Policy Spillovers to Biodiversity: The Impact of Clean Water Act Investments on Bird Diversity and Abundance

August 14, 2026

Research Question

Ecological signals $\overset{?}{\longleftrightarrow}$ federal expenditures to wastewater infrastructure

What are the ecological impacts to biodiversity from federal expenditures to wastewater treatment sanitation plants?

1. {Broadly}: How does effluent cleanup from grant-funded sanitation plant construction under the Clean Water Act (CWA)'s Title II program affect bird populations and community compositions?
2. {Specifically}: Can water-quality-sensitive species **downstream of a treatment plant** be linked to that plant's **grant** - a largely unmeasured **ecological return** of the policy?

Why it matters: the U.S. has spent $\sim$ \$650 Billion on CWA construction grants; Americans spend $\sim$ \$108 Billion/year on birding (Carver, 2024). Omitting that benefit biases the program's measured returns *downward* - the ledger has few measures of its ecosystem services

Contents

Introduction

Background

Data

Framework

Empirical Approach

Where This Project is Heading

Background - Why the Clean Water Act?

The *single largest public works program in the United States: $100/person-year ($650 billion over infrastructure lifetime)* (Keiser and Shapiro, 2019a).

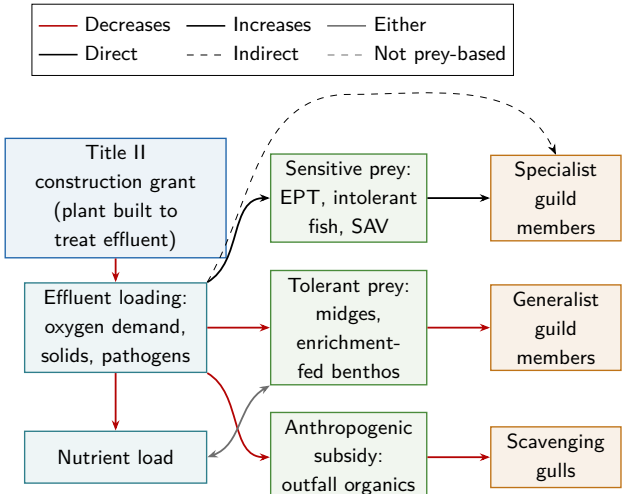


The Cuyahoga fire, 1969

Previous Research	Findings
What are the existing cost-benefits? (Lyon and Farrow, 1995; Keiser et al., 2019)	median benefit-cost ratio **0.37** in ≈20 evaluations (*many undercounts*)
Why the low ratio? (Keiser and Shapiro, 2019b; Shapiro, 2022)	surface water **only** category scoring < 1: *abstract from ecological consequence*
Did the grants clean the water? (Keiser and Shapiro, 2019a)	pollution down 25 miles from source (30 years); capitalized gain $\approx$ 0.25 cost
Did the grants improve health? (Flynn and Marcus, 2025)	birth weight +8 grams; lifts the ratio to $\leq$ 0.45
Did the grants reach the pipes? (Flynn and Smith, 2022; Jerch, 2026)	≈\$0.45 per federal dollar to *new* sewerage capital ; 3x-fees yet bound municipalities grew 13-30% faster
Ecosystem services? (Taylor and Druckenmiller, 2022) *(Section 404 wetlands)*	Flood mitigation: \$1,840 per hectare lost/year in flood-insurance claims (falls below agency estimates that count *wildlife habitat*)
What did the CWA do to the *ecology*?	**This project.** Tools to price nature *input* (Kolstoe and Cameron, 2017)

Background - Birds & Water Quality

- **Why birds?**
  - Sensitive ecosystem-health indicators (Ormerod and Tyler, 1993); 11-23% use inland waters
  - Functional extinctions within niches (Frank and Sudarshan, 2024) → lost ecosystem services: top-down fish regulation, nutrient transport, seed dispersal, bioindication (Green and Elmberg, 2014).
  - *People value them - A LOT* (Carver, 2024).
  - Water quality channels not capitalized, though technology shows up in recreation demand & ecosystem services (Olmstead, 2010).
- **Why water quality?**
  - Heterogeneous contamination + enforcement .
  - Degradation/recovery runs through the food web (Streeter–Phelps oxygen sag → macroinvertebrates → fish/SAV → birds).
  - Direct exposure difficult to scale: no well-defined indices
  - Habitat possibilities: IUCN habitat/threat codings, ECOTOX (not always quality/integrity specific)



Not *currently* modeled: bioaccumulative contaminants; habitat dependence.

Background - Motivating Descriptives

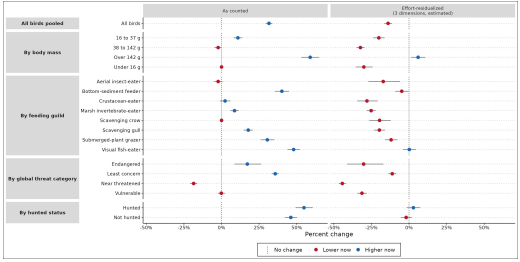


Figure: Long Differences in Species Counts

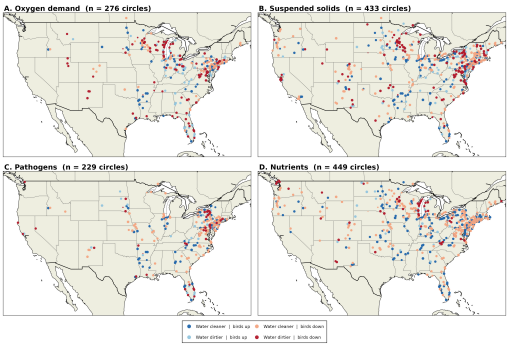
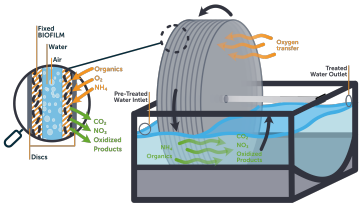


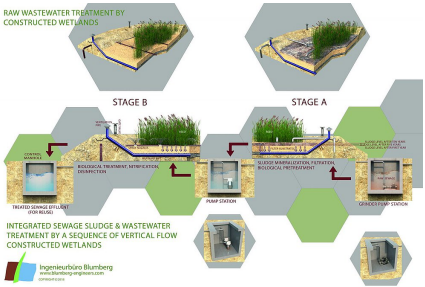
Figure: Bird-Water Concordance in Species Richness

Birds declining is well-backed (Johnston et al., 2025). Suggestions of movement in water quality.

Background - Technologies Funded



Secondary treatment: Rotating biological contactors - conventional projects drew a 75% federal share



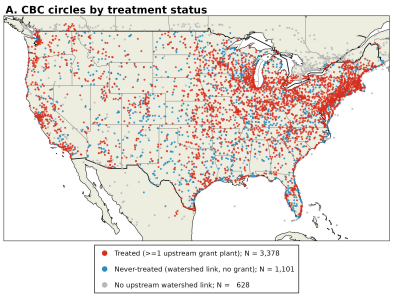
Innovative treatment: Constructed wetlands - an 85% federal share, and one Virginia town saved >65% vs conventional construction (Congressional Budget Office, 1985)

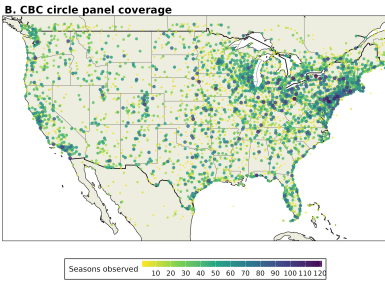
- **Why the CWA?**
  - Costs are VERY well measured - benefits are contested.
  - Title II construction grants, staggered across plants and decades, some forms of treatment more innovative than others.
  - Funding needs remain high. (Copeland, 2016)

A policy whose ranking criteria - pollution severity, population affected, protection of higher-quality waters - didn't explicitly reference wildlife.

Data - Constructed Panel

- **Birds:** Audubon **Christmas Bird Count** (1900-2024): >100k reports, >1k species; 24-km circles, fixed protocol, effort recorded.
- **Treatment:** EPA **GICS** grants × **CWNS** plant characteristics: (commissioning date, design flow, treatment level).
- **Network: NHDPlus** - 25-mile upstream-downstream networks map plants to circles.
- **Controls:** water quality (analytes the grants do not target); ERA5 winter climate; historical housing density (HHUUD10 census tracts).





▸ CBC Protocol

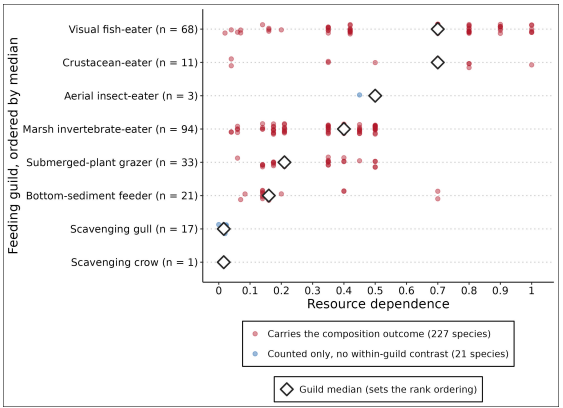
Data - Building up a Measure for Dependence

No single registry assigns birds a water quality sensitivity ranking: *built* from trait databases.

Source	Supplies	Data Limitations
EltonTraits 1.0 (Wilman et al., 2014)	diet proportions + foraging stratum, 9,993 spp.	coded from natural-history *records* in 10% steps (not measured)
AVONET (Tobias et al., 2022)	measured morphology, 11,009 spp. / 90,020 individuals	*ecological* fields derive from the same diet data
De Graaf et al. (De Graaf et al., 1985)	food × substrate × technique, with a **non-breeding** column	web-scraped from PDFs 669 spp., 242 water-linked records
Avibase (Lepage et al., 2014)	1983 names → the eBird spine	splits and low-confidence matches are hand-checked
GloBI (Poelen et al., 2014)	observed feeding interactions	records don't list strength/dependence

Water-linked = forages *on or in* water (EltonTraits foraging strata) **or** wetland-affiliated foraging off the surface

Data - Dependence Measure



Three quantities live in this picture, and they do different jobs.

- ρ_j: **resource dependence** per species: how much of its diet is the affected prey, how narrow that diet is, and whether it needs clear water to catch it.

$$\rho_j = \underbrace{\text{dependence}}_{\substack{\text{diet share on the guild's prey} \\ \text{(EltonTraits; 0-1 in 0.1 steps)}}} \times \underbrace{\text{substitutability}}_{\substack{\text{diet narrowness: 0.4 broad,} \\ \text{0.7 mixed, 1.0 strictly bound}}} \times \underbrace{\text{foraging sensitivity}}_{\substack{\text{capture needs clarity: 0.2 tactile,} \\ \text{0.5 facultative, 1.0 visual-obligate}}}$$

- **The median.** Split to define *specialist* and *generalist* poles.
- $\rho_g = \frac{1}{|g|} \sum_{j \in g} \rho_j$; used *across* guilds. Ranks the guilds against one another as a **moderator**: the axis the response is hypothesized to climb.

Ecological Heterogeneity - one row per species ρ_j.
Median splits a guild *internally*.
ρ_g lets guilds be ordered *against each other*.
Sources: dependence + substitutability = EltonTraits diet data (share; concentration); foraging sensitivity is in *no* database - guild detection mode + hand-coded, cited exceptions.

Data - Prey Base and Hypothesized Direction

Column	How it is set	When the case is close
sensitivity pole	ρ_j, split at the guild median	tie → specialist ($\geq$); trait databases disagree → pre-specified **sensitivity set**
Prey Base		
(what birds feed on)	fixed in advance with the guild (shared prey × foraging arena); in the data: *one EltonTraits diet column per guild*, feeding ρ_j's dependence factor	mechanical from the trait data where sources agree; unclear cases follow a *documented hand pass*
hypothesized sign	inherited from *per-taxon* tolerance grades (Karr fish, Hilsenhoff 0-10 invertebrates), read on the guild's *arena-specific* prey set	channels conflict → **contested by construction** (the scaup); no index grades the prey → **pre-set default**

Outcomes, with **generalists as the within-guild baseline**:

$$s_{igt} = N^{\mathcal{S}}_{igt}/(N^{\mathcal{S}}_{igt} + N^{\mathcal{G}}_{igt}) = \text{share of specialists}, \qquad R_{igt} = N^{\mathcal{S}}_{igt}/N^{\mathcal{G}}_{igt} = s_{igt}/(1 - s_{igt}) = \text{odds of specialists}$$

Ecological Literature: the community-specialization CWM meets bioassessment's sensitive-taxa share; tolerance grades are *per prey taxon*, bridged to the guild by its foraging arena.

Built-in effort control: any common factor - party-hours, habitat area - cancels *inside the cell*: $R = c\,\mu^{\mathcal{S}}/(c\,\mu^{\mathcal{G}}) = \mu^{\mathcal{S}}/\mu^{\mathcal{G}}$.

CBC Effort Parameters

Default: ungraded prey (macrophytes, carrion) gets its sign fixed in advance from the mechanism - solids → light → submerged plants gain (Dennison et al., 1993). The hand-pass procedure, and what each trait database contributes, are spelled out on *how it was fixed*.

Data - Example

Same product, four different deciding factors

Common Loon - *visual piscivore*
dependence 0.7 × substitutability 1.0 × foraging **1.0** (visual-obligate)
$\rho = 0.70$ = guild median → **specialist**
prey: clarity-sensitive fish → sign + *the clarity-obligate*



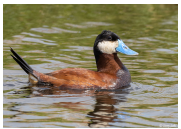
Double-crested Cormorant - *visual piscivore*
dependence **1.0** × substitutability 1.0 × forage **0.2** (close-quarter)
$\rho = 0.20$ → **generalist**
prey: the same fish → same guild sign
eats MORE fish, ranks lower: ***sensitive prey, tolerant bird***



Lesser Scaup - *benthivore*
dependence 0.8 × substitutability **1.0** × forage 0.5 (facultative)
$\rho = 0.40 \gg$ median 0.16 → **specialist**
prey: enrichment-fed benthos → sign **contested**
obligate on a ***tolerant*** *prey base: the sign is a separate axis*

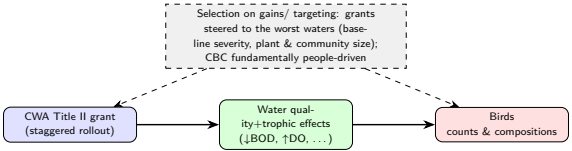


Ruddy Duck - *benthivore*
dependence 0.6 × substitutability **0.7** × forage 0.2 (tactile)
$\rho = 0.084 <$ median 0.16 → **generalist**
prey: the same benthos → same guild sign
broad diet, can switch off the benthos



Why does it matter we know which birds benefited? Why is more of one bird not good enough?
Ecosystem services differ by role, and recreation demand prices specific kinds of birds in the winter and the different amount of birds in the summer (Kolstoe and Cameron, 2017).

Framework - Causal Chain



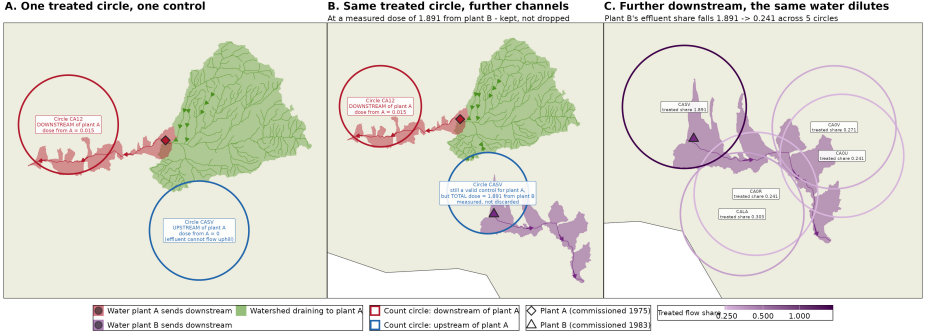
Treatment = staggered commissioning × dilution of the **effluent treated flow share** from plant p to circle i:

$$\text{FS}_{pi} = \frac{\overbrace{\text{plant design flow}}^{\text{chosen: capacity, technology}}}{\underbrace{\text{mean-annual river flow at the outlet}}_{\text{hydrology: how big the river is}}} \qquad D_{it} = \sum_{p} \text{FS}_{pi} \cdot \mathbf{1}[\text{commissioned}_p \leq t]$$

- **Why dilution?** Mass balance: the water-quality change a circle sees is $\Delta c_{\text{circle}} = \text{FS}_{pi} \times \Delta c_{\text{effluent}}$ The Dose
- FS: *the fraction of the circle's water that passed through the plant*; a grant event becomes measured, **continuous** exposure (Callaway et al., 2025; Fan, 2025), comparable across rivers.
- **Denominator = exogenous**: the river's size is set by hydrology.
- **Numerator = chosen (potentially endogenously)**: capacity tracks expected growth, where technology can *be* habitat (a constructed wetland treats effluent **and** is a marsh) - **selection on gains**.
- $\Delta c_{\text{effluent}} \approx$ **common by policy**: the uniform secondary standard (85-95% BOD removal) compresses it; the residual is *observed* (treatment-level bins; CWNS pre/post effluent concentrations) and enters as controls + falsification.

▸ Identification Threats

Framework - Visualization



- **A - gravity picks the control.** Plant A splits its own neighborhood: the circle below it drinks its water (1.5% of the river); the circle above it never can - *effluent cannot flow uphill.* Same valley, same winter, same counters; the river constructs the control (Fan, 2025).
- **B - overlaps are measured.** That same upstream control is itself dosed by plant B - at 1.891, *more treated water than natural flow.* **Treatment is continuous**: a circle's dose is the *sum* over its upstream plants. Treatment is a number, not a label.
- **C - distance is dilution.** Follow plant B's water down: 1.891 → 0.241 across five circles, because the river keeps gathering clean water.
- **Overlap in space, rarely in time.** Only the two oldest CALA (running 1938-2023) and CASV (1957-2024) genuinely coexist *(watershed-clustered errors; redistribution tested by spatial spillover gradient).*

The Dose | Dose QC

Empirical Approach - The Specification

A continuous, effluent-treated dose share for circle i, guild g, at CBC season t:

$$\underbrace{\Delta^{\text{dir}} Y_{igt}}_{\substack{\text{down-minus-up}\\ \text{in outcomes}}} = \underbrace{\alpha_i + \lambda_{gt}}_{\substack{\text{circle,}\\ \text{guild-season FE}}} + \sum_v \Big(\underbrace{\beta_v}_{\substack{\text{common}\\ \text{curve}}} + \underbrace{\gamma_v \, \rho_g}_{\substack{\text{dependence}\\ \text{tilt}}} + \underbrace{\xi_v' \, \mathbf{X}_i}_{\substack{\text{controls,}\\ \text{pre-treatment}}} \Big) \underbrace{b_v(D_i)}_{\substack{\text{dose}\\ \text{basis}}} \underbrace{\text{Post}_{pt}}_{\substack{\text{post-grant}}} + \underbrace{u_{igt}}_{\substack{\text{error}}} \qquad (1)$$

where,

- $\Delta^{\text{dir}} Y_{igt}$ = the downstream Y_{igt} outcome - the upstream Y_{igt} outcome across the *same* plant-CBC season.
  Anything common to the river reach (regional trends, weather, basin-wide shocks) subtracts out *before* estimation
- $v = 1, \ldots, V$ indexes the pieces of the **sieve**: the "dose basis" $b_v(D_i)$ is a set of simple functions of the dose (spline segments)
  i.e., the data choose the *shape* of the dose response rather than a presumed straight line
- β_v = the number multiplying piece v, shared by *every* guild: $\sum_v \beta_v \, b_v(d)$ traces the **common dose-response curve**
- $\gamma_v \, \rho_g$ **tilts** that common curve up the dependence ladder (ρ_g = how far guild g depends on the resource the policy moves): higher-ρ guilds respond more steeply iff $\gamma_v > 0$
- $\mathbf{X}_i$ = *pre-treatment* development, cropland, non-target water quality, CBC effort, winter severity

ρ_g **multiplies the dose basis and nothing else**: moves the *slope*

Empirical Approach - Interpretation and Identification

differentiate Eq. 1 in d : $\mathrm{ACRT}^{\mathrm{dir}}(d \mid d, \rho_g, \mathbf{X}) = \sum_v \left(\beta_v + \gamma_v \rho_g + \xi_v' \mathbf{X}_i\right) b_v'(d)$ i.e. the **average causal response** (Callaway et al., 2025)

(2)

differentiate Eq. 2 in ρ : $\partial_\rho \mathrm{ACRT}^{\mathrm{dir}} = \sum_v \gamma_v \, b_v'(d)$ i.e. that response's gradient in dependence (3)

- **Interpretation.** Controls × the Dose | Identification Threats | Scaling
  - $\mathrm{ACRT}^{\mathrm{dir}}(d \mid d, \rho_g, \mathbf{X})$ - **if one more dollar of upstream treatment raised the treated share of this circle's water, how much further would its community tilt toward the specialists?**
  - $\partial_\rho \mathrm{ACRT}^{\mathrm{dir}}$ - **does that recovery land where the ecology hypothesizes?**
- **Identification.**
  - **PT *in the dose*:** trends of the differenced outcome $\Delta^{\mathrm{dir}} Y$ common across *dose* groups identify the level curve; the slope $\mathrm{ACRT}^{\mathrm{dir}}$ needs *strong* PT - the same counterfactual trajectory at *every* dose (no selection on gains)
  - What Δ^{dir} buys: the *differencing* - both marginal parallel-trends conditions cancel identically; only the cohort-by-direction trend interaction survives as a requirement
  - **What could break it:** the dose *numerator* is chosen (capacity tracks growth; a technology can be habitat) - met with pre-treatment $\mathbf{X}_i \times$dose controls and the technology-margin falsification. The *denominator* is hydrology
  - The pre-trend test and the effort × dose diagnostic police the pole-specific residual

Immediate Steps After Model Estimates and Pipeline

- **Targeting of grants** - generalized random forests (Athey et al., 2019): where do investments matter most?
- **Birds don't stay still** - Checking for SUTVA violations (Butts, 2026)
- **From measurement to monetization** - using existing tools for recreation demand (Kolstoe and Cameron, 2017; Melstrom and Reeling, 2026; Wardle, 2025)

Thank You!

Contact: ap795@cornell.edu

Appendix - Functional Form

One community, two margins. Within guild g: $N^{\mathcal{S}}$ specialists, $N^{\mathcal{G}}$ generalists, dominance odds $R = N^{\mathcal{S}}/N^{\mathcal{G}}$; both arms Poisson pseudo-maximum-likelihood (PPML): **Each arm answers one question, and both model the *expected count* as rate $\times$ exposure:**

$$\textbf{how many birds?} \quad \mathbb{E}\left[N_{igt} \mid \cdot\right] = \underbrace{E_{it}}_{\substack{\text{exposure:}\\ \text{party-hours}}} \cdot \underbrace{e^{\eta_{igt}}}_{\substack{\text{rate: birds}\\ \text{per hour}}} \qquad \Rightarrow \textbf{ count arm}$$

$$\textbf{which birds?} \quad \mathbb{E}\left[N^{\mathcal{S}}_{igt} \mid \cdot\right] = \underbrace{N^{\mathcal{G}}_{igt}}_{\substack{\text{exposure:}\\ \text{generalists}}} \cdot \underbrace{e^{\eta_{igt}}}_{\substack{\text{rate: specialists}\\ \text{per generalist}}} \qquad \Rightarrow \textbf{ composition arm}$$

Specialists per generalist is exactly the odds $R = N^{\mathcal{S}}/N^{\mathcal{G}}$ - a dose term in η moves the odds proportionally. (Eq. 1's right-hand side); the exposure's coefficient is pinned at one (GLM: an *offset*).

- **Why a count model for an odds outcome**: the odds are never computed as data - zero-specialist cells stay ordinary zeros - and only the rate equation must be right on average. The Algebra
- **Neither arm takes a log** - $\log(1+y)$ is unit-dependent (Chen and Roth, 2024); PPML keeps the zeros and the proportional reading (Santos Silva and Tenreyro, 2006) - and **Hill diversity travels beside them** as the partition-free check. PPML Hill Numbers The Retired Relative Index

Appendix - The Retired Relative Index

The earlier design measured abundance/richness as a *relative index* = a year × geography fixed-effect coefficient (after Liang et al., 2020).

- **Degenerates on sparse cells:** a one-checklist circle-year is absorbed by its own FE - the index carries no information.
- **Quietly saps power:** a near-uninformative outcome rejects *no* pre-trend; the reassuring $p = 1.00$ is what low power looks like.
- **The fix is a closed form:** Hill numbers and one-stage PPML are functionals of the *raw counts* - nothing to degenerate.

Ill-posed extraction degeneracy ⇒ well-posed measurement problem with known corrections.

For the record: the retired index gave resident-abundance ATT $\approx +0.107$ SD ($z = 2.10$, pre-trend $p = 1.00$) - on a near-uninformative outcome, so $p = 1.00$ is low power as much as parallel trends.

back to Functional Form

Appendix - Hill Numbers I: The Rarity Dial

Notation, for one count cell: p_i = the share of the cell's birds belonging to species i (so $\sum_i p_i = 1$); S = the number of species present. One family of diversity measures (Hill, 1973; Jost, 2006), ${}^qD = \left(\sum_i p_i^q\right)^{1/(1-q)}$, all in units of the **effective number of species**; the dial q sets how much rare species count:

- $q = 0$: every species counts equally $\rightarrow$ richness S - most sample-dependent; true value not recoverable.
- $q = 1$: species weighted by abundance $\rightarrow$ **Hill-Shannon**, the **primary** member - asymptotically near-unbiased.
- $q = 2$: dominant species drive it $\rightarrow$ Hill-Simpson $1/\sum_i p_i^2$ - robustness.

Scale-invariant (a function of the shares p_i, not the counts) - silent about *quantity*, so it travels beside the arms as the partition-free check.

The $q = 1$ member travels under **two names** - Shannon entropy and the effective number of species. Part II defines both and shows they are the same number.

continue: Entropy vs Effective Number ◂ back to Functional Form

Appendix - Hill Numbers II: Entropy vs the Effective Number of Species

Part I's $q = 1$ member, unpacked - two names, one number. With p_i as before (species i's share of the cell's birds), Shannon entropy and Hill-Shannon diversity are exact transforms of each other:

$$\underbrace{H = -\sum_i p_i \ln p_i}_{\textbf{entropy} \text{ - average surprise, in nats}} \qquad \underbrace{{}^1D = \exp(H)}_{\textbf{effective number of species}} \qquad \Longrightarrow \qquad \underbrace{\ln {}^1D = H}_{\text{the identity: } \textit{exact, cell by cell}}$$

- **What H measures**: the average *surprise* of the next bird - how unpredictable the community is. Units: nats. Hard to read directly.
- **What ${}^1D$ measures**: *the number of equally common species that would produce the same entropy* - "effective species." Two species at 50/50: $H = \ln 2 = 0.693$, ${}^1D = 2$. Two species at 90/10: $H = 0.325$, ${}^1D = 1.38$ - the community *behaves like* about 1.4 equally common species.
- **Why report ${}^1D$**: it is linear in intuition - twice the diversity means twice the effective species, not twice the nats.
- **Why the identity matters for identification**: because $\ln {}^1D = H$, the apparent 2×2 menu of outcome (H or ${}^1D$) × trend assumption (levels or logs) has exactly **two** live cells - counterfactual effective species moving by a common *number*, or by a common *proportion*. (${}^1D$, logs) *is* (H, levels); (H, logs) is struck because $H = 0$ whenever one species holds the whole cell.

◂ back to Hill Numbers ◂ back to Functional Form

Appendix - The Measure's Lineage and Resolution

Lineage - each source's exact contribution. *s is* a community-weighted mean (CWM): the community trait mean of Garnier et al. (2004), trait restricted to specialist membership $\{0, 1\}$. The direct avian precedent is Julliard et al. (2006)'s community specialization index - an abundance-weighted mean of species-level specialization. The sign discipline is bioassessment's - grade taxa *ex ante* by tolerance, score the community by its sensitive share: Karr (1981)'s fish index of biotic integrity, Hilsenhoff (1987)'s invertebrate index (itself an abundance-weighted mean of tolerance grades: the *same functional*, tolerance as the trait), Davies and Jackson (2006)'s sensitive-vs-tolerant tiers. **New here**: the same functional *within* a resource guild, with the constructed ρ_j as the grade.

Resolution - the indices are per prey taxon; the arena is the bridge.

- The EltonTraits diet column is per *bird* and coarse: *how much* of bird j's diet is the prey class - all dependence needs (tolerance never enters dependence).
- The guild's foraging *arena* fixes *which taxa*: benthivore = sediment benthos; marsh invertivore = EPT; visual piscivore = water-column fish.
- The indices grade *those taxa* (Hilsenhoff: 0-10 per invertebrate family; Karr: fish); the guild's sign is the grade of its arena-specific set.
- **Proof in the lookup:** four guilds share the invertebrate diet column, four sign treatments - marsh invertivore + (EPT, 0-3), benthivore *contested* (7-10), crustacivore *contested* (mid), aerial insectivore *placebo* (8-10). Grades straddle → contested; no index → pre-set default.

◂ back to Prey Base

Appendix - How the Guild Table Was Fixed

Every discretionary choice was made in advance, in a written order - nothing is fitted to bird outcomes.

1. **Guilds and prey bases come first, from ecology.** A guild = one shared prey resource × one foraging arena (Root, 1967) - set before any estimation, and it pins the prey base: in the data, one EltonTraits diet column per guild.
2. Frozen Lookup (diet share and breadth read off EltonTraits; detection mode arrives with the guild). Where the databases lump - EltonTraits' invertebrate column mixes aquatic with terrestrial prey - species are flagged for a **documented hand pass**: consult the species' published natural-history account, record the citation in the frozen lookup. *State of play*: **108** flagged, **73** with completed reviews.
3. **Species the literature itself cannot settle** go to a pre-specified sensitivity list; estimates run with and without them.
4. Indices Behind the Signs: from the prey's published tolerance grades where an index exists; where none does (plants, carrion), the default is fixed in advance from the mechanism, and labeled as such.

What each database contributes: **EltonTraits** - the quantitative inputs (diet share, breadth); **AVONET** - measured morphology (mass, the foraging-reach input) + a cross-check (its ecology fields share Elton's diet lineage); **De Graaf** - an independent published guild scheme whose *disagreement* with González-Salazar defines the sensitivity list; **Avibase** - 1983 names → modern taxonomy; **GloBI** - observed feeding records, corroboration only. **Guardrails**: pole and sign are *derived*, not hand-set; the hand-coded factor carries a collapse-to-one robustness twin; every call keeps its citation.

◂ back to Prey Base

Appendix - The Frozen Lookup, Excerpted

Eleven of 248 rows:

Species	Guild	Dependence	Substitution	Foraging	ρ	Median Guild ρ	Pole
Common Loon	visual piscivore	0.7	1.0	1.0	0.70	0.70	specialist
Great Blue Heron	visual piscivore	0.5	0.7	1.0	0.35	0.70	generalist
Double-crested Cormorant	visual piscivore	1.0	1.0	0.2	0.20	0.70	generalist
Lesser Scaup	benthivore	0.8	1.0	0.5	0.40	0.16	specialist
Ruddy Duck	benthivore	0.6	0.7	0.2	0.084	0.16	generalist
Virginia Rail	marsh invertivore	0.2	0.4	0.5	0.04	0.40	generalist
Gadwall	submerged-plant grazer	0.7	1.0	0.5	0.35	0.21	specialist
Canvasback	submerged-plant grazer	0.6	0.7	0.5	0.21	0.21	specialist
American Coot	submerged-plant grazer	0.4	0.7	0.5	0.14	0.21	generalist
Tree Swallow	aerial insectivore	0.9	1.0	0.5	0.45	-	-
Ring-billed Gull	scavenging gull	0.2	0.4	0.2	0.016	-	-

Aerial insectivores and gulls carry no pole - their guilds serve the count arm only (placebo and negative-sign roles). The full 248-row table ships as a manuscript appendix table, emitted by the pipeline.

back to How It Was Fixed

Appendix - The Tolerance Indices Behind the Signs

The indices are published, field-calibrated grades.

Hilsenhoff assigns each invertebrate family a 0-10 tolerance value (0 = intolerant of organic enrichment, 10 = thrives in it) (Hilsenhoff, 1987);

Karr's index of biotic integrity grades fish (Karr, 1981); AMBI is the estuarine benthic analogue.

The rule. Look up the guild's prey set in those grades:

sensitive prey → +, tolerant prey → contested, no grade → pre-set default.

Guild	Prey base	Grade source	Tolerance	Sign
visual piscivore	clarity/oxygen-sensitive fish	Karr (intolerant fish)	sensitive	+
marsh invertivore	emergent-wetland invertebrates; estuarine benthos	Hilsenhoff 0-3 (EPT); AMBI (intertidal)	sensitive	+
benthivore	enrichment-fed benthos	Hilsenhoff 7-10	tolerant	contested
crustacivore	amphipods / crustaceans	Hilsenhoff mid-range	intermediate	contested
aerial insectivore	tolerant emergent midges	Hilsenhoff 8-10	tolerant	placebo
scavenging gull	subsidy + freshwater fish	Karr / Hilsenhoff	mixed	−
scavenging corvid	urban carrion subsidy	no aquatic grade exists	tolerant	placebo
submerged-plant grazer	submerged macrophytes	no grade exists; light channel (Dennison et al., 1993)	sensitive	pre-set default +

back to How It Was Fixed

Appendix - The Flow-Share Dose D_{it}

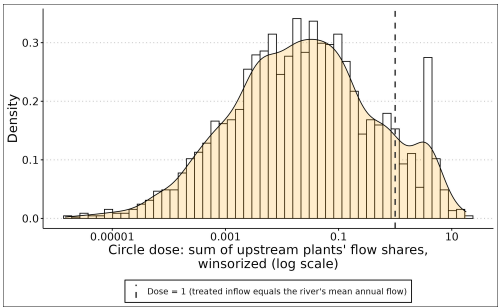
Per plant-circle pair:

$$\mathrm{FS}_{pi} = \frac{\text{plant design flow (MGD)}}{\text{mean-annual flow at the outlet (QA\_MA)}} \text{ i.e.,}$$

the treated-effluent fraction of the water where it enters the CBC circle.

- Circle-year dose
  $D_{it} = \sum_p \mathrm{FS}_{pi} \cdot \mathbf{1}[\text{commissioned}_p \leq t]$
  (raw + winsorized; binary $\mathbf{1}[D > 0]$ for the anchor).
- QA_MA = NHDPlus's mean-annual river flow, modeled by **EROM** (the Enhanced Runoff Method).
- **Quality control (QC)** - what the heavy tail is, and what screens it:

Dose QC



Distribution of the circle-year dose D_{it} (25-mile band).

back to Visualization

Appendix - Why PPML for Abundance

The open frontier: Wooldridge (2023) validates saturated PPML under an index parallel trends; the bridge to the Callaway-Sant'Anna clean-comparison structure for *staggered, continuous-dose* counts is not yet built.

- Counts carry many zeros; $\log(1+y)$ targets a unit-dependent estimand (Chen and Roth, 2024) and drops the extensive margin.
- One-stage **Poisson**: $\mathbb{E}[N_{it} \mid \cdot] = \exp(\theta D_{it} + \alpha_i + \delta_t + \cdots)$ - θ reads as a **proportional** effect, scale-free.
- Robust to count heteroskedasticity; keeps every zero.

back to Functional Form

Appendix - The Offset Model, Worked (I): Counts to Odds

Step 0 - the model. Poisson pseudo-ML on the *specialist count*, with the log generalist count entered at a coefficient *fixed at one* (the offset):

$$\mathbb{E}\left[N^{\mathcal{S}} \mid N^{\mathcal{G}}, X\right] = \exp\Big(\underbrace{\ln N^{\mathcal{G}}}_{\text{offset}} + \eta(X)\Big) = N^{\mathcal{G}} \cdot e^{\eta(X)}$$

Step 1 - from counts to odds

$$\frac{\mathbb{E}\left[N^{\mathcal{S}} \mid N^{\mathcal{G}}, X\right]}{N^{\mathcal{G}}} \overset{(a)}{=} \mathbb{E}\left[\frac{N^{\mathcal{S}}}{N^{\mathcal{G}}} \,\middle|\, N^{\mathcal{G}}, X\right] \overset{(b)}{=} \mathbb{E}[R \mid \cdot] \overset{(c)}{=} \frac{N^{\mathcal{G}}\, e^{\eta(X)}}{N^{\mathcal{G}}} = e^{\eta(X)} \quad \Longrightarrow \quad \eta(X) = \ln \mathbb{E}[R \mid \cdot]$$

(a) **The conditioning rule** ("taking out what is known"): $N^{\mathcal{G}}$ is one of the variables conditioned *on*, i.e., a fixed, known number in the exptection - and a constant passes through an expectation in either direction, $c\,\mathbb{E}[Z \mid \cdot] = \mathbb{E}[c\,Z \mid \cdot]$, here with $c = 1/N^{\mathcal{G}}$.

(b) **Definition**: $R \equiv N^{\mathcal{S}}/N^{\mathcal{G}}$.

(c) **Substitute the Step-0 model** for $\mathbb{E}[N^{\mathcal{S}} \mid \cdot]$ and cancel $N^{\mathcal{G}}$.

So the fitted index η *is* the log of the expected odds.

Part II ◂ back to Functional Form

Appendix - The Offset Model, Worked (II): Dose Effects

Step 2 - differentiate in the dose. Part I ended at $\eta(X) = \ln \mathbb{E}[R \mid \cdot]$. For a dose term θD inside η:

$$\frac{\partial \ln \mathbb{E}[R \mid \cdot]}{\partial D} = \theta \quad \Longrightarrow \quad \text{one dose unit multiplies the } \textit{expected odds} \text{ by } e^{\theta}$$

a dose coefficient reads as a *proportional* shift in the odds: $e^{\theta} = 1.05$ means five percent more specialists per generalist, at any baseline.

Step 3 - what the estimator actually solves. PPML picks η to zero the sample moment

$$\sum_i \left(N_i^S - N_i^G \, e^{\eta_i} \right) x_i = 0$$

- consistent whenever the Step-0 *mean* is right - no Poisson distribution is assumed (quasi-likelihood);
- a cell with zero specialists enters as $N_i^S = 0$, an ordinary value - nothing is dropped, no ratio is ever formed.

◂ back to Part I | What Cancels in the Odds | ◂ back to Functional Form

Appendix - What Cancels in the Odds, and What Does Not

Why effort divides out. Let a factor c scale both poles alike - survey effort, a habitat multiplier, baseline richness - so $N^S = c\,\mu_S$ and $N^G = c\,\mu_G$:

$$R = \frac{c\,\mu_S}{c\,\mu_G} = \frac{\mu_S}{\mu_G} \qquad \text{for } \textit{any } c > 0, \text{ in } \textit{every} \text{ cell, at } \textit{every} \text{ date.}$$

The cancellation is *contemporaneous* - inside one cell, before any differencing - so even a **time-varying** common factor never reaches the difference (a DiD alone removes only a time-invariant c).
The count arm, a level, carries c and must condition on effort.

Where it stops. If the factor is *pole-specific* - a cryptic specialist detected differently from a conspicuous generalist, $c_S \neq c_G$ - it does not cancel:

$$R = \frac{c_S\,\mu_S}{c_G\,\mu_G} \quad \Longrightarrow \quad \ln R = \ln(\mu_S/\mu_G) + \ln(c_S/c_G).$$

Case	What handles it
Common, $c_S = c_G$	cancels contemporaneously, before any differencing
Differential, time-invariant	a constant shift in $\ln R$; absorbed by the circle FE
Differential, trending with dose	**survives everything** - the one genuine residual

The pre-trend test and the effort $\times$ dose interaction are aimed at that third line.

◂ back to The Algebra ◂ back to Functional Form

Appendix - What Δ^{dir} Buys

Take one design cell - grant cohort G (the plant's grant year; ∞ = never) $\times$ direction q (1 = downstream, 0 = upstream) - and write how its outcome would drift with *no grant*. Any such drift has only four possible pieces:

$$\delta_{G,q} = \underbrace{\mu}_{\substack{\text{drift everyone}\\ \text{shares (national)}}} + \underbrace{\alpha_G}_{\substack{\text{drift of this}\\ \text{cohort's basins}}} + \underbrace{\beta_q}_{\substack{\text{down-vs-up drift,}\\ \text{true on any river}}} + \underbrace{\gamma_{G,q}}_{\substack{\text{drift of this}\\ \text{cell alone}}}$$

Two differences, in the pipeline's own order - each deletes one piece.

1. **Down minus up at the same plant** - the pre-model $\Delta^{\text{dir}} Y_{igt}$ - *the left-hand side of Eq. 1.* Both circles share the nation's drift and their basin's drift, so:

$$\delta_{G,1} - \delta_{G,0} = (\cancel{\mu} + \cancel{\alpha_G} + \beta_1 + \gamma_{G,1}) - (\cancel{\mu} + \cancel{\alpha_G} + \beta_0 + \gamma_{G,0}) = (\beta_1 - \beta_0) + (\gamma_{G,1} - \gamma_{G,0})$$

2. **Grant cohort minus never-grant** - the DiD the regression then runs on $\Delta^{\text{dir}} Y$. Both cohorts share the on-any-river down-vs-up drift, so:

$$\left[\cancel{(\beta_1 - \beta_0)} + \gamma_{G,1} - \gamma_{G,0}\right] - \left[\cancel{(\beta_1 - \beta_0)} + \gamma_{\infty,1} - \gamma_{\infty,0}\right] = (\gamma_{G,1} - \gamma_{G,0}) - (\gamma_{\infty,1} - \gamma_{\infty,0})$$

What survives is the whole bias. The design is clean unless a drift is *directional and grant-specific at once* - neither marginal parallel-trends condition is ever used. Steps 1-2 are Eq. 1's own two moves: its left-hand side is built pre-model; its Post $\times$ dose terms run the cohort contrast. The guild differential differences *again*, across guilds, deleting anything guild-common (Caron, 2025).

◂ back to the Model

Appendix - Scaling/Sensitivity to Functional Form

The share s, the odds R, and the log-odds are monotone transforms of one another and carry *identical* information.

What differs is the identifying assumption. Trends parallel in the odds are *not* trends parallel in the log-odds, and neither implies the other (Roth and Sant'Anna, 2023). Choosing the scale therefore *is* choosing the parallel-trends assumption (settled by a data checks).

Two halves to the check.

- **Definedness** - the share is defined on roughly **70%** of guild-circle-seasons, the odds on **62%**, the realised log-odds on **48%**.
- **A differential pre-trend estimated on the odds scale**, which discriminates between the branches rather than presupposing one.

All constructed scales are reported; definedness currently favors s, the differential pre-trends favor the log-odds.

Closed route: adding a constant inside a logarithm to admit the zeros makes the estimate depend on the outcome's units (Chen and Roth, 2024).

back to the Model

Appendix - Controls x the Dose

The rule. A control with no dose argument has zero dose-derivative: $\partial_d(\zeta X) = 0$. It shifts the *level*'s variance and reads a differential, but it *cannot touch the slope*. Interacted, it reaches both functionals:

$$\mathrm{ACRT}^{\mathrm{dir}}(d \mid d, \rho_g, \mathbf{X}) = \sum_v \left(\beta_v + \gamma_v \rho_g + \xi_v' \mathbf{X}\right) b_v'(d).$$

So the ACRT - the headline object - is defended *only* by $\mathbf{X}\times$ dose terms.

Must be pre-treatment: the realized post-grant value is the mediator - conditioning on it blocks the very effect being estimated (bad control).

The Controls ◂ back to the Model

Appendix - The Controls

How each planned control family lands on the composition arm, under the rule of the slide before:

Type of Control	Source	How it lands on the composition arm
Survey effort	CBC effort fields (party-hours, counters, miles)	a common detection multiplier - **cancels** in the ratio. Load-bearing on the count arm (the offset); here a named *diagnostic*, entered as effort $\times$ dose
Baseline nutrients	WQP, pre-grant, pooled to the HUC8 watershed	the common part cancels; what remains is a nutrient-driven *differential trend* - the residual the pre-trend test polices
Development / habitat	HHUUD10 historical housing units (census tracts, 1940-2019)	a common suitability multiplier cancels; a static snapshot is absorbed by the circle FE *and* differenced out of Δ^{dir}; only a gradient sorting the poles *and* trending with dose survives
Winter severity	ERA5 reanalysis	exogenous to grant timing; interacted - the response should be strongest where food is otherwise scarce, and conditioning on food directly is barred (the grant is what improves it)

This is the direct answer to "doesn't the ratio already account for these?" - largely yes, and provably so, down to the one residual that trends with the dose.

◂ back to Controls x the Dose ◂ back to the Model

Appendix - Identification Threats

Threat	How It's Addressed
Severity targeting: priority lists steered grants to the worst waters	Δ^{dir} - severity is reach-common, both circles sit on the same river; baseline severity $\times$ dose; the dose: **77-98%** of dose variance is hydrological flow-share magnitude vs 2-23% plant count, with only 13-20% commissioning timing
Chosen numerator: design capacity tracks community growth	*measured*: 66% of dose variation is river size; plant size is nearly orthogonal to river size (cor $\approx$ 0.10); 90% of variation is across plants. Growth controls $\times$ dose; a plant-by-year FE spec is kept as a *check* on the subset where within-plant dose actually varies
$\Delta c_{\text{effluent}}$ **not common**: upgrades differ in what they remove	its extent is *estimated*, not assumed: treatment-level bins + CWNS pre/post effluent levels enter as controls; **capacity-only class doubles as a growth placebo**: capacity that moves no water quality should move no birds
Technology IS habitat: a constructed wetland or lagoon treats effluent *and* is a marsh	the **technology-margin falsification**: *pre-program* bird trends (long differences over the full CBC span, effort-residualized) must not differ by the technology class a grant *later* funds - nor by grant receipt at all (in flight)
The counting eye moves: CBC is people-powered	effort placebo: effort itself shows no jump at commissioning ($p \geq 0.4$); the pole-specific residual is policed by effort $\times$ dose

Water quality is a mediator: baseline values only, and only analytes the grants do not move (nitrate, lead, sediment); BOD, DO, TSS, fecal are the treatment's own channels. ERA5 winter severity is exogenous to grant timing.

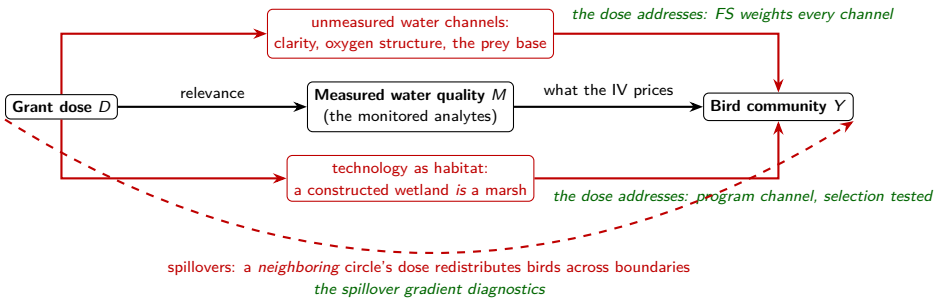
Why Not an IV? | The Decompositions | Water Placebo | Landbird Placebo | Effort Placebo | Extensive Margin

◂ back to Framework

Appendix - Why Not an IV? The Assumptions

The tempting two-stage design - grants *instrument* water quality; water quality prices birds - needs two assumptions:

- **Relevance** - grants must move the *measured* analytes M: $\mathrm{cov}(D, M) \neq 0$ given controls. Testable, true in direction (the mediator event studies) - but monitoring is sparse
- **Exclusion** - grants may touch birds *only* through M: every other path from dose to birds must be exactly *zero*. Untestable - and false on this project's own record:



"Isn't habitat your problem too?" ◂ back to Identification Threats

Appendix - Why Not an IV? The Habitat Objection

"A constructed wetland is a marsh - isn't that a problem for the dose design too?"

- **For the IV it is bias.** Exclusion demands the path be exactly *zero*; a real habitat channel makes the two-stage estimate wrong
- **For the dose it is estimand.** The reduced form prices the *program*, and a marsh the grant built is a real channel of the program.
  Threatens identification through *selection*: ponds and lagoons want cheap, flat, high-water-table land - also wetland-adjacent, bird-rich land - so technology choice can mark slow-moving site traits that predict bird trajectories on their own
- **Selection is testable:**
  - the **technology-margin falsification**: *pre-program* bird trends (long differences over the full count record, effort-residualized) must not differ by the technology class a grant *later* funds - selection on slow-moving site traits (in flight)
  - **extent-of-change controls**: treatment-level bins + plant-survey pre/post effluent levels absorb what the upgrade did to the water
  - the **capacity-only class** doubles as a growth placebo: capacity that moves no water quality should move no birds
- **If the falsification fails**, the pond class gets separate handling: dropped as robustness; habitat provision reported as its own labeled arm

The falsification needs the technology classes *measured*.

The Technology Margin in the Data | ◂ back to the DAG | ◂ Identification Threats

Appendix - Why Not an IV? Different Question

- **Two different questions.** The two-stage design prices the bird response *per unit of measured analyte*, averaged where the first stage is strongest (where monitoring caught the movement) - a water-quality *elasticity*. This project prices the *program*: every channel a grant dollar operates, at the incidence the river assigns. The reduced form in the dose *is* the policy parameter
- **The weights are the substance.** The dose weights each upstream plant by *hydrology*: its share of the receiving water, FS = plant design flow ÷ river flow, and mass balance carries $\Delta c_{\text{circle}} = \mathrm{FS} \times \Delta c_{\text{effluent}}$ - dilution *is* the weight, summed over all upstream plants, *whatever the channel*. 2SLS swaps that for a *statistical* weight - strongest where measured analytes moved most, *zero* on unmeasured channels - and boundary-crossing redistribution contaminates the exclusion instead of being estimated on its own
- **A valid application** (Jerch, 2026): IV targets *who is treated* - mandate exposure via downstream populations and state pollution-rights history - outcomes reduced-form. This design does the 'same' with within-river differences, decompositions, and technology falsifications

Kept from the IV idea: the 'would be first stage' *shown* - treated channels move, the exogenous class does not - as evidence of mechanism, not as identification.

The Decompositions | Water First Stage | ◂ back to Identification Threats

Appendix - The Technology Margin in the Data

Two margins of the technology record are already tabulated; the habitat-relevant one is owed:

A. What kind of project the grant funded

construction grants, $N = 20{,}067$

Project type	N	Share
unlabelled (mostly interceptors, pump stations)	9,191	45.8%
modification / upgrade	4,901	24.4%
new construction	2,902	14.5%
capacity + treatment-level increase	2,037	10.2%
capacity only (the growth placebo)	570	2.8%
treatment level only	464	2.3%

two singleton stray codes omitted

B. How far the plant treats (1978 survey)

$N = 30{,}670$ plants; the earliest wave coding this

Treatment level	N	Share
none	107	0.3%
primary	1,311	4.3%
partial secondary	2,976	9.7%
secondary	6,615	21.6%
advanced secondary	2,189	7.1%
advanced (tertiary)	706	2.3%
not reported	16,766	54.7%
of which: already secondary+, never construction-granted	7,077	23.1%

[Placeholder - table in production.] C. The unit process itself - the habitat-relevant margin. The plant survey records each facility's full treatment train ($\approx$ 9 unit processes per facility), including the pond and wetland family (stabilization pond, aerated lagoon, total containment pond, aquaculture / wetlands / marsh systems) and land-treatment processes (overland flow, infiltration). The composition table for grant-funded plants is queued with the technology-margin falsification build and will replace this box.

◂ back to the Habitat Objection ◂ Identification Threats

Appendix - The Two Dose-Variance Decompositions

Two questions about where the dose's variation comes from, answered separately:

A. Did grant decisions create the dose differences?

Split one: place vs timing	
differences across places (which plants sit upstream, how big their shares)	80-87%
differences from *when* plants switched on	13-20%
Split two: holding the number of upstream plants fixed	
the size of the shares - how much each river dilutes	**77-98%**
the number of upstream plants	2-23%
same split, extreme doses capped	77-88%

No - the dose varies mainly because *rivers* differ, not because grants were aimed or timed. This rebuts targeting; it cannot rule out big rivers also being different bird habitat (a separate check across channel sizes covers that).

B. How much of the dose is the part somebody *chose*?

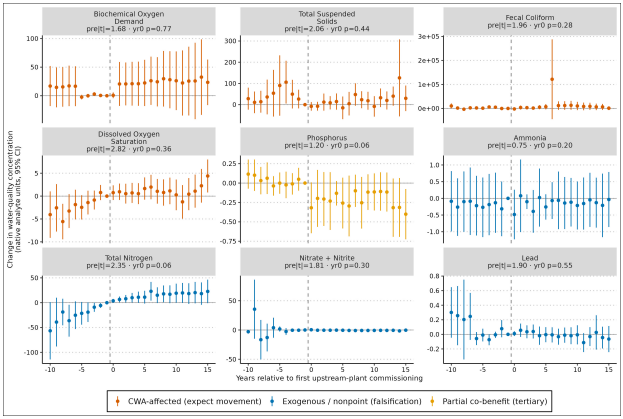
The dose is plant size ÷ river size; split its spread:	
from **plant size** - the chosen part	44.3%
from **river size** - nature's part	**66.1%**
overlap between the two (rows sum to 100)	-10.4%
correlation of plant size with river size	**0.10**
spread *within* one plant's own circles	9.9%

The chosen part is the minority, and plant size is nearly unrelated to river size. Panel A cannot see this (its "share size" bundles plant and river together) - which is why both panels are shown; the within/between split is computed below.

How these were computed (17,909 plant-circle shares, 6,299 plants, 25-mile band): split each log dose share's total spread by group - the standard between/within variance identity. **Panel B's 9.9%**: grouping *by plant*, the spread among one plant's own circles averages 0.183 of 1.842 total = 9.9% within, 90.1% between. Plant size is constant within a plant, so the within piece can only be river differences - and it equals the within-plant spread of river flow *exactly* (0.1830 = 0.1830). **Panel A**: the same identity on the circle-year dose - grouped by place vs commissioning year, then holding the upstream plant count fixed.

◂ back to Identification Threats

Appendix - Did the Grants Move the Water?



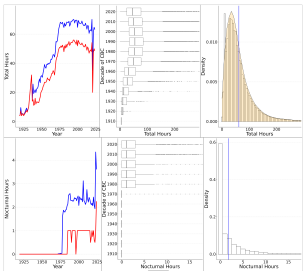
Water-quality analytes at circle outlets, in years relative to first upstream-plant commissioning.

- **Why this check exists:** the design assumes the grant moved the water.
- The **treatment's own channels** move in the treated direction after commissioning - clearest for dissolved-oxygen saturation (up) and phosphorus (down).
- The **exogenous class** the design uses as controls and falsification targets - nitrate, lead - **stays flat**: the analytes the grants do not touch, untouched.
- **Benchmark** (Keiser and Shapiro, 2019a): same direction as their grant first stage (a 50-million-reading, 240,000-site archive); our WQP pull is far sparser - hence the width. Read as **direction + falsification, not magnitudes**.
- **Phosphorus is in-between**: incidental under secondary treatment, but a tertiary-funded subset (Great Lakes / Chesapeake, post-1981) was *targeted* - a secondary outcome, tested separately there.

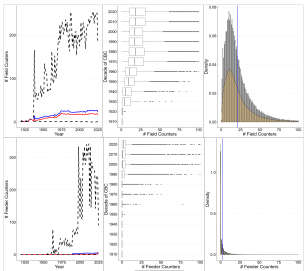
back to Identification Threats

Appendix - CBC Protocol and Effort Variation

Protocol: 24-km circles, non-overlapping; Dec 14 - Jan 5; teams count every bird seen *within* the circle; the organizer logs effort. **Presence-only + variable effort** $\Rightarrow$ counts are a lower bound, and effort (party-hours, counters, miles) must be controlled, not divided out.



Party-hours per count.



Field counters per count.

back to Data

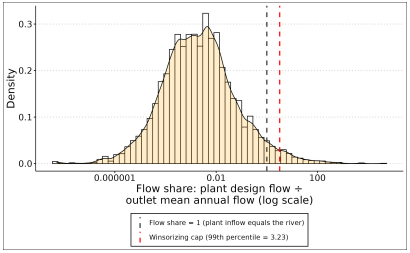
Appendix - Dose QC

Quality control (QC) issues:

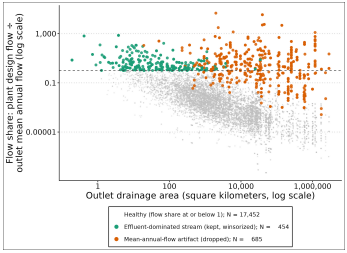
i) coastal snap errors, a plant matched to the wrong flowline (fixed upstream);

ii) **EROM artifacts**, where the flow model hands a braided or divergent channel an implausibly tiny mean-annual flow (screened by a specific-discharge floor).

Specific discharge = the outlet's mean-annual flow ÷ its upstream drainage area (flow per km^2 drained ≈ runoff depth). A reach draining thousands of km^2 cannot plausibly carry a trickle: artifacts sit near 10^{-5} mgd/km^2, normal reaches near 10^{-2} - the **floor** on this quantity is what screens them.



Per-pair flow share FS$_{pi}$ (25-mile band).



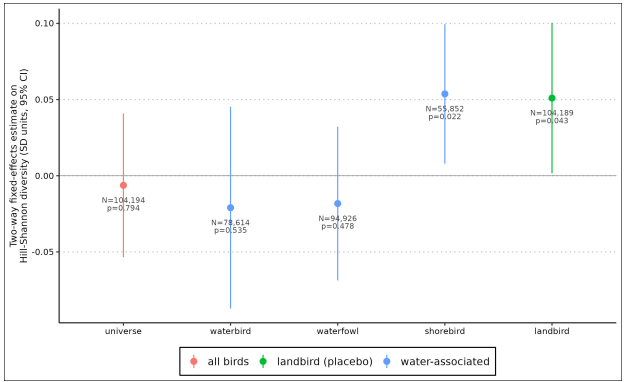
Outlet specific discharge discriminates the two tail populations (effluent-dominated vs. QA_MA artifacts).

Only ≈4% of FS-present pairs exceed 1 after QC, so the treated-discharge-equivalent reading holds; the specific-discharge floor removes the EROM artifacts.

back to The Dose

Appendix - Landbird Placebo

Why this check exists: if the story is water-borne, birds that never touch the waterway should not respond - so run the *naive* design (grant vs. no grant, no direction, no dose): does it pass its own placebo?

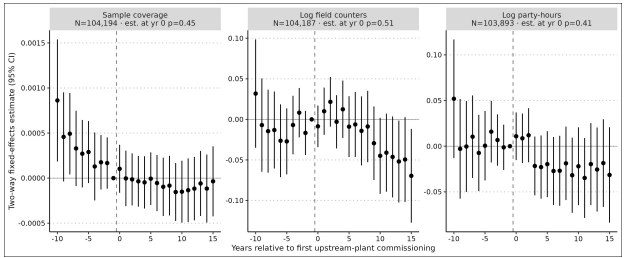


- Water-dependent **shorebirds** respond ($\approx +0.05$ SD, $p = 0.02$) - consistent with the discharge mechanism...
- ...but the terrestrial **landbirds move by the same amount** ($\approx +0.05$ SD, $p = 0.04$). Landbirds do not drink effluent - so the naive comparison picks up whatever *arrived with* the grants (growth, development, regional trends), not the water.

Binary $\mathbf{1}[D > 0]$ effect on Hill-Shannon diversity, guild by guild (two-way FE).
Read this slide as the design's motivation: the naive pass fails its own placebo; the hydrological third difference + dose-interacted controls exist to pass exactly this test. **Grouping caveat**: ran on the coarse pre-guild groups, *before* the mechanism-guild build; the guild-grain re-run rides with the placebo suite.

◂ back to Identification Threats

Appendix - Observation-Effort Placebo

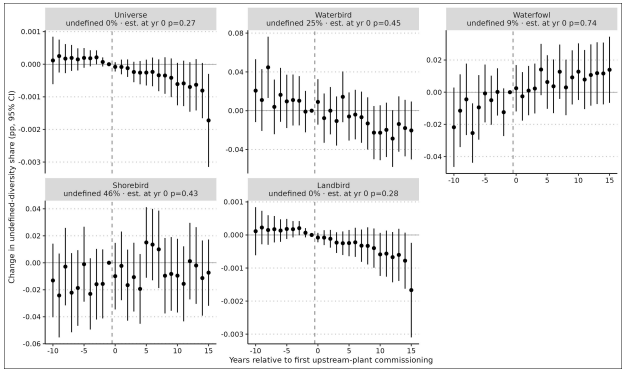


Event-study run on observation *effort* (sample coverage, field-counter counts, log party-hours).

- **Why this check exists:** CBC data are made by people - if grant towns also got more (or keener) birders, a "bird response" could be a *birder* response. So run the event study on **effort itself**, as if effort were the outcome.
- Nothing jumps at commissioning ($p \geq 0.4$ in every panel): **the counting eye did not change when the treatment arrived**.
- **What it does not cover**: detection shifting for *one pole only* (a cryptic specialist vs. a conspicuous generalist). Residual is policed by the effort $\times$ dose diagnostic.

back to Identification Threats

Appendix - Extensive-Margin Separation



Event-study on the share of circle-years with *undefined* diversity (the presence/absence margin), by guild.

- **Why this check exists:** the composition arm compares poles *within* a guild, so it only uses cells where the guild is present. If treatment moved presence itself, conditioning on it would select the sample *on the treatment*.
- So the event study runs on the share of *undefined* cells - the door into the sample. **Flat** through event time (no jump at zero, $p \geq 0.27$).
- **What flat buys**: presence/absence does not move with the grant, so the within-guild comparison is estimated on a selection-free sample - presence margin itself belongs to the *count* arm.

back to Identification Threats

References I

Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. *The Annals of Statistics 47*(2).

Butts, K. (2026, February). Difference-in-differences with spatial spillovers. Working paper, University of Arkansas. Revision of arXiv:2105.03737; no DOI.

Callaway, B., A. Goodman-Bacon, and P. H. C. Sant'Anna (2025). Difference-in-differences with a continuous treatment. NBER Working Paper 32117, National Bureau of Economic Research. arXiv:2107.02637, v8 (December 31, 2025).

Caron, L. (2025). Triple difference designs with heterogeneous treatment effects. arXiv preprint arXiv:2502.19620v2, 3 June 2025.

Carver, E. (2024). Birding in the United States: A demographic and economic analysis. Addendum to the 2022 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation Report 2022-4, U.S. Fish and Wildlife Service, Washington, D.C. Division of Policy, Economics, Risk Management, and Analytics.

Chen, J. and J. Roth (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics 139*(2), 891–938.

Congressional Budget Office (1985, June). Efficient investments in wastewater treatment plants. CBO Study 85-CBO-016, Congressional Budget Office, Washington, D.C. Also available via HathiTrust: https://catalog.hathitrust.org/Record/001818130 and Google Play Books: https://play.google.com/store/books/details?id=nDeALSwnxaEC.

Copeland, C. (2016). Clean water act: a summary of the law. Congressional research service, Library of Congress Washington, DC.

Davies, S. P. and S. K. Jackson (2006). The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications 16*(4), 1251–1266.

De Graaf, R. M., N. G. Tilghman, and S. H. Anderson (1985). Foraging guilds of north american birds. *Environmental Management 9*(6), 493–536.

Dennison, W. C., R. J. Orth, K. A. Moore, J. C. Stevenson, V. Carter, S. Kollar, P. W. Bergstrom, and R. A. Batiuk (1993). Assessing water quality with submersed aquatic vegetation. *BioScience 43*(2), 86–94.

Fan, C. (2025). Dam Thy Neighbor: International relations and the externalities of impounding rivers. Job Market Paper, University of Chicago Harris School of Public Policy.

Flynn, P. and M. Marcus (2025). A watershed moment: The Clean Water Act and birth weight. *Journal of Human Resources 60*(6), 1915–1936.

References II

Flynn, P. and T. Smith (2022). Rivers, lakes and revenue streams: The heterogeneous effects of Clean Water Act grants on local spending. *Journal of Public Economics 212*, 104711.

Frank, E. and A. Sudarshan (2024). The social costs of keystone species collapse: Evidence from the decline of vultures in India. *American Economic Review 114*(10), 3007–3040.

Garnier, E., J. Cortez, G. Billès, M.-L. Navas, C. Roumet, M. Debussche, G. Laurent, A. Blanchard, D. Aubry, A. Bellmann, C. Neill, and J.-P. Toussaint (2004). Plant functional markers capture ecosystem properties during secondary succession. *Ecology 85*(9), 2630–2637.

Green, A. J. and J. Elmberg (2014). Ecosystem services provided by waterbirds. *Biological Reviews 89*(1), 105–122.

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology 54*(2), 427–432.

Hilsenhoff, W. L. (1987). An improved biotic index of organic stream pollution. *The Great Lakes Entomologist 20*(1).

Jerch, R. L. (2026). Municipal governments under the Clean Water Act. *Journal of the Association of Environmental and Resource Economists 13*(3), 791–830.

Johnston, A., A. D. Rodewald, M. Strimas-Mackey, T. Auer, W. M. Hochachka, A. N. Stillman, C. L. Davis, V. Ruiz-Gutierrez, A. M. Dokter, E. T. Miller, O. Robinson, S. Ligocki, L. Oldham Jaromczyk, C. Crowley, C. L. Wood, and D. Fink (2025). North American bird declines are greatest where species are most abundant. *Science 388*(6746), 532–537.

Jost, L. (2006). Entropy and diversity. *Oikos 113*(2), 363–375.

Julliard, R., J. Clavel, V. Devictor, F. Jiguet, and D. Couvet (2006). Spatial segregation of specialists and generalists in bird communities. *Ecology Letters 9*(11), 1237–1244.

Karr, J. R. (1981). Assessment of biotic integrity using fish communities. *Fisheries 6*(6), 21–27.

Keiser, D. A., C. L. Kling, and J. S. Shapiro (2019). The low but uncertain measured benefits of US water quality policy. *Proceedings of the National Academy of Sciences 116*(12), 5262–5269.

Keiser, D. A. and J. S. Shapiro (2019a). Consequences of the Clean Water Act and the demand for water quality. *The Quarterly Journal of Economics 134*(1), 349–396.

References III

Keiser, D. A. and J. S. Shapiro (2019b). US water pollution regulation over the past half century: Burning waters to crystal springs? *Journal of Economic Perspectives 33*(4), 51–75.

Kolstoe, S. and T. A. Cameron (2017). The non-market value of birding sites and the marginal value of additional species: Biodiversity in a random utility model of site choice by eBird members. *Ecological Economics 137*, 1–12.

Lepage, D., G. Vaidya, and R. Guralnick (2014). Avibase – a database system for managing and organizing taxonomic concepts. *ZooKeys 420*, 117–135.

Liang, Y., I. Rudik, E. Y. Zou, A. Johnston, A. D. Rodewald, and C. L. Kling (2020). Conservation cobenefits from air pollution regulation: Evidence from birds. *Proceedings of the National Academy of Sciences 117*(49), 30900–30906.

Lyon, R. M. and S. Farrow (1995). An economic analysis of clean water act issues. *Water Resources Research 31*(1), 213–223.

Melstrom, R. T. and C. Reeling (2026). Modeling recreation and tourism demand using crowdsourced data: An application to visitation statistics from the ebird project. *Tourism Management 112*, 105276.

Olmstead, S. M. (2010). The economics of water quality. *Review of Environmental Economics and Policy 4*(1), 44–62.

Ormerod, S. and S. Tyler (1993). Birds as indicators of changes in water quality. In *Birds as monitors of environmental change*, pp. 179–216. London: Chapman & Hall.

Poelen, J. H., J. D. Simons, and C. J. Mungall (2014). Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics 24*, 148–159.

Root, R. B. (1967). The niche exploitation pattern of the blue-gray gnatcatcher. *Ecological Monographs 37*(4), 317–350.

Roth, J. and P. H. C. Sant'Anna (2023). When is parallel trends sensitive to functional form? *Econometrica 91*(2), 737–747.

Santos Silva, J. M. C. and S. Tenreyro (2006). The log of gravity. *Review of Economics and Statistics 88*(4), 641–658.

Shapiro, J. S. (2022). Pollution trends and US environmental policy: Lessons from the past half century. *Review of Environmental Economics and Policy 16*(1), 42–61.

Taylor, C. A. and H. Druckenmiller (2022). Wetlands, flooding, and the Clean Water Act. *American Economic Review 112*(4), 1334–1363.

References IV

Tobias, J. A., C. Sheard, A. L. Pigot, A. J. M. Devenish, et al. (2022). AVONET: morphological, ecological and geographical data for all birds. *Ecology Letters 25*(3), 581–597.

Wardle, A. R. (2025). The recreational value of rare species: Causal evidence from the cassia crossbill. *Journal of the Association of Environmental and Resource Economists 12*(5), 1133–1164.

Wilman, H., J. Belmaker, J. Simpson, C. de la Rosa, M. M. Rivadeneira, and W. Jetz (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology 95*(7), 2027.

Wooldridge, J. M. (2023). Simple approaches to nonlinear difference-in-differences with panel data. *Econometrics Journal 26*(3), C31–C66.