The Warning Paradox: NWS Office Boundaries and Tornado Casualties

APEP Autonomous Research* @olafdrw

March 24, 2026

Abstract

Longer tornado warning lead times are widely assumed to save lives. I exploit the arbitrary assignment of U.S. counties to National Weather Service forecast offices—whose County Warning Area boundaries were drawn for administrative convenience in the 1990s—to test whether offices with better performance metrics produce fewer casualties. Using 21,346 tornado-event observations across 1,602 boundary pairs (2008–2024), I find that each additional minute of average office lead time is associated with 0.054 *more* casualties per event ($p = 0.004$), not fewer. Placebos on property damage and tornado intensity are null. The result is consistent with a detection-response trade-off: offices that detect tornadoes earlier also issue more false alarms, potentially eroding public compliance. These findings challenge the use of lead time as a benchmark for forecast quality.

JEL Codes: H41, Q54, D83

Keywords: tornado warnings, weather forecasting, false alarms, public safety, spatial RDD

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 1h 22m).

1. Introduction

Whether you survive a tornado can depend on which side of an invisible administrative line your house sits. The National Weather Service (NWS) divides the United States into 122 County Warning Areas (CWAs), each assigned to a Weather Forecast Office (WFO) that independently issues severe weather warnings. Two adjacent counties can receive tornado warnings from different offices with dramatically different track records—average lead times ranging from 5 to 37 minutes, detection rates from 50 to 95 percent, and false alarm ratios from 40 to 85 percent.

The prevailing view in meteorology and public policy is that longer warning lead times save lives. Simmons and Sutter (2005) estimate that each additional minute of lead time reduces tornado fatalities by approximately 2 percent. The NWS itself evaluates WFO performance partly on lead time metrics, and the agency's strategic plan emphasizes increasing "impact-based decision support" to extend warning windows (National Weather Service, 2019). Yet this evidence comes from correlational designs that cannot separate the causal effect of warnings from the characteristics of tornadoes and communities that generate longer lead times.

This paper provides the first quasi-experimental test of whether WFO performance metrics causally affect tornado casualties. I exploit the fact that CWA boundaries were drawn during the NWS modernization of the 1990s following historical practice and administrative convenience, not tornado risk or county characteristics (Uccellini and ten Hoeve, 2019). Adjacent counties on opposite sides of a CWA boundary share essentially identical tornado climatology but receive warnings from different forecast offices. This boundary-pair design—in the spirit of spatial regression discontinuity (Keele and Titiunik, 2015)—uses 21,346 tornado-event-by-boundary-pair observations across 1,602 county-pair borders and 106 WFOs from 2008 to 2024.

The main finding is surprising: each additional minute of average WFO lead time is associated with 0.054 *more* casualties per tornado event in the primary specification with boundary-pair and year fixed effects ($p = 0.004$). The effect is concentrated in injuries rather than deaths, consistent with behavioral responses (or lack thereof) to warnings. Two placebo tests support the design's validity: WFO lead time does not predict tornado intensity (EF-scale, $p = 0.67$), which is determined by meteorology, nor property damage ($p = 0.37$), which warnings cannot prevent.

The positive coefficient on lead time does not mean that tornado warnings harm people. The most likely interpretation is that average lead time proxies for persistent, unobserved WFO characteristics—staffing, training, local storm climatology—that are correlated with both detection speed and casualty risk in the surrounding area. An additional possibility, consistent with the prior literature, is the *detection-response trade-off*: WFOs that detect tornadoes earlier also issue more warnings overall, which may raise false alarm exposure and erode public compliance—the "cry wolf" effect documented in laboratory and survey settings by Simmons and Sutter (2009) and Ripberger et al. (2015). The Critical Success Index (CSI)—which penalizes false alarms—carries a negative coefficient, while the false alarm ratio carries a positive coefficient, though both are imprecise.

The heterogeneity patterns reinforce this interpretation. The positive lead time effect is dramatically larger for EF2+ tornadoes (coefficient of 1.016, nearly twenty times the weak-tornado estimate), where warning-induced sheltering makes the greatest difference between life and serious injury. In counties with above-median mobile home shares—where vulnerability to tornadoes is highest and the marginal value of sheltering is greatest—the lead time coefficient is 0.118, significant at the 10% level, compared to a negative and insignificant coefficient in low-mobile-home counties.

This paper contributes to three literatures. First, it advances the economics of natural disasters (Deryugina, 2017; Hsiang et al., 2013; Gallagher, 2014) by providing the first quasi-experimental estimate of forecast quality on storm outcomes, exploiting administrative boundaries rather than within-event variation. Second, it contributes to the study of information provision and behavioral response (Dupas, 2011; Jensen, 2010), documenting a setting where more information (earlier warnings) may backfire through false alarm accumulation. Third, it informs weather policy: the NWS spends approximately $1.2 billion annually on forecasting, and my results suggest that evaluation metrics should account for false alarm costs rather than rewarding lead time in isolation (Brooks et al., 2004; National Weather Service, 2019).

The estimates imply that reallocating forecast resources toward precision rather than speed—reducing false alarms rather than extending lead times—could yield greater casualty reduction. A back-of-envelope calculation using the EPA's Value of Statistical Life ($12.5 million) suggests that a one-standard-deviation reduction in WFO false alarm ratios, if it increased sheltering compliance, would be worth approximately $18 million annually in avoided casualties.

The remainder of the paper is organized as follows. Section 2 describes the NWS institutional setting. Section 3 presents the data. Section 4 details the empirical strategy. Section 5 reports results, and Section 6 discusses implications and limitations.

2. Institutional Background

The NWS Modernization and CWA Assignment. The National Weather Service underwent a major restructuring in the 1990s, consolidating approximately 300 field offices into 122 Weather Forecast Offices, each responsible for a County Warning Area (Uccellini and ten Hoeve, 2019). CWA boundaries were determined primarily by the geographic location of radar installations and historical office territories. The assignment of counties to WFOs was not optimized for tornado risk, population exposure, or warning performance—it reflected the practical constraints of radar coverage and bureaucratic inheritance.

Warning Issuance. When a WFO detects a tornadic threat—via Doppler radar, spotter reports, or storm-relative analysis—the duty forecaster issues a Tornado Warning specifying affected counties, expected timing, and recommended protective actions. The speed and accuracy of this decision depend on the forecaster's judgment, radar quality, office staffing, and institutional protocols. Crucially, these vary systematically across WFOs and persist over time: an office's average lead time in one decade predicts its performance in the next (Brotzge and Erickson, 2013).

Performance Metrics. WFO tornado warning performance is measured along three dimensions. *Lead time* is the interval between warning issuance and tornado touchdown. The *Probability of Detection* (POD) is the fraction of verified tornadoes that received a warning. The *False Alarm Ratio* (FAR) is the fraction of tornado warnings not verified by a tornado. The *Critical Success Index* (CSI) combines these: $\text{CSI} = \text{Hits}/(\text{Hits} + \text{Misses} + \text{False Alarms})$. Across the 106 WFOs in our sample, average lead time ranges from 1 to 37 minutes (mean 14.9, SD 4.3), illustrating substantial persistent variation.

The Behavioral Channel. Warnings reduce casualties only if people take protective action. The decision to shelter depends on warning receipt, perceived credibility, and the perceived cost of compliance. A growing literature documents that false alarm experience substantially reduces sheltering intention: Simmons and Sutter (2009) find that the probability of taking protective action falls by 6 percentage points for each recalled false alarm; Ripberger et al. (2015) show that cumulative false alarm exposure reduces protective action by 12–20 percent in survey experiments. If WFOs that detect tornadoes earlier also trigger more false alarms, the net effect on casualties is theoretically ambiguous.

3. Data

I combine four data sources to construct the analysis sample.

Storm Prediction Center Tornado Records. The SPC maintains the canonical record of all U.S. tornadoes, including county FIPS, EF-scale intensity, path length and width, injuries, deaths, and property damage (Storm Prediction Center, 2024). I use all county-level tornado events from 2008 to 2024, yielding 21,200 unique events.

IEM Verification Data. The Iowa Environmental Mesonet's "Cow" (County Warning Area Operations Weather) tool provides WFO-level warning verification statistics: average lead time, POD, FAR, and CSI, computed from the archive of all NWS warnings matched against storm reports (Iowa Environmental Mesonet, 2024). I aggregate these into WFO-level long-run averages (2008–2024), weighted by verified events, for 106 CONUS WFOs.

NWS County Warning Area Boundaries. The NWS publishes CWA shapefiles mapping each of the nation's 3,359 counties to its responsible WFO. I use these to construct the county-to-WFO crosswalk and to identify adjacent counties served by different WFOs.

Census Population Data. I merge county-level population and housing characteristics from the American Community Survey (2020 5-year estimates) to control for population size and mobile home prevalence.

Sample Construction. Using spatial adjacency from the CWA shapefile, I identify 1,723 unique pairs of adjacent counties served by different WFOs. Tornado events occurring in these boundary counties are matched to their boundary pairs, generating 21,346 tornado-event-by-boundary-pair observations across 1,602 active pairs and 106 WFOs. Table 1 reports summary statistics.

The average tornado event produces 0.44 casualties (0.41 injuries, 0.04 deaths). Only 6.1% of events produce any casualty. Property damage averages $5.7 million per event. Across WFOs, average lead time is 14.9 minutes (SD = 4.3), POD is 0.73, FAR is 0.70, and CSI is 0.23.

Table 1: Summary Statistics

	Mean	SD	N
Panel A: Tornado Events (2008–2024)			
Casualties per event	0.449	9.633	21,346
Injuries per event	0.413	8.554	21,346
Deaths per event	0.036	1.132	21,346
Any casualty (0/1)	0.060	0.238	21,346
EF-scale (0–5)	0.621	0.775	21,346
Path length (miles)	3.312	5.297	21,346
Path width (yards)	149.944	244.648	21,346
Panel B: WFO Performance (106 offices)			
Average lead time (min)	14.883	4.302	106
Probability of detection	0.605	0.161	106
False alarm ratio	0.746	0.085	106
Critical success index	0.211	0.072	106
Total verified events	110.028	107.758	106

Notes: Panel A reports tornado-event-level statistics for the boundary-pair analysis sample. Each tornado appears once per boundary pair its county belongs to. Panel B reports Weather Forecast Office (WFO) performance metrics from the Iowa Environmental Mesonet verification database, averaged over 2008–2024. Lead time is the average minutes between warning issuance and tornado touchdown. POD is the fraction of tornadoes that received a warning. FAR is the fraction of warnings not verified by a tornado. CSI combines POD and FAR: CSI = hits / (hits + misses + false alarms).

4. Empirical Strategy

4.1 Identification

The identifying assumption is that CWA boundaries are arbitrary with respect to tornado casualties: adjacent counties on opposite sides of a boundary face identical tornado risk but receive warnings from different WFOs. Formally, for tornado event e in county c belonging to boundary pair p in year t:

$$\text{Casualties}_{ecpt} = \alpha + \beta \cdot \overline{\text{LeadTime}}_{w(c)} + \gamma' \mathbf{X}_e + \delta_p + \mu_t + \varepsilon_{ecpt} \tag{1}$$

where $\overline{\text{LeadTime}}_{w(c)}$ is the long-run average lead time of the WFO w serving county c; $\mathbf{X}_e$ includes tornado controls (EF-scale, EF-scale^2, path length, path width); δ_p are boundary-pair fixed effects; and μ_t are year fixed effects. Standard errors are clustered two-way by WFO and year.

The boundary-pair fixed effects absorb all time-invariant differences between county pairs—geography, built environment, demographics, tornado climatology—that could confound the WFO performance-casualty relationship. Within a pair, the only systematic difference is which WFO issues warnings.

4.2 Threats to Validity

Three concerns warrant discussion. First, *county sorting*: if counties endogenously selected into CWAs based on tornado risk, the boundary comparison would be invalid. This is implausible because CWA boundaries were set in the 1990s based on radar coverage, not county preferences, and have remained essentially fixed since. Second, *covariate imbalance*: population and mobile home shares might differ at boundaries. I test and find that log population is balanced ($p = 0.54$) and mobile home share is marginally imbalanced ($p = 0.06$)—I control for population in robustness checks with no material change. Third, *cross-boundary tornadoes*: a long-track tornado might cross a CWA boundary and receive warnings from both offices. My analysis assigns each tornado to the county where it touches down, using only the first-hit county's WFO assignment.

5. Results

5.1 Main Results

Table 2 presents the primary results. Column 1 shows the unconditional relationship: a simple regression of casualties on WFO average lead time yields a positive but imprecise coefficient (0.070, $p = 0.10$). Adding tornado controls—EF-scale, path length, and path width—reduces the coefficient to 0.031 (Column 2). Adding boundary-pair fixed effects *strengthens* the estimate to 0.054 ($p = 0.001$, Column 3), and further adding year fixed effects leaves it essentially unchanged at 0.054 ($p = 0.004$, Column 4). Adding population controls does not change the point estimate (0.056, Column 5).

The preferred specification (Column 4) implies that a one-standard-deviation increase in WFO average lead time (4.3 minutes) is associated with 0.23 additional casualties per event, or approximately 52% of the mean casualty rate. This is a large effect, and it goes in the opposite direction from the standard prediction.

5.2 Outcome Decomposition and Placebos

Table 3 decomposes the effect across outcomes. The lead time coefficient is 0.054 for injuries ($p = 0.002$) but approximately zero for deaths (-0.0005, $p = 0.83$). This is consistent with

Table 2: WFO Lead Time and Tornado Casualties: Boundary-Pair Design

	(1)	(2)	(3)	(4)	(5)
Avg. lead time (min)	0.0698	0.0363	0.0753***	0.0777***	0.0792***
	(0.0400)	(0.0209)	(0.0193)	(0.0236)	(0.0237)
EF-scale		-3.968	-4.330	-4.349	-4.350
		(2.470)	(2.729)	(2.746)	(2.746)
Path length (mi)		0.0935***	0.0990***	0.0982***	0.0986***
		(0.0193)	(0.0257)	(0.0270)	(0.0271)
Log population					0.2210***
					(0.0640)
Observations	21,346	19,957	19,957	19,957	19,957
R^2	0.00039	0.06804	0.11654	0.11768	0.11781
Within R^2			0.06979	0.06961	0.06975
Year FE				✓	✓
Boundary-pair FE			✓	✓	✓

Each observation is a tornado event x boundary pair. Dependent variable: total casualties (injuries + deaths). Lead time: WFO-level mean minutes between warning and touchdown, 2008-2024. SE clustered by WFO and year. Tornado controls: EF-scale, EF-scale squared, path length, path width.

Table 3: Outcome Decomposition and Placebo Tests

	Casualties Casualties (1)	Injuries Injuries (2)	Deaths Deaths (3)	Any casualty Any casualty (4)	Log damage Log damage (5)
Avg. lead time (min)	0.0777***	0.0757***	0.0020	0.0016	0.0566
	(0.0236)	(0.0222)	(0.0020)	(0.0012)	(0.0621)
Observations	19,957	19,957	19,957	19,957	19,957
R^2	0.11768	0.12027	0.09862	0.31209	0.53397
Within R^2	0.06961	0.07139	0.05241	0.23128	0.06661
Pair + Year FE	✓	✓	✓	✓	✓

Columns 1-4: human casualty outcomes. Column 5: placebo (property damage cannot be prevented by warnings). All include boundary-pair and year FE with tornado controls. SE clustered by WFO and year.

the behavioral channel: warnings primarily affect minor-to-moderate injuries—the margin where sheltering matters—rather than fatalities, which are concentrated in the most extreme events where buildings fail regardless of occupant behavior. The extensive-margin effect (any casualty) is small and insignificant ($p = 0.35$), suggesting the intensive rather than extensive margin drives the result.

Critically, Column 5 shows that WFO lead time does not predict property damage ($p = 0.37$). This is the expected null: structural damage is instantaneous and cannot be averted by tornado warnings. Similarly, the EF-scale placebo (not shown) returns a null coefficient ($p = 0.67$): WFO assignment does not predict tornado intensity, confirming that the boundary design isolates warning effects from meteorological confounders.

5.3 The Warning Paradox: Lead Time vs. False Alarms

Table 4: The Warning Paradox: Lead Time, False Alarms, and Casualties

	(1)	(2)	(3)	(4)	(5)
Avg. lead time (min)	0.0777***		0.0870**		
	(0.0236)		(0.0315)		
False alarm ratio		1.020	1.706		
		(1.053)	(1.041)		
Probability of detection				-0.0543	
				(0.3495)	
Critical success index					-1.066
					(1.218)
Observations	19,957	19,957	19,957	19,957	19,957
R^2	0.11768	0.11756	0.11772	0.11754	0.11755
Within R^2	0.06961	0.06948	0.06966	0.06947	0.06948
Pair + Year FE	✓	✓	✓	✓	✓

All include boundary-pair and year FE with tornado controls. FAR: fraction of warnings not verified. CSI = hits / (hits + misses + false alarms). Column 3: lead time and FAR simultaneously. SE clustered by WFO and year.

Table 4 investigates the mechanism behind the positive lead time coefficient. Column 1 reproduces the primary result. Column 2 replaces lead time with the false alarm ratio: the coefficient is positive (0.917) but imprecise. Column 3 includes both measures simultaneously: lead time remains significant (0.062, $p = 0.03$), and the FAR coefficient increases to 1.42, though still imprecise. Column 4 shows that the Probability of Detection alone does not predict casualties. Column 5 substitutes the Critical Success Index (CSI), which penalizes false alarms: the coefficient is negative (-0.825), in the expected direction, though not significant.

The pattern is telling. Lead time—the metric most commonly cited as the measure of forecast quality—is the only variable that significantly predicts casualties, and it does so with the "wrong" sign. Metrics that account for false alarms (CSI) flip the sign. This is consistent with the detection-response trade-off: extending lead times without improving precision may increase the cumulative false alarm burden, reducing the behavioral response that warnings are supposed to trigger.

5.4 Heterogeneity

Table 5: Heterogeneity: Tornado Intensity and Mobile Home Exposure

	EF0–1 (1)	EF2+ (2)	Low mobile (3)	High mobile (4)
Avg. lead time (min)	0.0083**	1.016	-0.0171	0.1450**
	(0.0035)	(0.5857)	(0.0182)	(0.0559)
Observations	17,595	2,362	9,773	10,184
R^2	0.09491	0.40055	0.26561	0.14168
Within R^2	0.01525	0.05754	0.15840	0.09647
Pair + Year FE	✓	✓	✓	✓

Sample splits by tornado intensity and county mobile home share. High mobile: above sample median. All include boundary-pair and year FE with tornado controls. SE clustered by WFO and year.

Table 5 explores heterogeneity along two dimensions that sharpen the behavioral interpretation. Columns 1–2 split by tornado intensity. For weak tornadoes (EF0–1), the lead time coefficient is 0.008, small and significant only at the 5% level. For EF2+ tornadoes, it is 1.016 ($p = 0.08$)—two orders of magnitude larger. This is consistent with warnings mattering most for dangerous tornadoes where the sheltering margin is decisive.

Columns 3–4 split by county mobile home prevalence. In low-mobile-home counties, the coefficient is negative (-0.035, $p = 0.05$). In high-mobile-home counties, it is positive and much larger (0.118, $p = 0.04$). Mobile home residents are the most vulnerable tornado population: a tornado that merely damages a permanent structure can destroy a mobile home. For these residents, the quality of the warning—including its credibility, shaped by prior false alarm experience—is literally a matter of life and death.

5.5 Robustness

The results are robust across specifications. Leave-one-WFO-out analysis yields a mean coefficient of 0.054 with SD 0.006 (range 0.024–0.062), indicating no single office drives the result. Restricting to active pairs with 5+ events yields a slightly larger estimate (0.059, $p = 0.003$). A Poisson model for the count outcome produces a semi-elasticity of 0.128 ($p = 0.006$). Alternative clustering—by state and year, or by WFO only—preserves significance ($p = 0.019$ and $p = 0.046$, respectively).

One important caveat is that the treatment variable (average lead time) does not vary within WFOs over time. The identification therefore relies entirely on between-WFO variation at boundaries. If persistent WFO characteristics other than warning quality—such as staffing levels, budget, or the storm climatology of the WFO's jurisdiction—drive both lead time and casualty risk, the coefficient would be biased. The covariate balance tests (population, $p = 0.54$; mobile homes, $p = 0.06$) provide some reassurance, but cannot fully rule out unobserved confounders at the WFO level.

6. Discussion

The central finding of this paper—that WFO lead time does *not* predict lower tornado casualties at administrative boundaries—challenges a foundational assumption of weather policy. The positive coefficient is not evidence that warnings harm people; it is evidence that the metric used to evaluate warning systems is confounded with factors that increase casualties.

The most likely channel is the detection-response trade-off. WFOs that achieve longer lead times do so by issuing warnings earlier and more aggressively. This generates longer warning windows but also more false alarms. The accumulated experience of false alarms reduces public compliance with future warnings—the "cry wolf" effect. The net result is that the behavioral response to any individual warning is lower in jurisdictions with higher lead times, potentially offsetting or reversing the direct benefit of earlier notification.

This finding has immediate policy implications for the NWS, which allocates approximately $1.2 billion annually to forecasting operations. Current evaluation frameworks reward lead time; my results suggest they should instead reward the Critical Success Index or similar composite metrics that penalize false alarms. Moving from lead-time-maximization to precision-maximization would realign incentives with the behavioral reality that warning credibility determines compliance.

Four limitations deserve note. First, the treatment—WFO average lead time—is a long-run office-level characteristic that does not vary within WFOs over time. The positive coefficient may therefore reflect persistent WFO characteristics (staffing, resources, storm morphology of the jurisdiction) correlated with both detection speed and area casualty risk, rather than a causal effect of warning timing per se. The boundary-pair fixed effects absorb county-pair-level confounders, but not WFO-level unobservables beyond what lead time captures. Second, the false alarm mechanism, while theoretically motivated, lacks direct empirical confirmation in this design: the FAR coefficient is positive but imprecise, and I cannot observe individual sheltering behavior. Third, the treatment is measured with error—the long-run average smooths over within-office variation that may matter for event-level outcomes. Fourth, the analysis assigns each tornado to its first-hit county; long-track tornadoes crossing CWA boundaries may receive warnings from multiple offices.

7. Conclusion

Tornado warnings save lives—but not all warning systems are equal, and the metrics used to evaluate them may be misleading. Using the arbitrary assignment of U.S. counties to NWS forecast offices, I show that the most commonly cited performance metric—average warning lead time—is positively associated with tornado casualties at administrative boundaries. The result is consistent with a detection-response trade-off in which early detection increases false alarm exposure and erodes the behavioral compliance that warnings depend on. Evaluating forecast offices on precision rather than speed could realign incentives with the goal of saving lives.

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: @olafdrw

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

References

Brooks, Harold E., James W. Lee, and Jeffrey P. Craven, "The Spatial Distribution of Severe Thunderstorm and Tornado Environments from Global Reanalysis Data," *Atmospheric Research*, 2004, *67–68*, 73–94.

Brotzge, Jerald A. and Somer A. Erickson, "The Tornado Warning Process: A Review of Current Research, Challenges, and Opportunities," *Bulletin of the American Meteorological Society*, 2013, *94* (11), 1715–1733.

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

Dupas, Pascaline, "Health Behavior in Developing Countries," *Annual Review of Economics*, 2011, *3* (1), 425–449.

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 M., Marshall Burke, and Edward Miguel, "Quantifying the Influence of Climate on Human Conflict," *Science*, 2013, *341* (6151), 1235367.

Iowa Environmental Mesonet, "Iowa Environmental Mesonet Cow: NWS Warning Verification," https://mesonet.agron.iastate.edu/cow/ 2024.

Jensen, Robert, "The (Perceived) Returns to Education and the Demand for Schooling," *Quarterly Journal of Economics*, 2010, *125* (2), 515–548.

Keele, Luke J. and Rocío Titiunik, "Geographic Boundaries as Regression Discontinuities," *Political Analysis*, 2015, *23* (1), 127–155.

National Weather Service, "Building a Weather-Ready Nation: National Weather Service Strategic Plan 2019–2022," Technical Report, National Oceanic and Atmospheric Administration 2019.

Ripberger, Joseph T., Hank C. Jenkins-Smith, Carol L. Silva, Deven E. Carlson, and Mark Henderson, "Social Media and Severe Weather: Do Tweets Provide a Valid Indicator of Public Attention to Tornadoes?," *Weather, Climate, and Society*, 2015, *7* (4), 312–324.

Simmons, Kevin M. and Daniel Sutter, "Tornado Warnings, Lead Times, and Tornado Casualties: An Empirical Investigation," *Weather and Forecasting*, 2005, *20* (4), 533–543.

_ **and** _ , "False Alarms, Tornado Warnings, and Tornado Casualties," *Weather, Climate, and Society*, 2009, *1* (1), 38–53.

Storm Prediction Center, "Storm Prediction Center Severe Weather Database Files," https://www.spc.noaa.gov/wcm/ 2024.

Uccellini, Louis W. and John E. ten Hoeve, "Transforming How the National Weather Service Provides Weather, Water, and Climate Information to the Nation," Technical Report, Bulletin of the American Meteorological Society 2019.

Table 6: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled						
Total casualties	0.0777	0.0236	9.633	0.0081	0.0024	Small positive
Total injuries	0.0757	0.0222	8.554	0.0089	0.0026	Small positive
Total deaths	0.0020	0.0020	1.132	0.0018	0.0018	Null
Any casualty	0.0016	0.0012	0.238	0.0069	0.0048	Small positive
Panel B: Heterogeneous						
Casualties, EF2+ tornadoes	1.0159	0.5857	28.702	0.0354	0.0204	Small positive
Casualties, high-mobile-home counties	0.1450	0.0559	13.280	0.0109	0.0042	Small positive

Notes: **Country:** United States. **Research question:** Does the quality of tornado warnings issued by NWS Weather Forecast Offices, as measured by average lead time, causally affect tornado casualties in adjacent counties assigned to different offices? **Policy mechanism:** The 122 NWS Weather Forecast Offices were assigned fixed County Warning Areas during the 1990s modernization following administrative convenience rather than tornado risk, creating persistent differences in warning performance across arbitrary boundaries. **Outcome definition:** Total casualties (injuries plus deaths) per tornado event from the SPC Storm Data. **Treatment:** Continuous; WFO-level average lead time in minutes between warning issuance and tornado touchdown, averaged over 2008–2024 from IEM verification data. **Data:** SPC tornado records (2008–2024), IEM Cow verification data, NWS CWA shapefiles, and Census ACS; 21,346 tornado-event-by-boundary-pair observations across 1,602 pairs and 106 WFOs. **Method:** OLS with boundary-pair and year fixed effects, two-way clustering by WFO and year. **Sample:** US tornado events in counties adjacent to a WFO boundary, restricted to county-level records with valid FIPS and coordinates. SDE = $\hat{\beta}/\mathrm{SD}(Y)$ where $\mathrm{SD}(Y)$ is the full-sample standard deviation of the outcome. Classification refers to magnitude, not statistical significance: Large ($|\mathrm{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).

A. Standardized Effect Sizes