How Effective Is PSPS at Preventing Wildfires?

Public-data estimates and the case for utility-held records — walking from naive comparisons to quasi-experiments, and why none of it is causal yet

Igor Geyn · June 2026

Full analysis: igorgeyn.com/blog

The Benefits of PSPS Are Simulated, Not Measured

PG&E's headline figure for 2024 — "up to 76,314 acres avoided," 10,841 buildings, 10,868 residents — comes from 24-hour fire simulations at damage locations. The counterfactual is generated inside a model.

76,314
acres "avoided" in 2024 — simulated

100%
effectiveness assumed by the best published EPSS study

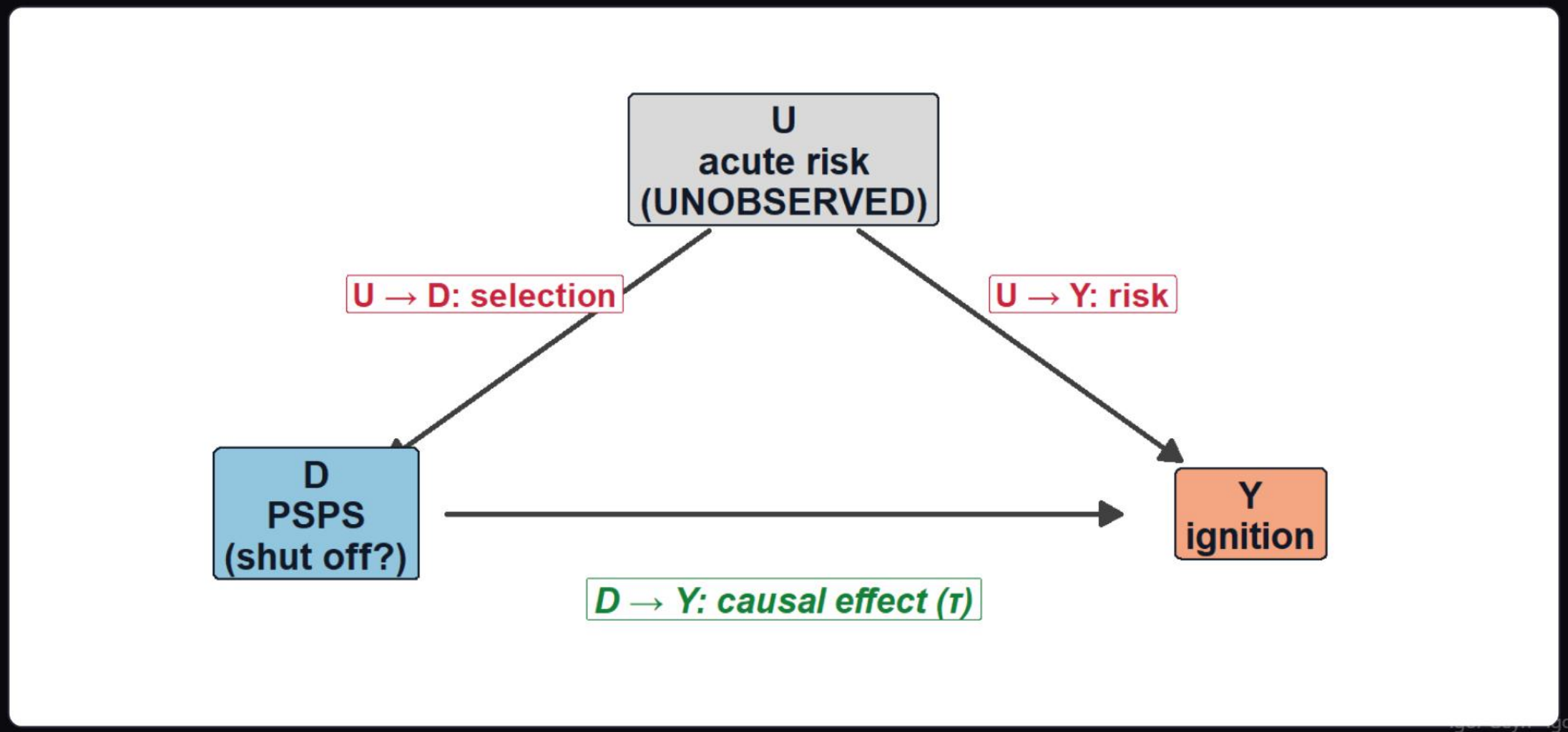
0
published checks against observed ignition outcomes

Why this matters: These numbers feed the cost-benefit ratios behind rate cases and wildfire-mitigation plans. Nobody appears to have estimated PSPS's effect against fires that actually happened. This analysis does, with public data.

PG&E 2024 PSPS Post-Season Report §IV.1; Warner, Callaway & Fowlie (2025), Nature Climate Change.

The Trap: Shut-Off Days Show 15× More Ignitions

Across ~12.2 million PG&E circuit-days, shut-off days ignite at 15.1× the rate of ordinary days (SCE: 24.5×). Read causally, shutoffs *cause* fires. They don't — utilities cut power exactly where and when risk peaks.

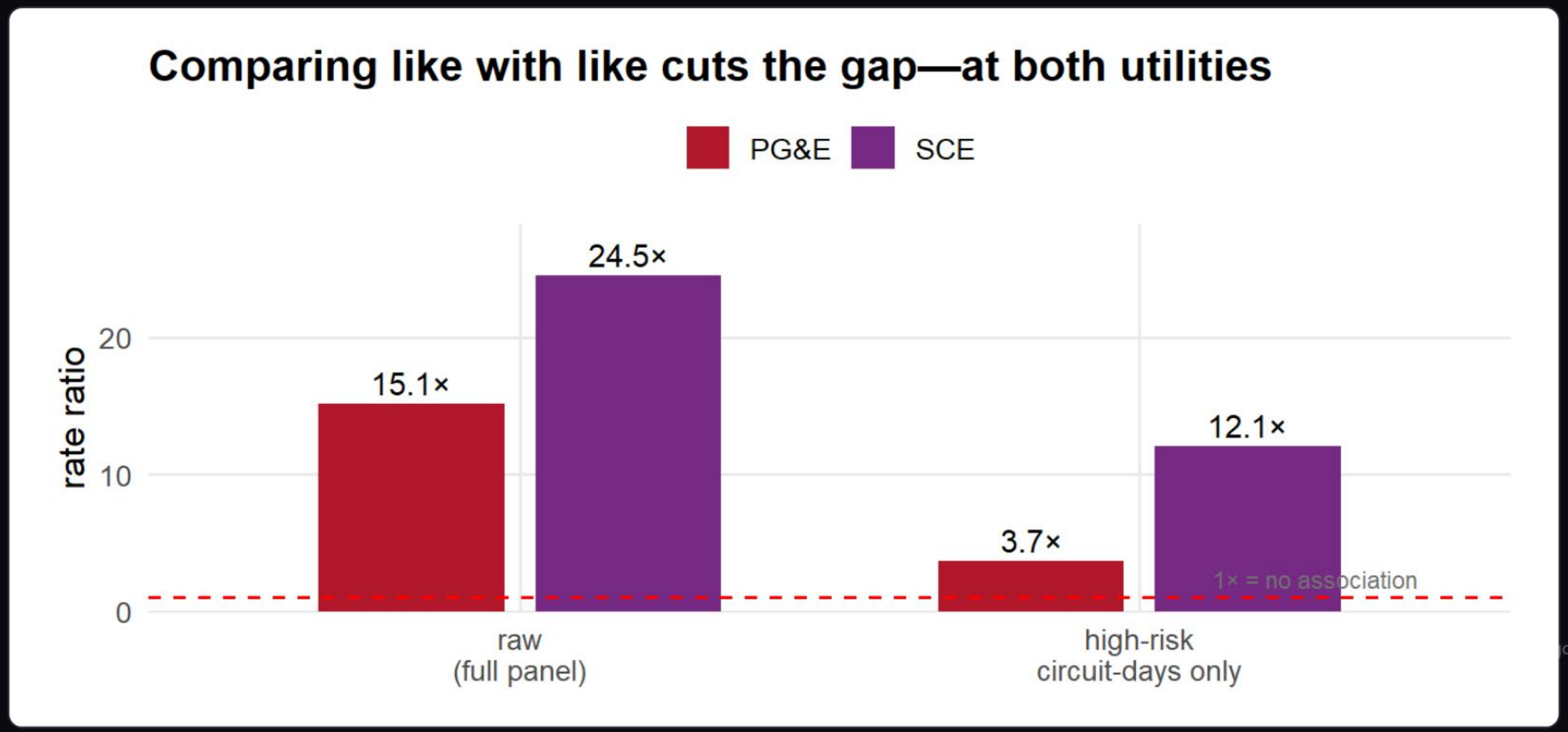


Why this matters: The back-door path D ← U → Y — the operator reacting to high winds, dry fuel, low humidity — is what every subsequent step tries to block. The raw comparison measures the risk, not the policy.

`PG&E panel: RR 15.1× [10.9, 21.0], RD +0.0048. SCE companion panel: RR 24.5× [7.9, 76.1].`

Step 1: Compare Only Where Shutoffs Were on the Table

Restrict to high fire-threat circuits on days a PSPS event was underway somewhere. Three-quarters of PG&E's raw gap evaporates (15.1× → 3.7×); half of SCE's (24.5× → 12.1×).



Why this matters: Much of the alarming raw gap was never about PSPS — just about which circuit-days get shut off. But the ratio sticks at 3.7×, not 1×: a binary risk tier is a coarse proxy for the operator's real signal.

`Within high-risk stratum: PG&E RD +0.0040 (from +0.0048); SCE RD +0.0016 (from +0.0017).`

Steps 2–3: Same Circuit, Same Storm — the Gap Survives

Compare each circuit only to itself (fixed effects): 2.8× [1.8, 4.3]. Then match each shut-off day to look-alike days in the same storm on wind, humidity, fuel, terrain: the gap persists at +0.0036 ignitions per circuit-day.

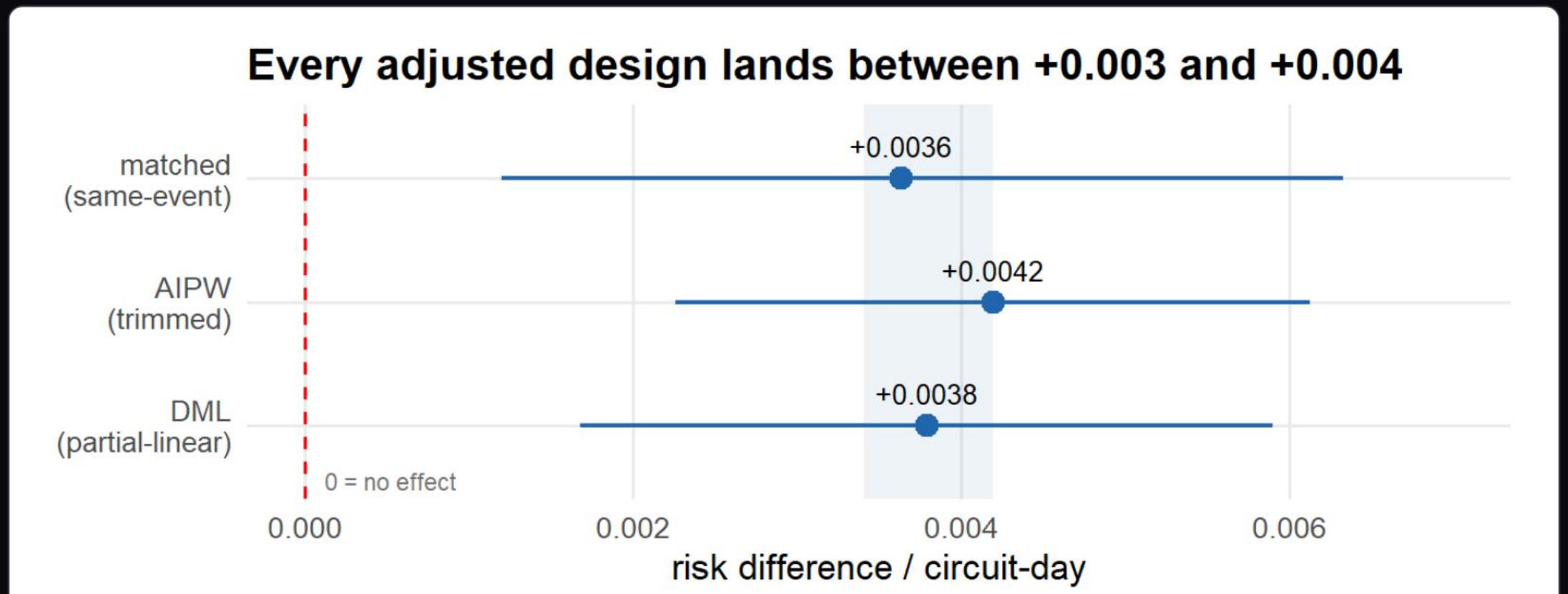
Design	Rate ratio	Risk difference
Raw difference in means	15.1×	+0.0048
High-risk circuit-days only	3.7×	+0.0040
Circuit fixed effects (within-circuit)	2.8×	+0.0034
Matched within storms (ATT)	4.4× [1.4, 24.1]	+0.0036

Why this matters: Erasing every permanent difference between places barely moves the estimate — the inflated gap is about which *days* get shut off, not which places. And matching can only balance what got written down.

PG&E panel. Matched design: same-event matching + Poisson model absorbing circuit and event.

Step 4: Flexible ML Lands in the Same Band

A doubly-robust AIPW estimator — lasso propensity and outcome models, cross-fit by circuit — removes as much observed-confounding bias as the measured covariates allow. It lands where simple matching did.



Why this matters: When richer adjustment stops moving the estimate, the remaining bias is not in the variables you can see. Every adjusted design lands between +0.003 and +0.004 — always positive, never near zero.


```
Double robustness is robustness to functional form, not to unobservedness. Overlap is
poor: ~half of control days have almost no chance of being shut off.
```

The Bounds Convict the Benchmark

Instead of removing the bias, bound it. Assumption-free Manski bounds run [-0.994, +0.006]. Add one mild assumption — a de-energized line cannot start a fire — and the harmful edge pins at zero.

[-0.994, +0.006]

assumption-free Manski band

≤ 0

effect range under monotone treatment response

+0.004

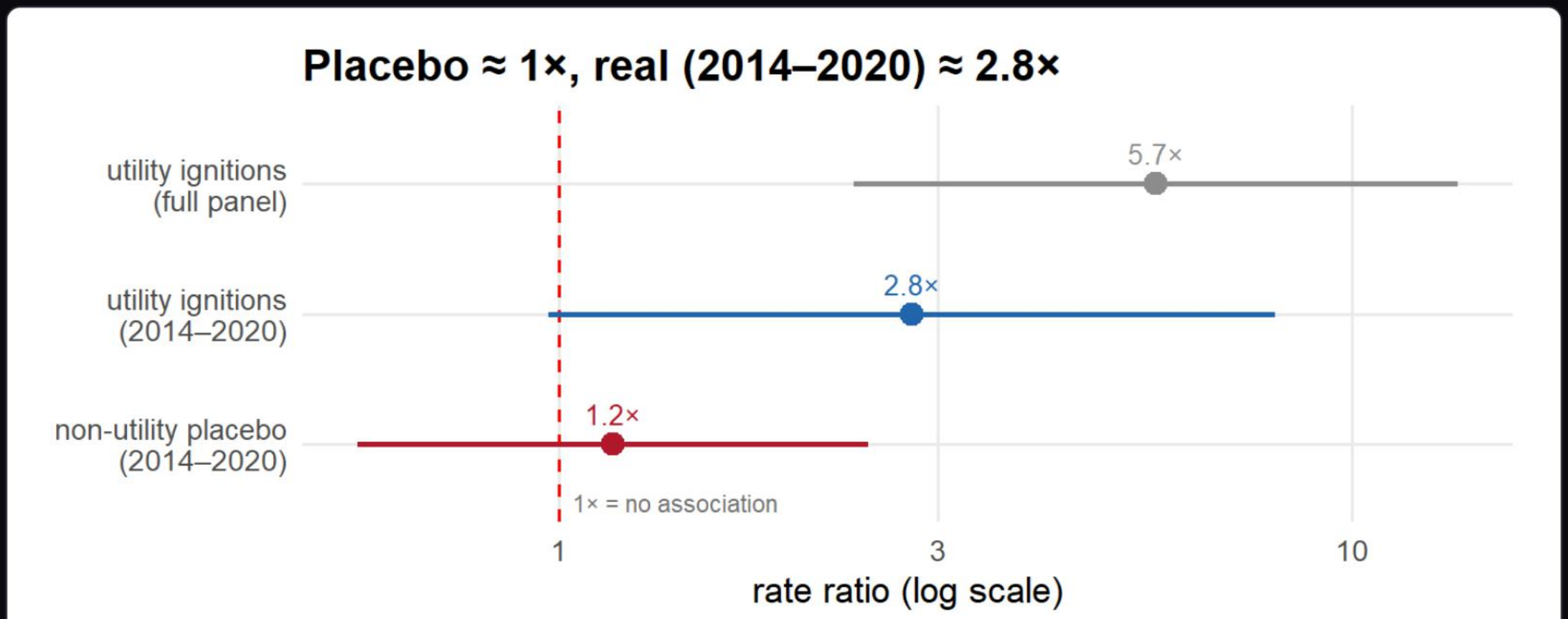
benchmark estimate — outside the MTR band

Why this matters: Under "cutting power doesn't start fires," a positive number cannot be the causal effect of PSPS — and the benchmark's entire CI [+0.0012, +0.0063] sits above zero. What survives adjustment is bias, not effect. Same verdict on SCE: +0.0017 sits outside its MTR band too.

Manski worst-case bounds on the high-risk analysis sample; MTR = monotone treatment response, Y(1) ≤ Y(0).

Falsification: A Clean Placebo, With a Blind Spot

Re-run the matched design on fires PSPS physically cannot prevent — lightning, arson, vehicle fires. The placebo lands near 1× while the real utility outcome over the same years (2014–2020) sits near 2.8×.

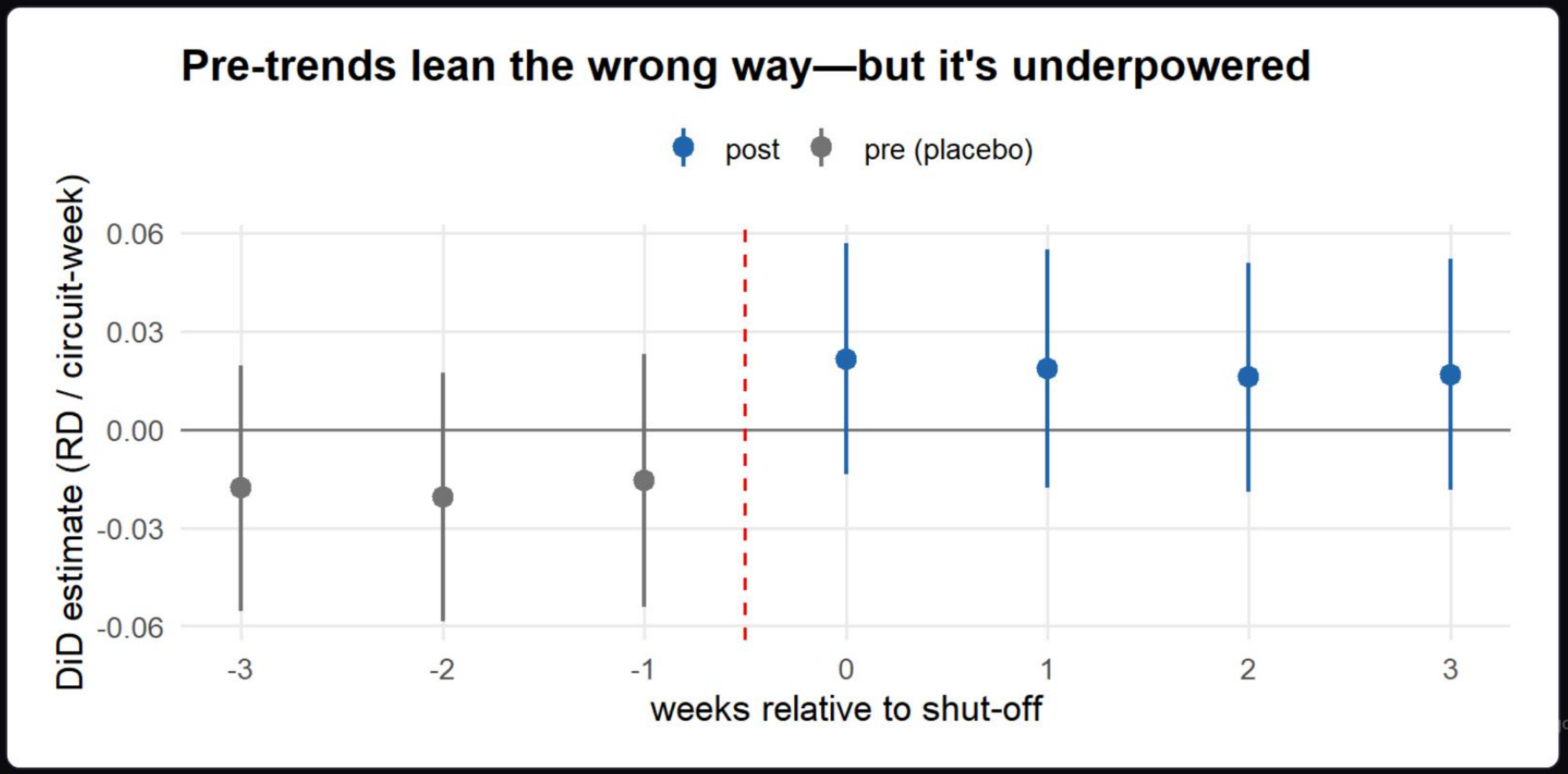


Why this matters: Reassuring — the method isn't conjuring gaps from bad fire weather alone. But a negative control only catches confounders that move both outcomes. The utility's selection on *equipment-specific* risk can't move lightning or arson, so the test is blind to precisely the bias that's biting.

Non-utility ignitions from FPA-FOD (2014–2020 overlap). Placebo RR statistically indistinguishable from 1×.

Two Quasi-Experiments, Two Dead Ends

A switching difference-in-differences (PanelMatch, 1,936 matched sets) returns intervals that all span zero — and pre/post lines that cross at the shutoff, the fingerprint of risk-based timing. The rule-change IV fails outright.

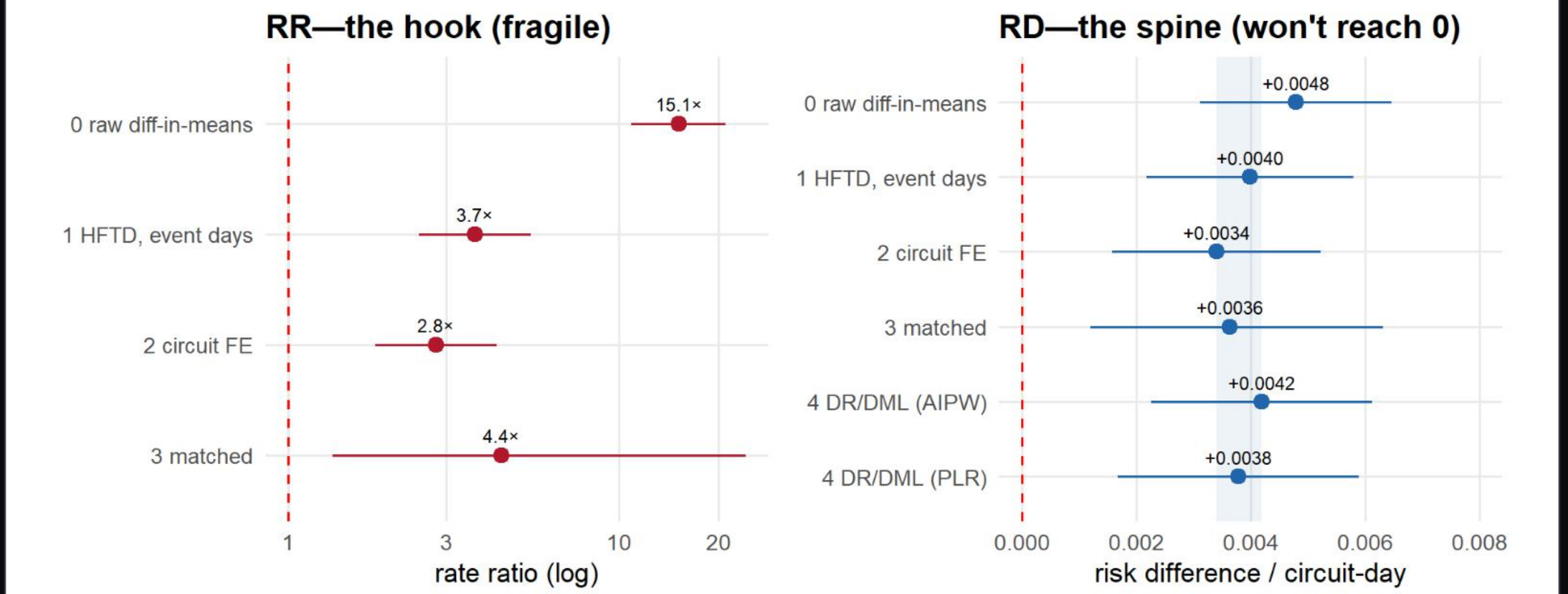


Why this matters: The IV's first stage is enormous (F ≈ 1000) because public "eligibility" is just a relabeled risk tier — and its pre-committed falsification test fails decisively (pre-trend p ≈ 10^{-8}). On public data, "rule" and "risk" cannot be pried apart.

```
DiD: circuit-week PanelMatch, 808 circuits. IV: HFTD-proxy 2019 rollout; ITT 0.65 and
```

Every Public-Data Approach Lands in the Same Place

Steps 0–4 on a common scale: a 15× gap that collapses, then plateaus in a narrow positive band no design can push to zero. Steps 5–9: bounded-but-wide, clean-but-blind, noisy, falsified.



Why this matters: The estimate is stuck on the wrong side of zero — not because PSPS is harmful, but because selection is, and public data can't see past it. The confounder is the operator's private, equipment-specific risk read — it decides who gets treated and never enters the public record.

PG&E panel, all steps. Replicated on SCE: same trap, wider intervals. SDG&E: no public circuit map, no estimate possible.

What Would Actually Answer the Question

Three designs deliver treatment variation that doesn't run through the operator's judgment — each gated on a specific piece of utility-held data.

Design	Identifying variation	Needs from the utility
Border RD	Adjacent locations, same weather — only one side loses power	Circuit & boundary geometry
Fuzzy RD at the FPI cutoff	Circuits just over vs. just under the de-energization threshold	Decision-time FPI values & thresholds
Reform-induced DiD, done right	Rule revisions whose timing was policy-driven, not risk-driven	Revision-rationale record

Why this matters: All three also need the one thing no public source contains: the circuits scoped for a shutoff but left energized — the controls. Utilities publish the rule and the treated circuits, never the controls. This is an industry-wide data-release ask, not a PG&E-specific one.

Public trigger rule (FPI > 0.7, sustained wind > 19 mph, RH < 30%, dead-fuel moisture < 9–11%) is necessary but nowhere near sufficient.

Key Takeaways

1	**PSPS benefits are simulated, not measured.**	The headline "acres avoided" figures are model counterfactuals; the best published study assumes 100% effectiveness. No one had checked against observed ignitions.
2	**The naive comparison is a trap.**	Shut-off days show 15× the ignition rate — but utilities de-energize exactly where risk peaks. The raw gap measures the risk, not the policy. Same pattern at SCE (25×).
3	**Adjustment hits a wall at +0.003 to +0.004.**	Restriction, fixed effects, matching, and doubly-robust ML all land in the same positive band — and under the mild assumption that de-energized lines can't spark, that number has to be bias, not effect.
4	**Utility data could settle it.**	Border RD, fuzzy RD at the FPI cutoff, and a defensible reform DiD are all feasible — but each needs records only utilities hold: scoped-but-not-shut-off controls, decision-time risk scores, boundary geometry, revision rationale.

Read the Full Analysis

igorgeyn.com/blog

Igor Geyn · June 2026