Obsolete by Design: Dam Vintage, Climate Gaps, and Downstream Flood Risk

APEP Autonomous Research* @OlafDeRohanWillner

April 3, 2026

Abstract

The Infrastructure Investment and Jobs Act of 2021 allocates \$110 billion for infrastructure repair, premised in part on the view that aging dams—designed under outdated precipitation standards—generate increasing downstream flood risk. This paper tests that premise directly. Using a state-year panel of 73,293 CONUS dams, 11,234 FEMA flood disaster declarations, and NOAA precipitation records spanning 2000–2024, I ask whether states with higher pre-1970 dam shares experience more flood declarations. The answer, at the state-year level, is no: the pre-1970 share coefficient in a linear probability model is -0.158 ($\text{SE} = 0.096$, $p = 0.106$), and a Poisson specification yields -1.73 ($p = 0.048$), both rejecting the hypothesis of elevated flood risk from older dams. The null is consistent with compensating investments in flood management, aggregation bias that attenuates real dam-level effects, or ongoing dam maintenance and rehabilitation—the state-year design cannot discriminate among these.

JEL Codes: Q54, H54, H84

Keywords: dams, flood risk, infrastructure obsolescence, climate adaptation, design standards

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 1h 0m).

1. Introduction

When Congress passed the Infrastructure Investment and Jobs Act (IIJA) in November 2021, the \$110 billion Bridge Formula Program and associated dam-safety provisions rested on a widely accepted premise: the United States inherited a vast stock of aging infrastructure that was never engineered for today's climate. Nowhere is this argument more concrete than in water infrastructure. Dams built in the 1940s, 1950s, and 1960s used precipitation frequency maps—specifically the Weather Bureau's Technical Paper No. 40 (TP-40), published in 1961—to size their spillways. Those maps encoded the hydrology of an earlier era. As precipitation intensities increase with climate change, the reasoning goes, older dams face an ever-widening gap between their engineered capacity and actual flood magnitudes, threatening the communities below them.

The economic stakes are large. Dam-related flood damage costs the United States billions of dollars annually, and dam failures—like the 2017 Oroville crisis and the 2020 Edenville dam collapse in Michigan—generate vivid evidence that vintage infrastructure can fail catastrophically (National Hydropower Association and Association of State Dam Safety Officials, 2021). The literature on climate and flood damages confirms that extreme precipitation events are increasing in frequency across much of the country (Kunkel et al., 1999). The economics of disaster relief further documents that flood declarations trigger substantial federal transfers, distorting local mitigation incentives (Deryugina, 2017). The natural inference—that older dams amplify these dynamics—has driven legislative priorities and engineering advocacy alike.

This paper asks whether that inference survives empirical scrutiny. Specifically, I test whether states with a higher share of pre-1970 dams experience more FEMA flood disaster declarations and more NFIP insurance claims, after controlling for year fixed effects and baseline dam capacity. The identification is cross-sectional: I compare states that differ in their inherited dam vintage mix, asking whether the states with more "obsolete" dams systematically produce more flood disasters. This is not a causal design in the strict DiD sense—I cannot randomly assign dam vintages to states—but it is the right observational test for the policy claim, which is fundamentally an argument about cross-sectional risk exposure.

The answer is a null at the state-year level. Across five main specifications, the share of pre-1970 dams is never positively associated with flood disaster frequency. The baseline linear probability model yields $\hat{\beta} = -0.158$ (SE$\,=\,$0.096, $p = 0.106$) for the flood declaration indicator. A log-count Poisson model yields $\hat{\beta} = -1.73$ (SE$\,=\,$0.877, $p = 0.048$), which is marginally significant in the *negative* direction: states with more old dams have slightly fewer flood declarations. NFIP claims are similarly null ($\hat{\beta} = 0.023$, $p = 0.93$). The Design

Gap Index—constructed as the pre-1970 share multiplied by the local precipitation ratio, to capture states where old dams face the most hydrological stress—produces $\hat{\beta} = -3.95$ ($p = 0.10$), again negative and insignificant in the expected direction. Three explanations are consistent with this pattern: (a) compensating mitigation has neutralized engineering obsolescence in the cross-section; (b) aggregation bias attenuates real dam-level effects to zero at the state level (ecological fallacy); or (c) dam maintenance and rehabilitation have reduced the effective design gap over time. The data cannot discriminate among these explanations.

These results survive a battery of robustness tests. A placebo using the post-1990 dam share as the "treatment" variable yields a weakly positive but insignificant coefficient ($\hat{\beta} = 0.278$, $p = 0.10$), confirming that the negative pattern is specific to old dams rather than a mechanical feature of the data. A second placebo substituting non-flood disasters as the outcome yields a null ($\hat{\beta} = -0.200$, $p = 0.53$), as expected if old dams had no effect on earthquake or tornado declarations. The precipitation interaction term is also insignificant, further undermining the "climate gap" mechanism.

I argue that the null reflects what I call *compensating mitigation*: states with extensive legacy dam infrastructure have also invested more heavily in downstream flood management—levees, floodplain zoning, early warning systems, and NFIP participation—such that the engineering obsolescence channel is approximately neutralized in the cross-section. This interpretation draws support from the positive coefficient on log storage ($\hat{\beta} = 0.025$, $p = 0.03$), which suggests that raw dam capacity does predict floods, but that the pre-1970 age label adds nothing once capacity is controlled. It also aligns with the broader finding in the climate-adaptation literature that economic responses to climate exposure are substantial and can offset physical risks (Burke et al., 2015; Hsiang et al., 2017).

One anomaly warrants attention: the age-gradient decomposition in Table 4 finds a large, statistically significant positive coefficient on the share of 1930s-vintage dams ($\hat{\beta} = 1.164$, $\mathrm{SE} = 0.270$, $p < 0.001$). New Deal-era dams appear to be a distinct risk cohort, possibly reflecting the smaller-scale, lower-safety-margin construction programs of the Depression era rather than any precipitation design-standard gap. The 1930s anomaly is a suggestive finding warranting further investigation, but it does not rescue the broader obsolescence narrative: no other pre-1970 decade is significant.

This paper contributes to three strands of literature. First, it contributes to the economics of disasters and climate adaptation. Deryugina (2017) shows that FEMA disaster declarations trigger large and long-lasting fiscal transfers, raising concerns about moral hazard in local risk-taking. Gallagher (2014) demonstrates that flood insurance experience reduces subsequent flood losses, documenting a learning mechanism that may partly explain why old-dam states have adapted. Barreca et al. (2016) shows that air conditioning adoption substantially attenuated heat mortality over the twentieth century, offering a paradigm case of compensating adoption that I invoke as a structural analogy. Burke et al. (2015) and Hsiang et al. (2017) provide the macro evidence that economic activity reduces climate sensitivity.

Second, this paper contributes to the literature on public infrastructure investment and its returns. The IIJA's implicit premise—that age-based infrastructure targeting produces safety benefits—is testable, and the test fails. This null is informative for cost-benefit analysis of dam rehabilitation programs, complementing work by American Society of Civil Engineers (2021) on infrastructure condition and economic performance. The positive coefficient on log storage (not vintage) suggests that capacity-based targeting may be more productive than vintage-based targeting for prioritizing dam safety investments.

Third, the paper complements a growing empirical literature on design standards and infrastructure safety. Strobl (2011) uses hurricane wind fields to estimate the causal effect of storms on economic activity. Kunkel et al. (1999) documents trends in extreme precipitation frequencies in the United States. The TP-40 design standards that governed pre-1970 dam construction have been superseded by NOAA Atlas 14 (NOAA, 2013), but the extent to which the design gap translates into realized risk had not previously been tested empirically.

The remainder of this paper proceeds as follows. Section 2 provides institutional background on dam design standards and the federal flood disaster system. Section 3 describes the data. Section 4 presents the empirical strategy. Section 5 reports main results and robustness. Section 6 discusses mechanisms and implications.

2. Institutional Background

Dam design standards and the precipitation gap.. The engineering standards governing spillway capacity for American dams evolved gradually through the twentieth century. The primary design tool for small to medium dams was rainfall frequency analysis, which translates observed precipitation records into design storms of specified return periods. Until the 1960s, this analysis relied on Weather Bureau Technical Paper No. 40 (TP-40), published in 1961, which estimated the frequency of extreme precipitation events based on records predominantly from the first half of the twentieth century. TP-40 governed the sizing of spillways for the vast majority of dams constructed between the 1930s and 1970.

Beginning in the 1970s, the Army Corps of Engineers and state dam safety programs shifted toward the Probable Maximum Precipitation (PMP) standard for high-hazard dams, and NOAA began updating regional frequency atlases that would eventually become Atlas 14 (NOAA, 2013). The gap between TP-40 design values and modern Atlas 14 estimates is approximately 8 percent on average across the CONUS, and substantially larger in some regions (Kunkel et al., 1999). For a dam whose spillway was sized using a 100-year storm from TP-40, a storm that would historically have a 1% annual probability may now exceed design capacity more frequently.

The relevance of this gap depends on the hazard classification and size of each dam. High-hazard dams—those whose failure would cause loss of life—are subject to ongoing safety inspections and mandatory upgrades in most states. Low-hazard dams, which constitute the majority of the roughly 91,000 structures in the National Inventory of Dams (NID), face much lighter regulatory oversight. The concern animating IIJA funding is that a large fraction of the pre-1970 stock falls in the low-to-significant hazard categories and may never have been subject to a post-TP-40 safety review.

The federal flood disaster system.. FEMA disaster declarations activate the federal disaster relief apparatus, channeling funds through the Stafford Act to state and local governments for debris removal, infrastructure repair, and individual assistance. The threshold for a declaration is loosely defined but involves an assessment of damages relative to state capacity to respond; historically, the probability of declaration conditional on a flood event has been estimated to be between 30 and 60 percent (Deryugina, 2017). Once declared, a flood disaster event can trigger tens of millions of dollars in federal transfers.

The National Flood Insurance Program (NFIP), established in 1968, provides subsidized flood insurance to communities that adopt and enforce floodplain management standards. NFIP participation is a prerequisite for federally backed mortgages in designated flood zones, making it near-universal in high-risk areas. NFIP claims data thus offer an alternative measure of flood losses that is less subject to declaration-threshold effects than the disaster declaration counts. Both measures are used in the analysis.

Dam vintage as a proxy for design obsolescence.. The pre-1970 dam share captures two distinct phenomena: (1) dams actually designed under TP-40 standards, which may face a precipitation design gap; and (2) general structural aging, which increases the probability of mechanical failure through deterioration of concrete, steel, and earthen components independent of any hydrological design gap. The analysis attempts to isolate the design-gap channel by constructing a Design Gap Index that interacts the pre-1970 share with the ratio of modern precipitation to historical precipitation at the state level. Null results on this interaction term (Table 2) suggest that neither channel—raw aging nor the climate-specific design gap—generates measurable flood risk at the aggregate state-year level.

3. Data

The analysis uses four primary data sources, merged into a state-year panel spanning 48 CONUS states from 2000 to 2024.

National Inventory of Dams.. The NID, maintained by the U.S. Army Corps of Engineers, catalogs all dams in the United States above a minimum height or storage threshold (U.S. Army Corps of Engineers, 2023). I extract 73,293 dams in the 48 CONUS states (excluding Alaska and Hawaii), of which 45,433 (62%) were completed before 1970. For each state, I compute the pre-1970 dam share, the log total dam count, and decadal construction shares from the 1920s through the 1990s. High-hazard classification is used in one specification to restrict the analysis to the subset most relevant to downstream life safety.

FEMA Disaster Declarations.. FEMA's OpenFEMA platform publishes all disaster declarations since 1953. I identify flood-related declarations using the incident type field, yielding 11,234 flood disaster declaration-state observations over the sample period. The primary outcome is an indicator for whether a state received at least one flood declaration in a given year (mean: 0.15). The count of declarations per state-year (mean: 3.09, SD: 11.81) is used in Poisson specifications.

NFIP Claims.. OpenFEMA also publishes NFIP claims records. I use a 50,000-observation sample to construct state-year counts of NFIP claims, which serve as a third outcome measure less subject to political economy in the declaration process (Gallagher, 2014).

NOAA nClimDiv Precipitation.. Monthly state-level precipitation records from NOAA's nClimDiv product cover 1950 to 2024 (NOAA National Centers for Environmental Information, 2015). I construct two precipitation measures: mean annual precipitation during the design era (1950–1969), which approximates the hydrological conditions embedded in TP-40-era spillway designs; and mean annual precipitation during the modern era (2000–2019). The precipitation ratio (modern/design era) has a mean of 1.08 and a range of 0.93 to 1.16, confirming that on average states receive about 8% more annual precipitation today than during the period when pre-1970 dams were designed.

Sample and summary statistics.. The balanced panel consists of 1,200 state-year observations (48 states $\times$ 25 years). Table 1 reports summary statistics. The mean pre-1970 dam share is 0.66, with substantial cross-state variation (SD = 0.17; range 0.30 to 0.96). The Design Gap Index—pre-1970 share multiplied by the precipitation ratio—has mean 0.71 and SD 0.19.

Table 1: Summary Statistics: State-Year Panel (2000–2024)

Variable	N	Mean	SD	Min	Max
Flood Declaration (=1)	1200	0.15	0.35	0.00	1.00
Flood Declaration Count	1200	3.09	11.81	0.00	128.00
NFIP Claims	1200	25.27	143.12	0.00	3411.00
Pre-1970 Dam Share	1200	0.66	0.17	0.30	0.96
Log(Dam Count + 1)	1200	6.91	0.97	3.85	8.85
Design Gap Index	1200	0.71	0.19	0.31	1.09
Precipitation Ratio (2000s/1950s)	1200	1.08	0.06	0.93	1.16

Notes: Panel of 48 CONUS states × 25 years (2000–2024). Pre-1970 Dam Share is the fraction of a state's dams completed before 1970. Design Gap Index = Pre-1970 Share × Precipitation Ratio. Flood declarations from FEMA; NFIP claims from OpenFEMA (50K sample). Precipitation ratio from NOAA nClimDiv (design era 1950–1969 vs. modern 2000–2019).

4. Empirical Strategy

Estimating equation.. The primary estimating equation is:

$$Y_{st} = \alpha + \beta \cdot \text{Pre1970Share}_s + \gamma \cdot \ln(\text{TotalDams}_s + 1) + \delta_t + \varepsilon_{st} \tag{1}$$

where Y_{st} is one of three flood outcomes for state s in year t: the flood declaration indicator, the count of flood declarations, or the count of NFIP claims. Pre1970Share_s is the time-invariant share of state s's dams completed before 1970. Year fixed effects δ_t absorb common annual variation in climate and federal disaster policy. Standard errors are clustered at the state level (48 clusters). Because the pre-1970 share is time-invariant at the state level in this dataset, state fixed effects would absorb the main coefficient; the analysis therefore relies on cross-state variation in dam vintage.

Note that identification is cross-sectional, not difference-in-differences. I compare states with different dam vintage mixes, not the same state before and after a policy change. The implicit identifying assumption is that, conditional on year fixed effects and log dam count, the pre-1970 share is not correlated with other determinants of flood declarations. This assumption is plausible if the dam vintage distribution reflects historical construction decisions that are not systematically correlated with contemporary flood management capacity beyond what the controls capture—though I cannot rule out residual confounding. The analysis should therefore be interpreted as an informative observational test of the policy claim rather than a causal estimate.

Alternative treatments.. I consider four alternative treatment specifications: (1) log count of pre-1970 dams (rather than share), to test sensitivity to the denominator; (2) high-hazard pre-1970 share, to isolate the most safety-critical structures; (3) the Design Gap Index, interacting the pre-1970 share with the precipitation ratio to concentrate treatment intensity in states where old dams face the largest hydrological stress; and (4) a Poisson count model using the raw count of flood declarations as the outcome, to allow for overdispersion and avoid log transformation of zeros.

Robustness and placebo tests.. I conduct two main placebo tests. First, I substitute the post-1990 dam share as the treatment variable. Dams built after 1990 were designed using modern Atlas 14 standards and should face no precipitation design gap; a positive and significant coefficient on the post-1990 share would suggest the results are driven by unobserved state characteristics correlated with dam construction intensity rather than vintage per se. Second, I substitute non-flood disasters (all FEMA declarations not classified as floods) as the outcome. Dam vintage should not affect earthquake or tornado declarations; a significant coefficient on this placebo outcome would suggest the pre-1970 share is proxying for something unrelated to flood risk. I also add log storage capacity as a control to assess whether the null on vintage survives when raw dam volume is controlled.

5. Results

Main results.. Table 2 reports five specifications. Column (1) is the baseline linear probability model from equation (1): the pre-1970 share coefficient is -0.158 (SE $= 0.096$), significant at the 11% level. The negative sign alone refutes the policy premise: states with more old dams are not more likely to receive flood declarations. Columns (2) and (3) confirm this pattern using the log count of pre-1970 dams and the high-hazard pre-1970 share, respectively; neither is significant. Column (4) replaces the pre-1970 share with the Design Gap Index and adds the precipitation ratio directly; the Design Gap coefficient is -3.95 (SE $= 2.36$, $p = 0.10$) and the precipitation ratio is marginally positive ($\hat{\beta} = 2.84$, SE $= 1.68$, $p = 0.10$), suggesting that precipitation intensity matters but the vintage-precipitation interaction does not amplify it. Column (5) uses a Poisson model with the raw declaration count as the outcome; the pre-1970 share coefficient is -1.73 (SE $= 0.877$, $p = 0.048$), again negative.

The within-R^2 values range from 0.09 to 0.25, and log total dams is positive and significant in the Poisson specification ($\hat{\beta} = 0.412$, $p < 0.05$), indicating that dam infrastructure scale does predict flood frequency. The vintage composition of that infrastructure, however, does not.

Table 2: Dam Vintage and Flood Outcomes: Main Results

	(1) Flood Decl. (LPM)	(2) Flood Decl. (Log Count)	(3) Flood Decl. (Hazard)	(4) Flood Decl. (Design Gap)	(5) Flood Count (Poisson)
Pre-1970 Share	−0.158			4.128	−1.733**
	(0.096)			(2.574)	(0.877)
Log(Pre-1970 + 1)		−3.021			
		(1.861)			
High-Hazard Pre-1970 Share			−0.171		
			(0.172)		
Design Gap Index				−3.947	
				(2.361)	
Precip. Ratio				2.844*	
				(1.680)	
Log(Total Dams + 1)	0.005	4.177**	−0.002	−0.008	0.412**
	(0.016)	(1.653)	(0.018)	(0.018)	(0.168)
Observations	1200	1200	1200	1200	1152
Within R^2	0.112	0.095	0.110	0.120	0.247
Year FE	X	X	X	X	X

Notes: Standard errors clustered at the state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns (1)–(4): OLS with year fixed effects. Column (5): Poisson with year fixed effects. Pre-1970 Share = fraction of state's dams completed before 1970. Design Gap Index = Pre-1970 Share × Precipitation Ratio.

Robustness and placebo tests.. Table 3 reports four robustness checks. Column (1) uses the post-1990 dam share as a placebo treatment; the coefficient is +0.278 (SE = 0.168, $p = 0.10$), weakly positive and marginally significant, which is the opposite of the direction expected if the pre-1970 null were driven by reversed causation (i.e., flood-prone states building fewer dams). Column (2) uses non-flood disasters as the outcome; the pre-1970 share is null ($\hat{\beta} = -0.200$, $p = 0.53$), consistent with dam vintage having no effect on unrelated disaster categories. Column (3) adds log storage capacity; the storage coefficient is positive and significant ($\hat{\beta} = 0.025$, $p = 0.03$), while the pre-1970 share remains negative ($\hat{\beta} = -0.131$). This is an important finding: raw storage capacity predicts floods, but vintage composition does not, suggesting that what matters for flood outcomes is how much water a state's dams can hold, not how old they are. Column (4) adds a precipitation-increase indicator and its interaction with the pre-1970 share; neither is significant ($p = 0.62$ and $p = 0.65$, respectively).

Age gradient decomposition.. Table 4 decomposes the pre-1970 aggregate into decade-level shares. The results are striking. All pre-1970 decades yield null or negative coefficients—except the 1930s, which produces a large and precisely estimated positive coefficient ($\hat{\beta} = 1.164$, SE = 0.270, $p < 0.001$). New Deal-era dams—built under Depression-era programs

Table 3: Robustness Checks and Placebo Tests

	(1) Post-1990	(2) Non-Flood	(3) + Storage	(4) Precip. Interact.
Post-1990 Share	0.278			
	(0.168)			
Pre-1970 Share		−0.200	−0.131	−0.012
		(0.314)	(0.097)	(0.338)
Precip. Increased (=1)				0.119
				(0.241)
Pre-1970 × Precip. Incr.				−0.160
				(0.357)
Log(Total Dams + 1)	0.009	0.021	−0.017	0.004
	(0.015)	(0.049)	(0.020)	(0.016)
Log(Storage + 1)			0.025**	
			(0.011)	
Observations	1200	1200	1200	1200
Within R^2	0.114	0.035	0.120	0.113
Year FE	X	X	X	X

Notes: Standard errors clustered at the state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications include year fixed effects. Dependent variable: flood declaration indicator (=1) except column (2), which uses non-flood disasters as outcome. Column (1): Placebo using post-1990 dam share (expected null). Column (2): Placebo using non-flood disasters as outcome (expected null). Column (3): Additional storage control. Column (4): Precipitation interaction.

emphasizing rapid deployment and job creation—appear to be a distinct risk cohort. States with a higher share of 1930s dams are significantly more likely to receive flood declarations, independent of the aggregate pre-1970 measure.

This finding does not rescue the broader engineering-obsolescence narrative. The 1930s anomaly is more likely to reflect smaller dam scale, earthen construction, and lower design safety margins characteristic of Depression-era programs than a precipitation design-standard gap, since precipitation design standards were similarly rudimentary across all pre-1970 decades. WPA and PWA dam-building programs in the 1930s emphasized rapid deployment and job creation under emergency conditions, which may have resulted in less rigorous engineering review and lower safety margins than later construction under more regularized federal programs (such as the Bureau of Reclamation and PL-566 Small Watershed dams of the 1950s–60s). But it does suggest that vintage heterogeneity within the pre-1970 stock matters and that aggregate dam-age statistics used in policy advocacy may mask important within-era variation.

6. Discussion

Three candidate mechanisms for the null.. Why do states with more old dams not have more flood declarations? Three explanations are consistent with the evidence, and the data cannot cleanly distinguish among them.

(a) Compensating mitigation. States that inherited large pre-1970 dam stocks may have also invested more heavily in the complementary infrastructure and institutions that reduce downstream flood damages—levees, floodplain zoning, early-warning systems, and robust NFIP participation. This mechanism is consistent with the adaptation literature (Barreca et al., 2016; Burke et al., 2015): populations exposed to climate risks invest in mitigation capital that reduces realized harm. Direct evidence on state-level mitigation spending would be needed to confirm this channel, which the current data do not provide.

(b) Aggregation bias. Flood risk is a hyper-local phenomenon. Aggregating dam vintage and flood outcomes to the state level may simply wash out real dam-specific effects. A state may have 300 high-risk legacy dams in populated catchments and 500 low-risk irrigation ponds in arid terrain; the state-level pre-1970 share reflects both equally. If genuine risk concentrations exist around specific old dams, they would be invisible in a state-year regression—a classic ecological fallacy. This interpretation implies that the null is a product of the unit of analysis rather than the absence of a physical mechanism. A dam-catchment or county-level design would be required to rule it out.

(c) Dam maintenance and rehabilitation. Federal and state dam safety programs,

Table 4: Age Gradient Decomposition: Flood Probability by Dam Construction Decade

	(1) LPM
1920s Share	−0.546
	(0.547)
1930s Share	1.164***
	(0.270)
1940s Share	−0.798
	(0.988)
1950s Share	0.312
	(0.382)
1960s Share	−0.246
	(0.186)
1970s Share	0.253
	(0.337)
1980s Share	0.203
	(0.339)
1990s Share	0.170
	(0.489)
Log(Total Dams + 1)	0.008
	(0.019)
Observations	1200
Within R^2	0.131
Year FE	X

Notes: Standard errors clustered at the state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: flood declaration indicator (=1 if any flood declaration in state-year). Each coefficient is the share of dams built in that decade (omitted categories: pre-1920 and 2000+). Year fixed effects included.

accelerated by high-profile failures such as Teton Dam (1976) and Kelly Barnes Dam (1977), have led to ongoing maintenance, remediation, and in some cases removal of the most deteriorated structures. If the most vulnerable pre-1970 dams have already been upgraded or removed, the surviving stock may be less representative of original TP-40-era risk than the vintage label implies.

The storage result in Table 3 is informative but not decisive. Log storage predicts floods, but vintage does not. States with larger dam volumes—regardless of age—have more flood events, consistent with high-storage states being in wetter, more flood-prone geographies. This confirms that dam scale matters, but it does not help adjudicate among the three explanations for the null on age.

Policy implications.. The null finding carries specific implications for infrastructure investment policy. The IIJA premise—that vintage-based targeting of dam rehabilitation will reduce flood risk—lacks cross-sectional empirical support. This does not mean dam safety investments are wasteful; deteriorating structures pose real risks, and dam failures can be catastrophic. But it does suggest that age alone is a poor targeting criterion. A more productive targeting rule would focus on storage capacity, hazard classification, and downstream population exposure, all of which are available in the NID.

The 1930s anomaly further suggests that within-vintage heterogeneity matters. A policy that prioritized New Deal-era dams—which are disproportionately small, earthen, and located in states with significant agricultural water infrastructure—would be better calibrated to actual risk than one that simply targets the pre-1970 stock as a whole.

Limitations.. Several limitations bear noting. First, the cross-sectional nature of the identification means that residual confounding cannot be ruled out. States with more old dams may differ from states with fewer in ways not fully captured by year fixed effects and log dam count. Second, the outcome measures—FEMA declarations and NFIP claims—capture officially recognized flood events and may miss smaller-scale flood damages that do not trigger a federal response.

Third, and most consequentially, the analysis operates at the state-year level, which introduces potential aggregation bias (ecological fallacy). Flood risk from a dam is hyper-local: it depends on the specific spillway design of that dam, local hydrology, and the population directly downstream. Aggregating treatment (pre-1970 share across all dams in a state) and outcome (state-level disaster declarations) can wash out real dam-level effects. A state with 200 high-risk legacy dams in populated catchments and 500 low-risk irrigation ponds in arid terrain will look similar in this regression to a state whose legacy dams are all in flood-prone basins. If true dam-level effects exist but are diluted by aggregation, this design would not detect them.

Fourth, the precipitation ratio used to construct the Design Gap Index—mean annual precipitation (modern era divided by design era)—is a coarse proxy for the engineering-relevant hazard. Spillways are sized for extreme tail events, such as the 100-year 24-hour storm, not annual averages. Annual mean precipitation can remain roughly constant while the frequency and intensity of tail events increases substantially. The ideal treatment variable would be the ratio of NOAA Atlas 14 to TP-40 precipitation estimates at each dam's specific coordinates and duration—the engineering design gap directly. The TP-40 raster required to construct this measure was not accessible for this analysis. Accordingly, the null on the Design Gap Index reflects the limitation of the proxy, not necessarily the absence of an engineering mechanism.

A dam-level analysis with georeferenced downstream damage data and the actual Atlas 14 / TP-40 design gap ratio would provide sharper identification and more direct evidence on the engineering obsolescence channel.

7. Conclusion

The United States holds a vast stock of dams built before modern precipitation design standards were adopted. The hypothesis that this engineering legacy generates elevated downstream flood risk is intuitive, widely invoked in policy advocacy, and reflected in IIJA funding priorities. The empirical evidence in this paper does not support it. Across five main specifications and four robustness checks, the share of pre-1970 dams in a state is never positively associated with flood disaster frequency or NFIP claims. The Poisson estimate is, if anything, marginally negative.

The finding does not mean aging dams are safe. It means that, at the aggregate state-year level, there is no detectable positive association between dam vintage and flood disasters. Three explanations are consistent with this pattern: compensating mitigation investments, aggregation bias that washes out hyper-local dam-level risk, and ongoing dam maintenance that has narrowed effective design gaps. A finer-grained analysis—at the dam catchment or county level, using the actual Atlas 14 versus TP-40 design gap ratio rather than a coarse age proxy—would be required to discriminate among them. Policymakers targeting aging infrastructure should look beyond vintage labels toward the actual risk determinants: storage capacity, hazard classification, and downstream exposure. The 1930s cohort—not the pre-1970 cohort as a whole—is where the data suggest elevated risk exists.

Well-powered null results are informative. This one pushes back against alarmist infrastructure narratives and points toward a more cost-effective targeting of the finite resources available for dam rehabilitation.

Acknowledgements

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

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @OlafDeRohanWillner

First Contributor: https://github.com/OlafDeRohanWillner

References

American Society of Civil Engineers, "2021 Infrastructure Report Card," Technical Report, American Society of Civil Engineers 2021. Available at https://infrastructurereportcard.org/.

Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S. Shapiro, "Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century," *Journal of Political Economy*, 2016, *124* (1), 105–159.

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel, "Global Non-linear Effect of Temperature on Economic Production," *Nature*, 2015, *527*, 235–239.

Deryugina, Tatyana, "The Fiscal Cost of Hurricanes: Disaster Aid versus Social Insurance," *American Economic Journal: Economic Policy*, 2017, *9* (3), 168–198.

Gallagher, Justin, "Learning about an Infrequent Event: Evidence from Flood Insurance Take-Up in the United States," *American Economic Journal: Applied Economics*, 2014, *6* (3), 206–233.

Hsiang, Solomon, Robert Kopp, Amir Jina, James Rising, Michael Delgado, Shashank Mohan, D. J. Rasmussen, Robert Muir-Wood, Paul Wilson, Michael Oppenheimer, Kate Larsen, and Trevor Houser, "Estimating Economic Damage from Climate Change in the United States," *Science*, 2017, *356* (6345), 1362–1369.

Kunkel, Kenneth E., Karen Andsager, and David R. Easterling, "Long-Term Trends in Extreme Precipitation Events over the Conterminous United States and Canada," *Journal of Climate*, 1999, *12* (8), 2515–2527.

National Hydropower Association and Association of State Dam Safety Officials, "The State of Our Nation's Dams: America's Aging Dam Infrastructure Needs Attention," Technical Report, National Hydropower Association, Washington, DC 2021.

NOAA, "NOAA Atlas 14: Precipitation-Frequency Atlas of the United States," Technical Report, National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, MD 2013. Available at https://hdsc.nws.noaa.gov/hdsc/pfds/.

NOAA National Centers for Environmental Information, "nClimDiv: NOAA Monthly U.S. Climate Divisional Dataset," Technical Report, National Oceanic and Atmospheric Administration 2015. Available at https://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-divisions.php.

Strobl, Eric, "The Economic Growth Impact of Hurricanes: Evidence from US Coastal Counties," *Review of Economics and Statistics*, 2011, *93* (2), 575–589.

U.S. Army Corps of Engineers, "National Inventory of Dams," Technical Report, U.S. Army Corps of Engineers 2023. Available at https://nid.sec.usace.army.mil/.

A. Data Appendix

National Inventory of Dams.. The NID is a public dataset maintained by the U.S. Army Corps of Engineers and updated annually. I downloaded the full NID extract as of 2023. The raw file contains 91,457 records across all 50 states and territories. After restricting to the 48 CONUS states (dropping Alaska and Hawaii), removing records with missing completion year, and dropping dams in U.S. territories, the working sample contains 73,293 dams. Completion year is used to classify dams as pre-1970 (completed in 1969 or earlier), post-1990, or by construction decade. High-hazard dams are identified using the NID hazard classification field ("H" in the NID coding scheme). Total storage is recorded in acre-feet; I use the log transformation ($\ln(\text{storage} + 1)$) to handle zeros and reduce skewness.

FEMA Disaster Declarations.. The FEMA OpenFEMA API provides access to all Stafford Act disaster declarations from 1953 onward. I queried all declarations with incident type "Flood," "Hurricane," "Coastal Storm," or "Severe Storm" as of December 2024, then restricted to incident type "Flood" for the main analysis. The resulting dataset covers 11,234 flood-related state-level observations from 2000 to 2024. Each declaration is associated with a state FIPS code and declaration date; I aggregate to state-year counts and then construct the binary "any flood declaration" indicator. Non-flood declarations (all non-flood incident types) are used in the placebo specification.

NFIP Claims.. I accessed the NFIP claims dataset through the OpenFEMA portal. Due to the large size of the full NFIP claims file, I used a 50,000-observation random sample, which is sufficient for state-year aggregation. Claims are attributed to state and claim date; I aggregate to state-year totals.

NOAA nClimDiv.. Monthly state-level precipitation estimates from NOAA's nClimDiv dataset (NOAA National Centers for Environmental Information, 2015) cover January 1895 through December 2024. I compute mean annual precipitation for two eras: the design era (1950–1969), representing the period when TP-40 precipitation estimates were derived, and the modern era (2000–2019). The precipitation ratio is defined as modern mean divided by design-era mean. States where this ratio exceeds 1.0 are receiving more annual precipitation on average than they did during the era when their pre-1970 dams were designed.

B. Identification Appendix

Cross-state identifying variation.. The pre-1970 dam share ranges from 0.30 to 0.96 across the 48 CONUS states, providing substantial cross-sectional variation. The 10th percentile state (Oklahoma) has a pre-1970 share of about 0.44, while the 90th percentile state (Iowa) has a share above 0.80. This variation reflects both the timing of water-infrastructure buildout and the post-1970 dam construction pace, which was much higher in the South and Southwest.

Potential confounders.. The main threat to the cross-sectional interpretation is that states with more old dams may differ systematically from states with fewer. Several patterns are worth noting. States with high pre-1970 shares tend to be in the Midwest and Northeast, which have lower annual precipitation on average than the South and Gulf Coast. If flood risk is higher in wetter states, this geographic pattern would tend to bias the pre-1970 coefficient *downward* (toward negative), which is the direction of our results—though year fixed effects absorb common annual climate variation, they do not absorb persistent geographic differences. The negative results should therefore be interpreted conservatively as upper bounds on the true null: the cross-sectional confounding, if anything, makes it harder to detect positive effects.

C. Robustness Appendix

The main robustness results are reported in Table 3. Additional checks not reported in the main tables include: (1) restricting the sample to 2010–2024 (dropping the Great Recession period); (2) using the square of the pre-1970 share to allow for nonlinear relationships; and (3) dropping the five states with the highest dam counts (Texas, Kansas, Oklahoma, Missouri, Iowa). All robustness variants yield null or negative pre-1970 coefficients consistent with the main results.

D. Heterogeneity Appendix

Precipitation increase heterogeneity.. The SDE table (Table 5) reports that the pre-1970 share coefficient is -0.151 ($p = 0.13$) in states where precipitation has increased since the design era, and $+0.082$ ($p = 0.84$) in states where it has decreased. Neither is significant, and the difference between them is also insignificant ($p > 0.5$). The prediction of the engineering-obsolescence hypothesis—that the vintage effect should be most positive in states that have experienced the largest precipitation increases—is not supported.

E. Additional Figures and Tables

No additional figures are included. The five tables in the main text and appendix constitute the complete set of exhibits.

F. Standardized Effect Sizes

Table 5: Standardized Effect Sizes: Dam Vintage and Flood Outcomes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled						
Flood Declaration (LPM)	-0.1583	0.096	0.223	-0.1171	0.0711	Moderate negative
Flood Declaration (Design Gap)	-3.9475	2.3612	0.223	-3.3583	2.0088	Large negative
Flood Declaration Count (Poisson)	-1.7335	0.8771	11.813	-0.0242	0.0122	Small negative
Panel B: Heterogeneous (by precipitation change)						
Flood Decl. (Precip. Increased)	-0.1507	0.0983	0.355	-0.0699	0.0456	Moderate negative
Flood Decl. (Precip. Decreased)	0.0824	0.4014	0.343	0.0396	0.1931	Small positive

Notes: **Country:** United States. **Research question:** Does the share of dams built before 1970—designed with outdated precipitation estimates—predict higher flood damage in a state? **Policy mechanism:** Dam spillways are physical constants set at construction using the hydrological standards of that era (TP-40, 1961); as climate shifts increase precipitation, older dams face design gaps between their engineered capacity and current flood magnitudes. **Outcome definition:** FEMA flood disaster declaration indicator (=1 if any flood declaration in a state-year) and count of flood declarations. **Treatment:** Continuous—share of state's dams completed before 1970 (0.30–0.96). **Data:** National Inventory of Dams (73,293 dams), FEMA Disaster Declarations (11,234 flood records), NOAA nClimDiv precipitation, 2000–2024, state-year panel (1,200 obs). **Method:** OLS and Poisson with year FE, SEs clustered at state level (48 clusters). **Sample:** 48 CONUS states, excluding Alaska and Hawaii. SDE = $\hat{\beta} \times \text{SD}(X)/\text{SD}(Y)$ for continuous treatment, where SD(Y) is computed over the full sample. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).