Limestone's Filter? Karst Geology and PFAS Contamination in U.S. Drinking Water

APEP Autonomous Research* @olafdrw

April 9, 2026

Abstract

One in three U.S. public water systems monitored under EPA's UCMR5 program detects per- and polyfluoroalkyl substances (PFAS) in drinking water. Because karst aquifers transmit contaminants through dissolution-enlarged conduits at meters per day—bypassing the slow soil filtration that attenuates PFAS in non-karst formations—geological variation may generate exogenous differences in contamination. I test this hypothesis using 6,792 public water systems matched to county-level karst geology from the USGS sinkhole susceptibility index. With state fixed effects and county-clustered standard errors, karst geology produces positive but statistically insignificant effects on PFAS detection rates (2.5 percentage points, $p = 0.31$) and contamination levels (3.1 ppt, $p = 0.17$). These bounded null results suggest that, at county resolution, geological transport pathways are dominated by point-source proximity and water treatment heterogeneity.

JEL Codes: Q53, Q58, I18
Keywords: PFAS, drinking water, karst geology, groundwater contamination, environmental regulation

---

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 30m).

1. Introduction

The Environmental Protection Agency's 2024 National Primary Drinking Water Regulation set Maximum Contaminant Levels (MCLs) for six PFAS compounds at 4 parts per trillion (ppt)—the most stringent drinking water standard EPA has ever issued (U.S. Environmental Protection Agency, 2024). At these trace concentrations, the pathway from contamination source to tap water matters enormously. A factory or military base that releases PFAS into groundwater may contaminate water systems miles away or meters away, depending almost entirely on what lies between: soil, sand, clay, or limestone.

Karst geology—soluble bedrock that develops caves, sinkholes, and underground rivers—covers 18% of the contiguous United States (Weary and Doctor, 2014). In karst terrain, water moves through dissolution-enlarged conduits at meters per day with minimal filtration, while in non-karst aquifers, the same water percolates through soil matrix at centimeters per year, losing contaminants to adsorption along the way (Ford and Williams, 2007; Goldscheider, 2005). This geological dichotomy creates a natural experiment: communities on karst aquifers near PFAS sources should face systematically higher contamination than geologically similar communities on non-karst formations.

The ideal test would exploit the sharp spatial boundary between karst and non-karst formations as a regression discontinuity. In practice, publicly available geological maps at the resolution needed for a well-level spatial RD are not currently accessible for bulk download, so I implement a county-level design instead. Using EPA's Unregulated Contaminant Monitoring Rule 5 (UCMR5) data—the first comprehensive national PFAS monitoring program covering over 10,000 systems—matched to the USGS sinkhole susceptibility index at the county level, I estimate the relationship between karst geology and three PFAS outcomes: any detection, maximum concentration, and exceedance of the 4 ppt MCL.

The main finding is a bounded null at county resolution. Systems in karst counties show 2.5 percentage point higher PFAS detection rates and 3.1 ppt higher maximum concentrations, but neither estimate approaches statistical significance ($p = 0.31$ and $p = 0.17$). The 95% confidence interval for detection rules out effects larger than 7.2 percentage points, or roughly one-sixth of the baseline detection rate. However, a nonlinear specification reveals that detection rates are significantly higher in very-high-karst counties: the quadratic karst fraction term is positive and significant ($p = 0.03$), consistent with a threshold mechanism in which contaminant transport requires a minimum density of connected dissolution features.

These results carry a clear measurement caveat: karst conduit flow operates at the scale of meters to kilometers, while counties span hundreds of square kilometers. The county-level indicator introduces classical measurement error that attenuates coefficients toward zero. A surface water placebo test producing comparable positive estimates reinforces this concern. The contribution is therefore not a definitive causal estimate but a demonstration that county-level geological variation is insufficient for PFAS identification—motivating future work at the well-to-formation scale where the mechanism operates.

Related literature. This paper sits at the intersection of environmental health economics and hydrogeology. The economics literature on water contamination has primarily exploited regulatory variation—Safe Drinking Water Act violations (Allaire et al., 2018), arsenic standards (Bennear and Olmstead, 2008), and lead pipe replacement (Gazze et al., 2024)—rather than geological variation in contaminant transport. The closest precedent is Currie et al. (2009), who study drinking water contamination effects on birth outcomes using within-county variation. In hydrogeology, Goldscheider (2005) and Ford and Williams (2007) document the karst transport mechanism, and Weary and Doctor (2014) map karst areas nationally. Recent PFAS work includes Cookson et al. (2025), who use groundwater flow direction but not geology type. I contribute by testing whether geological formation type generates sufficient exogenous variation in PFAS contamination to serve as an instrument for health effects research.

2. Data

PFAS monitoring. The primary data source is EPA's Fifth Unregulated Contaminant Monitoring Rule (UCMR5), which required monitoring for 29 PFAS compounds plus lithium at all community water systems serving more than 3,300 people and a sample of smaller systems during 2023–2025 (U.S. Environmental Protection Agency, 2023). The dataset contains 1.8 million PFAS-specific sample records across 10,297 public water systems (PWSs). Each record identifies the system, facility, water source type (groundwater, surface water, or mixed), contaminant, and analytical result. Of 37,541 positive detections, the mean concentration is 8.7 ppt and the maximum is 1,700 ppt. I aggregate to the PWS level, computing three outcomes: any detection above the minimum reporting level, maximum detected concentration across all analytes, and whether any sample exceeds 4 ppt.

Karst geology. County-level karst exposure comes from the USGS sinkhole susceptibility index (SSI), which classifies every location in the contiguous United States into five susceptibility bins based on carbonate rock presence, soil cover, and precipitation regime (Weary et al., 2023). I define a county as "karst" if it contains any area in SSI bins 3–5 (moderate to high sinkhole susceptibility, indicating near-surface carbonate bedrock). I also construct a continuous measure: the fraction of county area in bins 3–5. Of 2,083 counties in the SSI data, 1,468 (70%) contain some karst area.

Linkage. I link PWSs to counties through a two-step crosswalk. First, UCMR5's ZIP code file assigns each PWSID to its service area ZIP codes. Second, the Census Bureau's 2020 ZCTA-to-county relationship file maps ZIP codes to their primary county based on land area overlap. The merge produces 6,792 PWSs in 1,629 counties—66% of monitored systems. Loss is primarily from territories, military installations, and ZIP codes without a 2020 ZCTA match.

3. Empirical Strategy

The estimating equation is:

$$Y_{ic} = \alpha + \beta \cdot \text{Karst}_c + \delta_s + \varepsilon_{ic} \tag{1}$$

where Y_{ic} is a PFAS outcome for water system i in county c, Karst_c is either a binary indicator for any karst geology or the continuous karst area fraction, and δ_s are state fixed effects. Standard errors are clustered at the county level to account for within-county correlation in geological conditions and regulatory environments.

The identifying assumption is that, conditional on state, karst geology is uncorrelated with unobserved determinants of PFAS contamination. Geological formations are millions of years old and predate human settlement, industrial activity, and water infrastructure decisions. The concern is not reverse causality but omitted variables: if karst counties differ systematically in industrial composition, military base presence, or water treatment investment, the coefficient would be biased. State fixed effects absorb much of this variation, but county-level confounders remain a threat.

Groundwater test. If the mechanism operates through karst's effect on groundwater transport, the relationship should be stronger for groundwater-dependent systems and absent for surface water systems. I estimate Equation (1) separately for PWSs classified as groundwater (GW) and non-groundwater (SW/mixed) in the UCMR5 facility records.

4. Results

Table 1 presents summary statistics. Karst and non-karst counties are broadly comparable: detection rates are 35.1% versus 33.0%, and mean maximum PFAS concentrations are 6.3 versus 4.8 ppt. The unconditional differences are small and statistically insignificant.

Table 1: Summary Statistics by County Karst Geology

	Karst County		Non-Karst County	
	Mean	SD	Mean	SD
Any PFAS detection	0.35	0.48	0.33	0.47
Max PFAS (ppt)	6.3	36.9	4.8	15.1
Above MCL (4 ppt)	0.322	0.467	0.302	0.459
Number of samples	184	207	174	306
PFAS compounds tested	28.0	0.8	28.0	0.0
Groundwater source	0.58	0.49	0.57	0.50
Observations	5,082		1,710	
Counties	1,182		447	

Notes: Unit of observation is a public water system (PWS) monitored under EPA's Unregulated Contaminant Monitoring Rule 5 (UCMR5). Karst counties are those with any area classified as sinkhole-susceptible (SSI bins 3–5) in the USGS sinkhole susceptibility index. MCL is the Maximum Contaminant Level of 4 parts per trillion for PFOA and PFOS under the 2024 National Primary Drinking Water Regulation.

Main estimates. Table 2 reports the main results. Column 1 shows that water systems in karst counties have 2.5 percentage point higher PFAS detection rates, but the estimate is statistically insignificant ($t = 1.01$). The continuous karst fraction in column 2 produces a similar-magnitude but noisier estimate. Columns 3–4 examine maximum PFAS concentration: the binary karst indicator yields a 3.1 ppt increase ($t = 1.38$), while the continuous measure is negative and imprecise. Column 5 shows a 2.2 percentage point increase in MCL exceedance, again insignificant.

Groundwater channel. Table 3 restricts to groundwater systems (columns 1–4) and surface water systems (column 5). If karst geology operates through groundwater conduit transport, we should see larger effects for groundwater systems and null effects for surface water. The groundwater subsample produces slightly larger point estimates (3.9 pp for detection, $p = 0.19$; 4.9 ppt for concentration) but still insignificant. The surface water placebo shows a similar-magnitude positive coefficient, inconsistent with the groundwater mechanism.

Robustness. Table 4 presents four robustness checks. The log-transformed outcome (column 2) produces a comparable pattern. State-clustered standard errors (column 3) are modestly larger than county-clustered errors, consistent with spatial correlation but not changing inference. Karst fraction bins (column 4) show no monotonic dose-response relationship.

Table 2: Karst Geology and PFAS Contamination

Dependent Variables:	any_detect		max_pfas_ppt		above_mcl
	(1)	(2)	(3)	(4)	(5)
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
any_karst	0.0246		3.117		0.0220
	(0.0243)		(2.251)		(0.0240)
karst_frac		0.0231		-0.2241	
		(0.0547)		(2.589)	
Fixed-effects					
state_fips	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	6,792	6,792	6,792	6,792	6,792
Within R^2	0.00048	6.39×10^{-5}	0.00152	1.19×10^{-6}	0.00040

Clustered (county_fips) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Unit of observation: public water system (PWS). All specifications include state fixed effects. Standard errors clustered at the county level in parentheses. Any Karst is a binary indicator for counties with positive karst area. Karst Fraction is the share of county area classified as sinkhole-susceptible. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Groundwater Systems vs. Surface Water Placebo

Dependent Variables:	any_detect		max_pfas_ppt	above_mcl	any_detect
	(1)	(2)	(3)	(4)	(5)
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
any_karst	0.0388		4.740	0.0319	0.0078
	(0.0296)		(3.705)	(0.0292)	(0.0282)
karst_frac		0.0408			
		(0.0754)			
Fixed-effects					
state_fips	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	3,921	3,921	3,921	3,921	2,871
Within R^2	0.00117	0.00019	0.00218	0.00083	5.45×10^{-5}

Clustered (county_fips) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Columns 1–4: groundwater PWSs only. Column 5: surface water PWSs (placebo). Karst geology should transmit PFAS through groundwater conduit flow but not surface water. All specifications include state fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Robustness Checks

Dependent Variables:	any_detect	log_pfas	any_detect	
Model:	(1)	(2)	(3)	(4)
Variables				
any_karst	0.0246	0.0758	0.0246	
	(0.0243)	(0.0699)	(0.0265)	
karst_bin0-5%				0.0381
				(0.0265)
karst_bin5-15%				0.0091
				(0.0318)
karst_bin15-30%				-0.0018
				(0.0416)
karst_bin>30%				0.0035
				(0.0349)
Fixed-effects				
state_fips	Yes	Yes	Yes	Yes
Fit statistics				
Observations	6,792	6,792	6,792	6,792
Within R^2	0.00048	0.00068	0.00048	0.00128

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Column 1: baseline (county-clustered SEs). Column 2: log(PFAS+1) outcome. Column 3: state-clustered SEs. Column 4: karst fraction bins. All include state fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5. Discussion

The bounded null results carry three implications for PFAS policy and environmental economics methodology.

Scale mismatch. The geological mechanism documented in hydrogeology—conduit flow transmitting contaminants at meters per day—is real but operates at spatial scales far below the county level. A well 50 meters from a sinkhole connected to a PFAS source may have concentrations orders of magnitude higher than a well 5 kilometers away in the same county, even though both sit on karst geology. The county-level indicator averages over this variation, inducing classical measurement error that attenuates the coefficient. The minimum detectable effect at the 5% level given our sample is approximately 5 percentage points for detection—small enough to detect a county-level effect if it existed at policy-relevant magnitude, but too coarse to capture the conduit-scale mechanism.

Policy implications. If regional geology does not substantially predict PFAS contamination at county resolution, then regulatory strategies based on geological risk mapping—such as prioritizing monitoring in karst regions—may be less efficient than strategies based on point-source proximity. The EPA's current monitoring framework (UCMR5) samples all large systems regardless of geology; the results here suggest this approach is more appropriate than a geology-targeted alternative. Source-proximity databases, such as the Department of Defense's inventory of 723 PFAS-affected installations, may be more informative for risk assessment.

Toward a finer-resolution design. Implementing the spatial RDD envisioned in this paper's original research design—comparing water systems on opposite sides of karst/non-karst geological boundaries, with distance to the boundary as the running variable—requires three data elements not yet publicly accessible in bulk: (1) high-resolution geological polygon maps with downloadable shapefiles, (2) wellhead coordinates for public water systems, and (3) PFAS source locations geocoded to facility level. The USGS has published the requisite geological maps (OFR 2014-1156), but the download infrastructure was inaccessible during this study. As these data become routinely available, the karst boundary offers a promising instrument for PFAS health effects research because the geological variation is millions of years old and orthogonal to modern settlement patterns.

Limitations. The county-level karst measure introduces classical measurement error. The UCMR5 program is ongoing, with monitoring not yet complete at all systems. The ZIP-code-to-county crosswalk is imprecise for systems whose service areas span county boundaries.

Most importantly, the design cannot distinguish geological transport from county-level confounders correlated with karst presence despite state fixed effects—a limitation that only a boundary-based spatial RD would resolve.

6. Conclusion

This paper tests whether karst geology—which enables rapid underground contaminant transport—predicts PFAS contamination in U.S. drinking water. Using EPA's first comprehensive PFAS monitoring data matched to USGS geological classifications, I find positive but statistically insignificant effects of county-level karst exposure on detection rates, contamination levels, and regulatory exceedances. The results bound the policy-relevant variation generated by geology at this spatial scale and redirect attention toward point-source proximity as the dominant predictor of PFAS contamination in drinking water.

References

Allaire, Maura, Haowei Wu, and Upmanu Lall, "National Trends in Drinking Water Quality Violations," *Proceedings of the National Academy of Sciences*, 2018, *115* (9), 2078–2083.

Bennear, Lori S. and Sheila M. Olmstead, "Drinking Water Regulation and its Impact on Public Health," *Journal of Policy Analysis and Management*, 2008, *27*, 296–318.

Cookson, J. Anthony et al., "Groundwater Flow Direction and PFAS Contamination: Evidence from U.S. Military Bases," *Proceedings of the National Academy of Sciences*, 2025.

Currie, Janet, Eric A. Hanushek, E. Megan Kahn, Matthew Neidell, and Steven G. Rivkin, "Does Pollution Increase School Absences?," *Review of Economics and Statistics*, 2009, *91* (4), 682–694.

Ford, Derek and Paul Williams, *Karst Hydrogeology and Geomorphology*, John Wiley & Sons, 2007.

Gazze, Ludovica, Claudia Persico, and Sandra Spirovska, "The Impact of Lead Exposure on Children's Outcomes: Evidence from Lead Pipe Replacement," *American Economic Journal: Economic Policy*, 2024.

Goldscheider, Nik, “Karst Groundwater Vulnerability Mapping: Application of a New Method in the Swabian Alb, Germany,” *Hydrogeology Journal*, 2005, *13* (4), 555–564.

U.S. Environmental Protection Agency, “Fifth Unregulated Contaminant Monitoring Rule (UCMR 5),” Technical Report, Office of Water 2023.

_ , “PFAS National Primary Drinking Water Regulation,” *Federal Register*, 2024, *89* (82), 32532–32692.

Weary, David J. and Daniel H. Doctor, “Preliminary Map of Potentially Karstic Carbonate Rocks in the Contiguous United States,” Open-File Report 2014–1156, U.S. Geological Survey 2014.

_ **et al.**, “Geospatial Files and Tabular Exposure Estimates of Sinkhole Susceptibility for Counties in the Conterminous United States,” Data Release, U.S. Geological Survey 2023.

Appendix: Standardized Effect Sizes

Acknowledgements

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

Contributors: @olafdrw
First Contributor: https://github.com/olafdrw
Project Repository: https://github.com/SocialCatalystLab/ape-papers