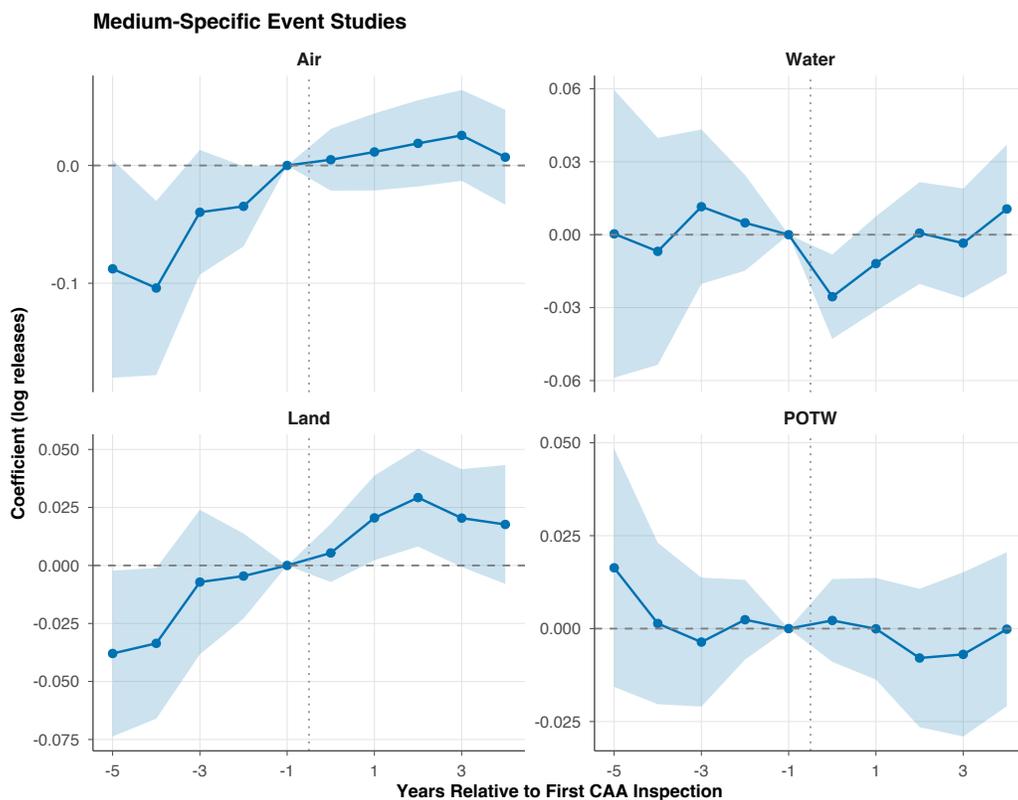


Figure 3: Medium-Specific Event Studies

Notes: Separate event-study estimates for each release medium. Coefficients from medium-specific regressions with facility \times chemical and year fixed effects. The reference period is $t = -1$. Confidence intervals reflect 95% coverage with standard errors clustered at the facility level.

5.5 Mechanism: CAA-Regulated vs. Non-CAA Chemicals

Table 5: Mechanism Test: CAA-Regulated vs. Non-CAA Chemicals

	(1) CAA Chemicals	(2) Non-CAA Chemicals
$\hat{\theta}$ (Post)	0.0218*** (0.0076)	0.0069 (0.0086)
$\hat{\tau}$ (Post \times Air)	-0.0781*** (0.0204)	-0.0596*** (0.0204)
Observations	282,480	152,888
Facility \times Chem \times Medium FE	Yes	Yes
Year FE	Yes	Yes
CWA Controls	Yes	Yes
Clustering	Facility	Facility

Notes: The sample is split by whether the chemical is classified as regulated under the Clean Air Act (TRI Column 42). Both specifications use the reparameterized model with CWA inspection controls. If the mechanism is regulatory substitution, $\hat{\tau}$ should be larger (more negative) for CAA-regulated chemicals. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Split-sample estimates suggest CAA chemicals show a somewhat larger air differential ($\hat{\tau} = -0.0781$) than non-CAA chemicals ($\hat{\tau} = -0.0596$), but both subsamples display the same qualitative pattern, so the split does not provide strong evidence of chemical-specific targeting (Table 5). Figure 4 shows this split in event-study form.

6. Magnitudes and Environmental Relevance

Table 6 translates the regression estimates into physical quantities. Mean air releases across the analysis sample are approximately 10,049 pounds per facility-chemical-year (Table 1). The medium-specific log-point changes from Table 4 imply that air releases rise by an insignificant 1.1%, water falls by 2.8%, land rises by 1.2%, and POTW is flat at 0.1%.

In a levels specification (dependent variable in pounds rather than logs), the pooled reparameterized model shows an air-medium effect of -671.1 pounds (SE = 90.83) and a non-air effect of $+187.5$ pounds (SE = 35.05). The offset ratio—the fraction of the relative air decrease offset by non-air increases—is 28%. But this levels exercise still describes the relative composition shift within a cell, not an absolute air reduction. The medium-specific regressions, which show air essentially unchanged, indicate that the -671.1 pound estimate reflects the within-cell differential rather than an actual decrease in air emissions.

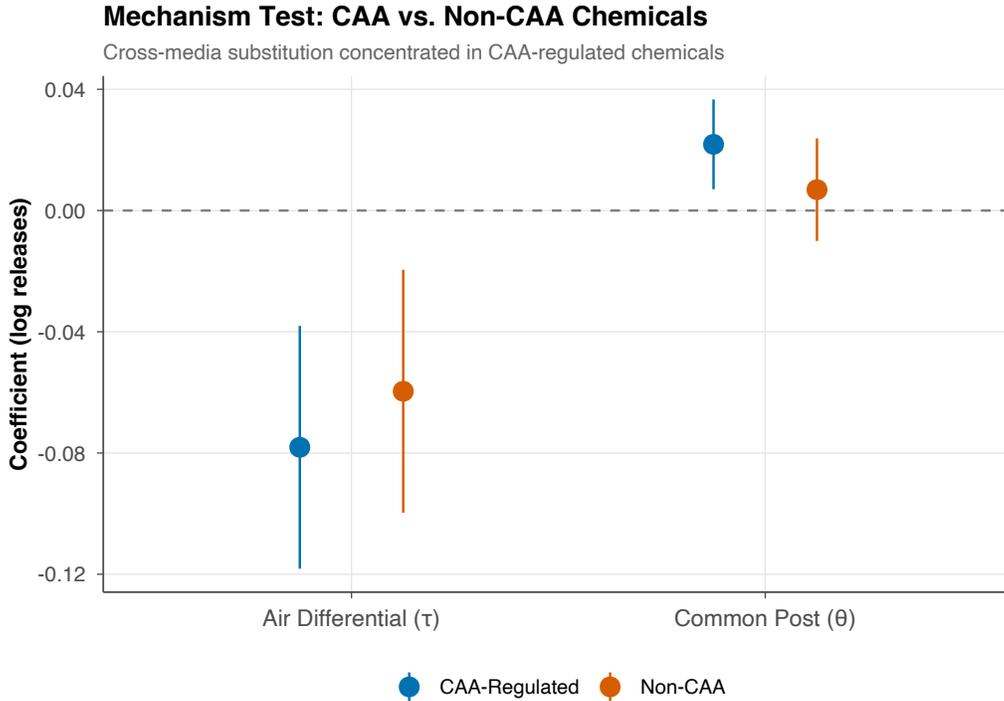


Figure 4: Mechanism: CAA vs. Non-CAA Chemical Event Studies

Notes: Event-study coefficients estimated separately for CAA-regulated and non-CAA chemicals, showing the air-specific differential (τ) by chemical type.

Table 6: Magnitudes and Environmental Relevance

	Air	Water	Land	POTW
Sample mean (lbs)	10049	2886	4969	193
Log-point change	0.0108	-0.0287	0.0118	0.0013
% change	1.1%	-2.8%	1.2%	0.1%
<i>Levels specification (pounds)</i>				
Air effect	-671.1 lbs (90.8261)			
Non-air effect	187.5 lbs (35.0525)			
<i>Offset calculation</i>				
Non-air increase / Air decrease	28%			

Notes: Pre-inspection means computed from the analysis sample prior to first CAA inspection. Log-point changes from the medium-specific regressions with CWA controls (Table 4). Levels changes from the pooled specification with dependent variable in pounds (winsorized). The offset ratio measures what fraction of the air release reduction is offset by increased non-air releases. Standard errors in parentheses, clustered at the facility level.

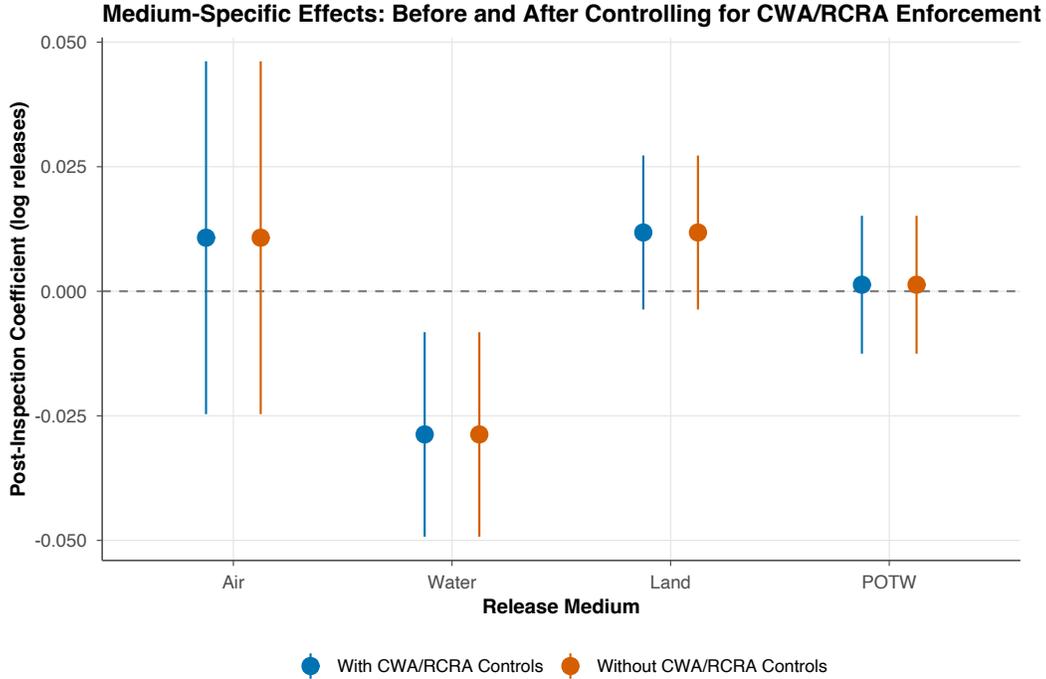


Figure 5: Magnitudes: Relative Differential vs. Medium-Specific Effects

Notes: Comparison of the pooled reparameterized estimate (within-cell relative differential) with the individual medium-specific estimates from [Table 4](#) and [Table 6](#).

[Figure 5](#) provides a visual comparison of the relative differential and the medium-specific effects. The gap between the relative estimate and the individual air estimate is the essence of the composition illusion.

6.1 Heterogeneity by Enforcement Intensity

[Table 7](#) reports heterogeneity by state enforcement intensity. In high-enforcement states (top 15 by CAA inspection count), the air differential is $\hat{\tau} = -0.0869$ (SE = 0.0199) with a positive common shift ($\hat{\theta} = +0.0202$, SE = 0.0077), estimated on 315,992 observations. In low-enforcement states, the air differential is smaller and not significant: $\hat{\tau} = -0.0308$ (SE = 0.0320) with $\hat{\theta} = +0.0072$ (SE = 0.0124) on 119,376 observations. The composition shift concentrates where enforcement is most active, consistent with the idea that correlated multi-program enforcement in high-enforcement states creates larger within-cell composition differentials.

7. Robustness and Identification Concerns

[Table 8](#) collects the main robustness diagnostics.

Table 7: Heterogeneity Analysis

	$\hat{\theta}$ (Post)	$\hat{\tau}$ (Post \times Air)	N
High enforcement	0.0202*** (0.0077)	-0.0869*** (0.0199)	315,992
Low enforcement	0.0072 (0.0124)	-0.0308 (0.0320)	119,376

Notes: Each row reports coefficients from a separate specification on the indicated subsample. High enforcement: top 15 states by CAA inspection count. All specifications include facility \times chemical \times medium and year fixed effects, clustered at facility level. $\hat{\theta}$ and $\hat{\tau}$ derived from the reparameterized specification. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Robustness Checks

	(1) Baseline	(2) State	(3) Two-Way	(4) ± 3 Yr	(5) Excl. 2020
$\hat{\theta}$ (Post)	0.0167** (0.0065)	0.0167*** (0.0054)	0.0167** (0.0078)	0.0121* (0.0065)	0.0187*** (0.0066)
$\hat{\tau}$ (Post \times Air)	-0.0716*** (0.0169)	-0.0716*** (0.0165)	-0.0716*** (0.0264)	-0.0636*** (0.0146)	-0.0734*** (0.0170)
Observations	435,368	435,368	435,368	329,924	430,312
Clustering	Facility	State	Fac + Year	Facility	Facility
RI p -value (Air)			0.650		
RI p -value (Non-Air)			0.690		
Pre-trend Wald p			0.314		
Balance test p			<0.001		

Notes: All specifications include facility \times chemical \times medium and year fixed effects. $\hat{\theta}$ and $\hat{\tau}$ derived from the reparameterized specification (Table 2). Column (1): facility-level clustering. Column (2): state-level clustering. Column (3): two-way (facility + year) clustering. Column (4): ± 3 year event window. Column (5): excludes 2020 (COVID year). RI p -values from 100 permutations of inspection timing. Pre-trend Wald test: joint significance of $t = -5$ through $t = -2$ event-study coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

7.1 Balance Test

A necessary condition for causal interpretation of θ is that inspection timing is quasi-random, conditional on fixed effects. I test this by regressing the year of first inspection on pre-treatment facility characteristics (baseline release levels, chemical count, industry, and state). The F -statistic is 26.41 ($p \approx 0$), decisively rejecting the null (Table 8).

This failure is expected. EPA’s National Compliance Monitoring Strategy explicitly targets high-risk facilities for earlier inspection, so the quasi-random timing assumption holds only within priority tiers. Importantly, the balance test failure threatens the identification of θ (the common post-inspection shift) more than τ (the air-specific differential). θ relies on between-facility variation in inspection timing, which is clearly non-random. τ is identified from the within-cell differential across media: a facility-level confounder must affect air and non-air releases *differentially* to bias τ . The institutional separation of CAA and CWA enforcement makes such medium-specific confounders less plausible, though not impossible.

The stacked design provides additional protection. By comparing each treatment cohort to not-yet-treated facilities and including cohort-specific fixed effects, the stacked estimator constructs cleaner control groups. The convergence of $\hat{\tau}_{\text{TWFE}} = -0.0716$ and $\hat{\tau}_{\text{stacked}} = -0.0671$ is reassuring.

7.2 Pre-Trends

The pre-trend Wald test—joint significance of event-study coefficients at $t = -5$ through $t = -2$ —yields $p = 0.314$ (Table 8). This provides no evidence against the parallel-trends assumption for the air-vs-non-air differential. Visual inspection of Figure 2 is consistent: pre-period coefficients are close to zero without a monotonic pattern that would mimic the post-treatment effect.

7.3 Randomization Inference

I conduct a permutation exercise, randomly reassigning facility-level treatment timing 100 times to construct a reference distribution of the medium-specific estimates under random treatment assignment. The RI p -values are 0.650 for air and 0.690 for non-air (Table 8). This permutation fundamentally tests the *between*-facility variation in medium-specific effects, not the *within*-cell variation in τ . Because the paper’s central finding rests on the relative differential τ rather than the levels of individual medium-specific effects, the RI exercise is a conservative test of a different estimand. The stacked event-study design—which constructs clean not-yet-treated comparisons for each cohort—provides stronger protection against the concerns that motivate permutation-based inference.

7.4 Alternative Clustering and Event Windows

Table 8 reports the main estimates under alternative inference procedures. State-level clustering (column 2) produces somewhat smaller standard errors and retains high significance for $\hat{\tau}$. Two-way clustering by facility and year (column 3) yields wider standard errors but does not change the qualitative conclusions: $\hat{\tau} = -0.0716$ (SE = 0.0264). The ± 3 year event window (column 4) attenuates the differential to $\hat{\tau} = -0.0636$ (SE = 0.0146), consistent with the composition shift building over time. Excluding 2020 (column 5) leaves estimates essentially unchanged ($\hat{\tau} = -0.0734$, SE = 0.0170), indicating that COVID-era disruptions do not drive the results.

7.5 Event-Window Sensitivity

The choice of event window involves a tradeoff. A wider window (± 5 years, the baseline) includes more pre- and post-inspection variation, improving precision, but also increases the risk that unrelated events contaminate the treatment effect. A narrower window (± 3 years, column 4 of Table 8) restricts the comparison to a tighter neighborhood around the first inspection, at the cost of fewer observations (329,924 vs. 435,368).

The narrower window attenuates $\hat{\tau}$ from -0.0716 to -0.0636 , a reduction of approximately 11%. This attenuation is consistent with the event-study evidence in Figure 2, which shows the air–non-air gap widening gradually over the post-inspection period rather than appearing as an immediate discrete jump at $t = 0$. If the composition shift accumulates as facilities experience multiple years of overlapping enforcement, a shorter post-period mechanically captures less of the cumulative differential. Conversely, the common post-inspection shift $\hat{\theta}$ moves from $+0.0167$ (baseline) to $+0.0121$ (narrower window), with the narrower estimate losing significance at the 5% level. This is consistent with θ being the less robustly identified parameter, as established by the balance test failure.

The qualitative conclusion is invariant to the window choice. Under both windows, $\hat{\tau}$ is negative and highly significant, and the medium-specific decomposition continues to show flat air releases with a declining water coefficient. The composition illusion is not an artifact of a particular event-window specification. If anything, the attenuation in the narrower window provides additional information about timing: the relative differential builds gradually rather than arriving as an instantaneous shock, which is more consistent with overlapping enforcement programs operating on different schedules than with a single discrete deterrence event.

7.6 Functional Form

The baseline specification uses $\log(Y + 1)$ to handle the many zeros in non-air media (85.2% of water, 95.8% of land, and 91.2% of POTW observations are zero). This transformation is common but not innocuous: [Chen and Roth \(2024\)](#) show that the $\log(Y + 1)$ transformation can produce estimates that are sensitive to the units of measurement and difficult to interpret when zeros are prevalent. [Table 10](#) reports estimates under two alternative functional forms.

The inverse hyperbolic sine (IHS) transformation, which approximates the log for large values but is defined at zero, yields $\hat{\tau} = -0.0698$ (SE = 0.0184) and $\hat{\theta} = +0.0157$ (SE = 0.0072). These are close to the baseline estimates (-0.0716 and $+0.0167$), with the difference falling well within the confidence intervals. The levels specification (dependent variable in pounds) yields $\hat{\tau} = -0.859$ thousand pounds (SE = 0.112) and $\hat{\theta} = +0.188$ thousand pounds (SE = 0.035), both highly significant. These magnitudes are consistent with the sample means ([Table 1](#)): mean air releases are approximately 10,000 pounds, so the relative differential corresponds to roughly 8.6% of the mean—close to the log-point estimate of 7.2%. In physical units, the relative air differential corresponds to roughly 859 fewer pounds of air releases relative to non-air releases per facility-chemical-year, while the non-air composite rises by 188 pounds. As noted in [Section 6](#), this levels estimate still describes the within-cell relative differential, not an absolute air reduction—the medium-specific air coefficient is essentially zero. The stability of the sign, significance, and approximate magnitude of $\hat{\tau}$ across all three functional forms indicates that the composition illusion is not an artifact of how zeros are handled.

7.7 Extensive Margin

The intensive-margin specifications measure changes in the quantity of releases conditional on any reporting. But enforcement could also operate at the extensive margin—causing facilities to open or close release pathways. [Table 11](#) reports linear probability models for the probability of any positive release in land and water, estimated at the facility-chemical-year level (108,515 observations).

The land coefficient is $+0.0030$ (SE = 0.0017), significant at the 10% level, suggesting that some facility-chemical pairs begin reporting positive land releases after a CAA inspection. The magnitude is small: a 0.3 percentage point increase against a base rate that includes 95.8% zeros. The water coefficient is -0.0037 (SE = 0.0031), negative but not statistically significant. Interestingly, the CWA-inspected indicator is significant for water ($+0.0104$, SE = 0.0049, $p < 0.05$) but not for land ($+0.0013$, SE = 0.0032), suggesting that CWA enforcement increases the probability of water reporting—plausibly through enhanced monitoring requirements

rather than increased discharges.

These extensive-margin results reinforce the main finding. The composition illusion operates primarily through heterogeneous intensive-margin movements across media (flat air, declining water, drifting land) rather than through wholesale opening or closing of release pathways. The limited evidence of new land reporting is consistent with some facilities reclassifying or initiating small land-disposal activities after inspection, but the effect is economically small and statistically marginal.

7.8 Staggered DiD Considerations

The TWFE estimator can produce biased estimates under heterogeneous treatment effects with staggered adoption (Callaway and Sant’Anna, 2021; de Chaisemartin and D’Haultfoeuille, 2020; Sun and Abraham, 2021). The stacked design directly addresses this concern. The convergence of the TWFE ($\hat{\tau} = -0.0716$) and stacked ($\hat{\tau} = -0.0671$) estimates indicates that treatment-effect heterogeneity across cohorts is not a first-order concern for the air differential. The attenuation of $\hat{\theta}$ in the stacked design (from +0.0167 to +0.0072, no longer significant) suggests mild heterogeneity in the common post-inspection shift, consistent with θ being less robustly identified.

7.9 Summary of Identification Assessment

The diagnostics present a picture with two distinct components. First, the relative air differential τ is robust across specifications (TWFE, stacked, alternative clustering, shortened windows). It reflects a genuine within-cell composition shift. Second, the medium-specific regressions show that individual air releases are approximately unchanged, and the relative shift reflects heterogeneous movements across media—flat air, declining water, drifting land. The balance test failure and the conservative RI exercise qualify the causal interpretation of θ but are less threatening to τ , which is identified from within-cell variation. The pre-trend Wald test ($p = 0.314$) supports the identifying assumption for the differential.

8. Discussion

What the evidence shows. The evidence is strongest for three claims. First, within facility-chemical cells, air releases fall relative to non-air releases after a CAA inspection: $\hat{\tau} = -0.0716$ in the TWFE specification and $\hat{\tau} = -0.0671$ in the stacked design, both highly significant (Tables 2 and 3). Second, this relative differential does not correspond to an absolute air reduction: the medium-specific air coefficient is +0.0108 (not significant), and

composition outcomes—air share (+0.0008), log total (+0.0048), log air-only (+0.0086)—are all indistinguishable from zero (Tables 4 and 9). Third, the only individually significant medium-specific coefficient is water (-0.0287 , $p = 0.006$), and 839 of 3,504 facilities face simultaneous CWA enforcement (Table 1).

What the evidence does not show. The evidence does not cleanly identify what generates the relative differential. Adding the CWA control does not move $\hat{\tau}$ (Table 2), so we cannot attribute the differential to CWA enforcement through a control-variable strategy. The balance test failure ($F = 26.41$, $p \approx 0$) means that inspection timing is endogenous, qualifying the causal interpretation of θ . The split-sample mechanism test (Table 5) shows a relative differential for both CAA and non-CAA chemicals, which is inconsistent with a pure strategic-substitution story but does not identify the precise channel. The honest summary is that the composition shift is real, its decomposition is suggestive, and the causal attribution to specific enforcement programs remains imprecise.

This is an estimand-interpretation paper, not a large quantitative bias estimate. The contribution is showing that a robust and precisely estimated relative differential does not correspond to any single-medium absolute effect. The stakes are therefore conceptual rather than large in magnitude: the paper identifies a failure of mapping from relative to absolute effects, not a large measured gap in aggregate pollution.

Why this matters for the literature. The enforcement effectiveness literature has established that inspections reduce emissions in the targeted medium (Gray and Shadbegian, 2005; Shimshack and Ward, 2007; Gray and Shimshack, 2011). Many of these studies use within-facility variation and measure the targeted medium only. The present paper shows that relative within-facility multi-medium estimates can be mistaken for medium-specific effects when enforcement programs are correlated. The concern does not require cross-media substitution by firms—it requires only that multiple enforcement programs operate simultaneously on the same facilities, which is the institutional reality.

The composition illusion connects to a broader theme in the enforcement evaluation literature. Shimshack (2014) surveys the economics of environmental monitoring and enforcement, emphasizing that credible identification of enforcement effects requires careful attention to what the comparison group experiences. When the “untreated” comparison within a facility is non-air releases that are themselves affected by CWA enforcement, the relative design does not isolate what policymakers want to know: did the air inspection reduce air pollution? Gray and Shimshack (2011) similarly stress that enforcement evaluations must account for the full regulatory environment facing each facility, not just the program under study. The present paper provides a concrete, empirically grounded example of what can go wrong when

this advice is not followed. The relative differential $\hat{\tau} = -0.0716$ is precisely estimated and robust, yet it does not correspond to a detectable air-specific effect. This is not a hypothetical concern—it is an empirical demonstration in the exact setting where enforcement effectiveness is most commonly studied.

Generalizability. The composition illusion is not specific to air and water enforcement. The mechanism requires two ingredients: (i) correlated enforcement programs that target overlapping sets of regulated entities, and (ii) relative within-unit designs that benchmark the targeted outcome against other outcomes within the same entity. These conditions hold in many regulatory settings. Occupational safety enforcement (OSHA) and workers' compensation regulation overlap on the same establishments; financial regulators (SEC, CFTC, state banking authorities) share jurisdiction over overlapping sets of institutions; and food safety inspections by FDA and USDA target many of the same processing facilities. In each case, an evaluator using relative within-establishment variation to assess one program's effectiveness risks producing estimates that reflect the joint operation of multiple programs. The issue extends beyond enforcement to any evaluation context where multiple interventions operate simultaneously on the same units and outcomes are measured across multiple dimensions within each unit. The remedy proposed here—pairing relative estimates with dimension-specific and aggregate outcomes—applies equally to these settings.

Substitution vs. composition. The prior version of this paper framed the question around cross-media substitution: do firms strategically reroute pollution from inspected to uninspected media? (Sigman, 1996, 2001). The medium-specific decomposition provides limited support for that framing. The extensive-margin analysis (Table 11) shows limited evidence of new land pathways opening (+0.0030, significant at the 10% level), but the intensive-margin air coefficient is essentially zero. What the data show is a relative composition shift that does not map cleanly onto either strategic substitution or medium-specific deterrence.

Policy implications. EPA's National Compliance Monitoring Strategy evaluates CAA and CWA programs separately. If relative multi-medium estimates can be mistaken for medium-specific effects, program evaluations may overstate the effectiveness of single-medium enforcement. Evaluators should report medium-specific and total-release outcomes alongside relative within-cell comparisons, particularly when enforcement programs overlap across media. The TRI provides exactly the data needed for such reporting, but current practice does not exploit it.

Consider the practical consequence. If an evaluator uses a relative within-facility air-vs.-non-air design that yields $\hat{\tau} = -0.0716$ and interprets this as evidence that CAA inspections

reduce air pollution by 7 log points, the evaluator attributes to CAA inspectors an air-specific effect that the medium-specific regressions do not support. This has implications for resource allocation: if policymakers allocate inspection budgets based on relative multi-medium effectiveness measures, they may misallocate resources across programs. The composition illusion is not about fraud or manipulation—it is about the gap between what relative within-facility multi-medium estimates measure and what evaluators want them to mean.

Limitations. Several limitations warrant emphasis. First, RCRA enforcement data are unavailable, limiting the interpretation of land-specific results. If land-disposal enforcement is also correlated with CAA inspections, the land coefficient may be similarly confounded. Second, the log transformation with many zeros may not fully capture extensive-margin responses; Poisson pseudo-maximum-likelihood estimation (Santos Silva and Tenreyro, 2006; Chen and Roth, 2024) is a natural extension (the IHS transformation yields qualitatively similar results: $\hat{\tau} = -0.0698$; Table 10). Third, the TRI covers only facilities above reporting thresholds; smaller facilities may respond differently to enforcement. Fourth, the analysis links enforcement to self-reported releases, which may themselves change in response to inspection (a reporting rather than pollution response). Fifth, the balance test failure means that θ should not be interpreted causally without strong assumptions. Future work should explore chemical-specific toxicity weights: a pound of benzene diverted from air to groundwater has very different health consequences than a pound shifted to a lined landfill (Currie et al., 2020). Risk-weighted composition metrics would transform this measurement finding into a welfare statement.

9. Conclusion

Environmental enforcement evaluation can suffer from a composition illusion. When enforcement programs are correlated and outcomes are measured across multiple media, a relative multi-medium estimate can be mistaken for a medium-specific effect. I document this using EPA enforcement data linked to chemical-level releases across four media for 3,504 manufacturing facilities over 17 TRI years.

Within facility-chemical cells, air releases fall relative to non-air releases after a CAA inspection ($\hat{\tau} = -0.0716$, confirmed at -0.0671 in the stacked design). But medium-specific regressions show that individual air releases are approximately unchanged ($+0.0108$, $p = 0.55$), while water releases decline (-0.0287 , $p = 0.006$), and composition outcomes are all null. The relative differential does not correspond to an absolute air-specific effect.

The composition illusion operates through a simple mechanism: within-cell designs bench-

mark the targeted medium against a non-air composite that may itself be shifting under correlated enforcement. The result is a relative differential that exists in the pooled specification but does not survive medium-specific decomposition. This finding has implications beyond the specific setting. Any enforcement evaluation that uses within-unit relative variation across outcome dimensions—medium, sector, jurisdiction—risks producing relative estimates that are mistaken for dimension-specific effects when enforcement programs are correlated across dimensions.

The remedy is straightforward: evaluators should report medium-specific and total-release outcomes alongside relative within-cell comparisons, particularly when enforcement programs overlap across media. Relative multi-medium contrasts are not enough: evaluators should pair them with medium-specific and aggregate outcomes before interpreting any estimate as targeted-medium deterrence.

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: @ai1scl

First Contributor: <https://github.com/ai1scl>

References

- Callaway, Brantly and Pedro H. C. Sant’Anna**, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2021, *225* (2), 200–230.
- Chay, Kenneth Y. and Michael Greenstone**, “The Impact of Air Pollution on Infant Mortality,” *Quarterly Journal of Economics*, 2003, *118* (3), 1121–1167.
- Chen, Jiafeng and Jonathan Roth**, “Logs with Zeros? Some Problems and Solutions,” *Quarterly Journal of Economics*, 2024, *139* (2), 891–936.
- Currie, Janet, John Voorheis, and Reed Walker**, “What Caused Racial Disparities in Particulate Exposure to Fall?,” *American Economic Review*, 2020, *110* (12), 3821–3849.
- de Chaisemartin, Clément and Xavier D’Haultfœuille**, “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 2020, *110* (9), 2964–2996.
- Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan**, “Truth-telling by Third-party Auditors and the Response of Polluting Firms,” *Quarterly Journal of Economics*, 2013, *128* (4), 1499–1545.
- EPA**, “National Compliance Monitoring Strategy,” Technical Report, U.S. Environmental Protection Agency 2014.
- Foulon, J., P. Lanoie, and B. Laplante**, “Incentives for Pollution Control: Regulation or Information?,” *Journal of Environmental Economics and Management*, 2002, *44*, 169–187.
- Gray, Wayne B. and Jay P. Shimshack**, “Effectiveness of Environmental Enforcement,” *Review of Environmental Economics and Policy*, 2011, *5* (1), 47–66.
- **and Ronald J. Shadbegian**, “The Consequences of Environmental Enforcement Actions,” *Journal of Law and Economics*, 2005, *48* (1), 235–266.
- Greenstone, Michael**, “Does Ambient Air Quality Regulation Affect Heavy Industry?,” *Quarterly Journal of Economics*, 2002, *117* (4), 1209–1257.
- 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.
- Shimshack, Jay P.**, “The Economics of Environmental Monitoring and Enforcement,” *Annual Review of Resource Economics*, 2014, *6*, 339–360.

– **and Michael B. Ward**, “Enforcement and Over-Compliance,” *Journal of Environmental Economics and Management*, 2007, *55* (1), 90–105.

Sigman, Hilary, “Cross-Media Pollution: Responses to Restrictions on Chlorinated Solvent Releases,” *Land Economics*, 1996, *72* (3), 298–312.

– , “The Pace of Progress at Superfund Sites,” *Journal of Law and Economics*, 2001, *44* (1), 315–344.

Silva, J.M.C. Santos and Silvana Tenreiro, “The Log of Gravity,” *Review of Economics and Statistics*, 2006, *88* (4), 641–658.

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

A. Data Appendix

ICIS-Air. The Integrated Compliance Information System for Air (ICIS-Air) records all compliance monitoring activities for Clean Air Act stationary sources. Data were downloaded from EPA’s Enforcement and Compliance History Online (ECHO) bulk download system (https://echo.epa.gov/files/echodownloads/ICIS-AIR_downloads.zip) in March 2026. The file ICIS-AIR_FCES_PCES.csv contains compliance monitoring events. I filter to COMP_MONITOR_TYPE_CODE = “FOO” (FCE On-Site), yielding 635,967 inspections across 279,000 facilities. Each record is linked to a facility via PGM_SYS_ID, which maps to REGISTRY_ID in the ICIS-AIR_FACILITIES.csv file for cross-database linkage.

ICIS-NPDES. The Integrated Compliance Information System for the National Pollutant Discharge Elimination System (ICIS-NPDES) records Clean Water Act compliance monitoring activities. Data were downloaded from EPA ECHO (https://echo.epa.gov/files/echodownloads/npdes_downloads.zip) in March 2026. The dataset contains 709,000 CWA inspections. Facilities are linked to TRI records via FRS REGISTRY_ID. Of TRI facilities in the analysis sample, 839 received at least one CWA inspection during the study period.

RCRAInfo. Resource Conservation and Recovery Act inspection data was sought from EPA’s RCRAInfo database, but the download returned an error page rather than data; RCRA enforcement controls are therefore unavailable.

TRI Basic Plus. The Toxics Release Inventory Basic Plus data files contain the 100 most-used fields from Forms R and A for each reporting year. Annual national files were downloaded from EPA’s Envirofacts system (https://data.epa.gov/efservice/downloads/tri/mv_tri_basic_download/). ICIS-Air and ICIS-NPDES data were downloaded from EPA ECHO in March 2026. TRI data covers reporting years 2005–2022; the most recent TRI release available at time of download covered reporting year 2022. Data were obtained for 17 years (2005–2022, excluding 2012 due to server availability). Release quantities are disaggregated by medium (Sections 5.1–5.5.4 and 6.1 of Form R). I restrict to records reported in pounds and convert missing release quantities to zero.

Linkage. The FRS ID field in TRI corresponds to the REGISTRY_ID field in ICIS-Air and ICIS-NPDES. The matched sample, after applying event-window restrictions (± 5 years around first CAA inspection, requiring at least 2 years before and after), contains 435,420 observations (four media \times facility-chemical-years; 435,368 in regressions after singleton fixed-effect removal) across 3,504 facilities and 323 chemicals.

Variable construction. The dependent variable is $\log(\text{releases} + 1)$ where releases are measured in pounds per facility-chemical-medium-year, winsorized at the 99th percentile within each medium. The four release media are: air (sum of fugitive and stack air emissions, Sections 5.1 and 5.2), water (surface water discharges, Section 5.3), land (sum of landfill, land treatment, surface impoundment, and other on-site disposal, Sections 5.5.1–5.5.4), and POTW (transfers to publicly owned treatment works, Section 6.1). $\text{Post}_{i,t}$ equals one for all years at or after facility i 's first FCE on-site inspection. The CAA chemical indicator is drawn from TRI Column 42 (Clean Air Act chemical flag).

B. Composition Outcomes

Table 9: Composition Outcomes: Air Share, Total Releases, and Air-Only Releases

	(1)	(2)	(3)
	Air Share	Log Total	Log Air
Post	0.0008 (0.0019)	0.0048 (0.0184)	0.0086 (0.0182)
CWA Inspected	0.0009 (0.0029)	0.0102 (0.0206)	0.0137 (0.0198)
Observations	82,845	108,515	108,515
Facility \times Chemical FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Measurement bias	0.0039 log points		

Notes: Column (1): dependent variable is $\text{Air}/(\text{Air} + \text{Water} + \text{Land} + \text{POTW})$, the air share of total releases. Column (2): $\log(\text{total releases} + 1)$. Column (3): $\log(\text{air releases} + 1)$. The measurement bias is the difference between the air-only and total log-point changes: if air falls faster than total, the gap reflects compositional reallocation rather than true abatement. All specifications at the facility \times chemical level with CWA controls. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9 reports regressions with composition-related dependent variables. Column (1) uses the air share of total releases ($\text{Air} / [\text{Air} + \text{Water} + \text{Land} + \text{POTW}]$), column (2) uses log total releases, and column (3) uses log air-only releases. All three post-inspection coefficients are small and statistically insignificant: +0.0008 (SE = 0.0019) for air share, +0.0048 (SE = 0.0184) for log total, and +0.0086 (SE = 0.0182) for log air-only. The measurement gap between the air-only and total log-point changes is 0.0039 log points—small in absolute terms. These null results confirm that air releases do not decline after a CAA inspection and that the aggregate composition of releases barely shifts.

C. Functional Form Robustness

Table 10: Functional Form Robustness

	$\hat{\theta}$ (Post)	$\hat{\tau}$ (Post \times Air)	N
Log(Y+1) [baseline]	0.0167** (0.0065)	-0.0716*** (0.0169)	435,368
IHS(Y)	0.0157** (0.0072)	-0.0698*** (0.0184)	435,368
Levels (1000 lbs)	0.1876*** (0.0351)	-0.8587*** (0.1117)	435,368

Notes: All specifications include facility \times chemical \times medium and year fixed effects, clustered at the facility level. IHS = inverse hyperbolic sine transformation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10 reports the reparameterized estimates under alternative functional forms. The inverse hyperbolic sine (IHS) transformation yields $\hat{\tau} = -0.0698$ (SE = 0.0184), close to the baseline $\log(Y + 1)$ estimate of -0.0716 . The levels specification (dependent variable in thousands of pounds) yields $\hat{\tau} = -0.859$ (SE = 0.112), with a common shift $\hat{\theta} = +0.188$ (SE = 0.035). The qualitative conclusions are unchanged across functional forms.

D. Extensive-Margin Results

Table 11: Extensive Margin: Probability of Any Release by Medium

	(1) Pr(Land > 0)	(2) Pr(Water > 0)
Post	0.0030* (0.0017)	-0.0037 (0.0031)
CWA Inspected	0.0013 (0.0032)	0.0104** (0.0049)
Observations	108,515	108,515
Facility \times Chemical FE	Yes	Yes
Year FE	Yes	Yes
Clustering	Facility	Facility

Notes: Linear probability models for whether a facility-chemical has any positive release in a given medium-year. Unit of analysis is facility \times chemical \times year. A positive coefficient on Post indicates that CAA inspections increase the probability of initiating releases through that medium. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 11 reports linear probability models for the probability of any positive release in land and water. The land coefficient is $+0.0030$ ($SE = 0.0017$), significant at the 10% level—limited evidence that some facilities begin land releases after a CAA inspection. The water coefficient is -0.0037 ($SE = 0.0031$), not significant. These results provide limited evidence of extensive-margin adjustment, consistent with the composition illusion operating primarily through heterogeneous intensive-margin movements across media rather than through the opening of new discharge channels.