

Do Small Releases Warn of Serious Chemical Accidents?

Testing Emission Events as a Leading Indicator at Texas Plants

Daniel Okwor

Orygn LLC

daniel@orygn.tech

ORCID: 0009-0005-2546-181X

Preprint, October 2026

DOI: [10.5281/zenodo.23348378](https://doi.org/10.5281/zenodo.23348378)

Abstract

Process safety standards treat small releases as leading indicators of serious accidents, but the link has rarely been tested with records kept outside the companies themselves. Texas publishes every report of an emission event, an unauthorized release of air pollution above a reportable quantity that a plant must report within 24 hours, and the U.S. Environmental Protection Agency’s Risk Management Program (RMP) records serious chemical accidents at covered plants. We collected all 103,614 Texas event reports filed from 2003 to 2026, linked those from RMP plants to the same plants’ accident records, and asked, in a preregistered test, whether a plant’s unplanned emission events become more frequent in the three months before a serious accident, stopping one week before it. We found no sign that they do. Across 545 accidents at 216 plants from 2003 to 2021, comparing each accident day with the same plant’s other days in the same year, the registered estimate was 0.88 times the accident risk per doubling of recent events (95% confidence interval 0.75 to 1.02). After seeing the registered result, we found that it is biased downward. An accident generates its own event reports, 22 times the plant’s usual daily rate on the accident day, and those reports fall into the look-back counts of the comparison days that follow. Removing each plant’s reports from the accident day and the two days after gives 1.05 (0.85 to 1.29). Rises larger than about 30% per doubling are therefore unlikely, though smaller ones cannot be excluded. In a simulation, the bias hid a real 30% rise and the correction recovered it. Serious incidents reported to the National Response Center, OSHA injury records, events that release the plant’s own regulated chemicals, and satellite heat detections at the plant also show no clear rise. Over longer spans there are hints of a link. Across a plant’s whole history, periods with more events have modestly more accidents (1.24 per doubling, an exploratory estimate), and between plants, more events in one year go with slightly more accidents the next (1.12, 0.99 to 1.26), although events add little to prediction beyond a plant’s size and its past accidents. One exploratory lead, found after the main results were seen, is that accidents were about 1.6 times as likely in the week after a plant reported planned work, mainly startups and shutdowns. A first further test was inconclusive. The registered plan, the code and the derived data are public.

1 Introduction

A common idea in process safety is that small failures come before big ones, so counting the small ones should give warning of the big ones. The idea goes back at least to Heinrich’s accident pyramid [1], and current standards still rely on it. The American Petroleum Institute’s standard for process safety indicators, RP 754, sorts loss-of-containment events into tiers by their consequences and states that “it is believed that a similar predictive relationship exists between lower and higher consequence events that relate to process safety” [2]. The Center for Chemical Process Safety

describes its second tier of events as potential precursors that “act as a leading indicator” for the first [3]. Community groups make the same argument with public data. A 2026 report on Harris County, Texas, calls the frequency of reported emission events “another key metric that helps identify facilities at high risk of accidents, fires, explosions, malfunctions, shutdowns, and other incidents” [4].

Despite its wide use, the idea has rarely been tested. A review of the process safety indicator literature found “little published empirical research” on the subject [5]. There are two practical obstacles. The smaller events in RP 754, Tiers 3 and 4, are “intended for internal use at individual sites” [2], so outsiders cannot see them. Within one site, serious accidents are also too rare for a count to show much. An analysis by RAND researchers for the U.S. Chemical Safety Board found that lagging process safety indicators have low statistical power and are unlikely to show significant changes at a single refinery [6, 7]. A real test needs records of minor events and of serious accidents for many plants, kept by someone other than the companies, with the plants identified in both. Such data are unusual; the European Union’s major accident database, for example, does not disclose site names or locations [8].

Texas is one place where such records exist. Since 2003 the Texas Commission on Environmental Quality (TCEQ) has published every report of an emission event, an unauthorized release of air pollution above a reportable quantity, which a plant must report within 24 hours of discovering it [9, 10]. Separately, plants that hold large amounts of listed toxic or flammable chemicals file Risk Management Plans (RMPs) with the U.S. Environmental Protection Agency (EPA), and each plan must include a five-year history of accidental releases that caused deaths, injuries or other serious harm [11]. Both sets of records identify the site. Linking them lets us ask a direct question: does a plant report more emission events in the months before it has a serious accident?

We answer it with a test whose plan was written, frozen and registered before any accident data were compared with events. For each accident, we compare the plant’s recent emission events with its events at other times in the same year, so that the plant’s size, equipment and reporting habits cancel out. Our results:

- **No sign of a rise before accidents.** Across 545 accidents at 216 plants from 2003 to 2021, unplanned emission events in the three months before an accident (stopping a week before it) were not more frequent than on the plant’s other days of the same year. Our corrected estimate, adopted after the registered result was seen, is 1.05 times the accident risk per doubling of recent events (95% confidence interval 0.85 to 1.29). Rises larger than about 30% per doubling are unlikely, and smaller ones cannot be excluded (Section 5).
- **Accidents report themselves.** An accident generates its own event reports, about 22 times the plant’s usual daily rate on the accident day. In the standard design those reports make the weeks after an accident look like accident-free days with many events, which pulls estimates below 1 and can hide a real effect. A correction, checked by simulation, recovers it (Section 6). The same problem applies to any study that tests self-reported events as warning signs of accidents that themselves generate reports.
- **Hints of a long-run link, no short-term warning.** Across a plant’s whole history, periods with more events have modestly more accidents, and between plants, more events in one year go with slightly more accidents the next. Both estimates are uncertain, and events add little to prediction beyond a plant’s size and its past accidents (Section 7).
- **Other records agree.** Serious incidents reported to the National Response Center, OSHA injury records, events that release the plant’s own regulated chemicals, and satellite heat detections at the plant show no clear rise. Of 1,152 variations of the main test, none shows a rise

outside one narrow corner (Section 8).

- **One lead, found after looking.** Following up that corner, we found that accidents were about 1.6 times as likely in the week after a plant reported planned work, mainly startups and shutdowns. This is exploratory, and a first further test was inconclusive (Section 9).
- **A public, linked dataset.** The code that builds the linked event and accident tables from public sources, every derived table and the full deviation log are released (Section 13).

Section 10 says what this means for plant process safety teams, regulators, people who live near these plants, and researchers.

2 Background and related work

Leading indicators. RP 754 counts process safety events in four tiers, from Tier 1 releases with serious consequences down to Tier 4 measures of operating discipline [2]. Hopkins argued that injury rates do not measure process safety and that, because catastrophic accidents are too rare to count at any one site, loss of containment events may be frequent enough to serve instead [12]. The independent panel that reviewed one company’s U.S. refineries after a 2005 explosion found that the company’s personal injury rates “were not predictive of process safety performance” [13]. The CCPS guide also cautions that once a near-miss program starts, reports of small events can rise while serious incidents fall. A raw count therefore need not rise before accidents even when the small events are related to them [3].

Tests of the pyramid. The closest tests come from occupational safety. Using data from more than 27,000 U.S. mines, Yorio and Moore found that lower-severity incidents in one year were associated with fatalities the next, although after adjusting for mine size only two incident types remained significant [14]. Marshall and colleagues found, across more than 50,000 Chilean firms, that minor accidents were a useful signal for forecasting a firm’s overall safety performance [15]. Elliott and colleagues linked OSHA injury rates to RMP accident histories and found “only weak evidence” that everyday safety performance prevents rare major accidents [16]. These studies compare firms or years. None asks whether minor events rise in the weeks before a major accident at the same facility.

What predicts RMP accidents. Analyses of RMP accident histories found facility-specific factors associated with accident rates [17], chiefly a facility’s hazard, size and location, and that it is “not the small facilities per se that are the primary sources of accidents” [18]. A small share of facilities have repeat accidents: 2.2% of the facilities filing in 1999 and 2000 reported more than one accident in five years [18]. EPA has called a past accident “one of the best predictors of future accidents” [19], and its own analysis found that of 801 facilities with an accident in 2004 to 2013, 279 had another within 10 years [20]. For pipelines, a federal analysis found incident history about as useful as infrastructure or operator performance in predicting an operator’s incidents the next year [21]. All of this work is between facilities or from one year to the next.

Texas emission event data. TCEQ’s event reports have been used to study ozone formation in Houston [22], the size and sources of excess emissions [23], and their health effects, including deaths among older residents [24]. Watchdog groups use them to track pollution and enforcement [25, 4]. We found no study that links them to accident records.

When the outcome changes later exposure. Comparing each accident day with other days of the same plant in the same year is a case-crossover design [26] with time-stratified comparison days [27], and a conditional Poisson model with one stratum per plant-year gives identical estimates [28]. Using comparison days after the event is valid only when the event cannot change later exposure, as with air pollution and heart attacks [29, 30]. Studies of vaccine safety state the same condition: “the occurrence of an event must not alter the probability of subsequent exposure” [31], and special methods exist for when it does [32]. Econometrics calls the condition strict exogeneity [33]. An industrial accident breaks it, because the accident itself generates emission event reports. Section 6 shows how large the resulting bias is and how to correct it.

3 Data

3.1 Emission events

TCEQ’s Air Emission Event Report Database holds every report filed since January 2003 [9]. A plant must notify TCEQ within 24 hours of discovering an emission event that releases more than a reportable quantity, and must keep records of smaller ones [10]. Plants also report planned maintenance, startup and shutdown work that will release more than their permits allow, usually before the work begins. We retrieved every report page on October 2 and 3, 2026, kept a byte-for-byte copy with a SHA-256 fingerprint of each page, and confirmed that the list of reports was complete three ways (by start date, by end date, and by region with no dates). The database’s search accepts dates from January 31, 2003. It held 103,614 reports: 71,027 unplanned emission events, 12,812 maintenance reports, 7,650 startup reports, 4,885 shutdown reports and 7,240 excess opacity reports. Of all reports, 53,230 (51%) come from plants in the RMP program. Reports per year fell from about 7,500 in 2004 to about 2,000 in 2024, and planned work reports fell steeply after 2010.

The exposure in the main test is the count of unplanned emission events. Counting rules were fixed in advance. Reports from the same site with the same start date, start time and event type count as one occurrence, as when one outage trips several units and each gets its own report (4,070 extra reports, 4% of the total). Reports that say they duplicate another report are not counted. Reports later found to be below the reportable amount (about 4.5%) or marked as filed in error (0.8%) are counted, because an upset still happened and small upsets are exactly the warning sign under study. A robustness check drops them.

Do the reports record real upsets? If the reports were mostly paperwork, a null result would mean little. Before any accident data were opened, we checked the reports against other sources. At sites with no other site within 1 km, satellites detected heat on 14.9% of nights during reported flaring events and on 2.9% of ordinary nights. Comparing each site with itself, the odds of a detection were 6.6 times higher on event nights (95% interval 4.0 to 10.5, 128 sites) [34]. Downwind air monitors recorded sulfur dioxide at 1.9 times its usual level during reported events (1.6 to 2.2) and 1,3-butadiene at 3.0 times (2.0 to 7.5), with no rise in a placebo comparison a year away. Nitrogen oxides were only slightly above their placebo (1.17 against 1.12), and benzene and hydrogen sulfide showed weak or no signal. Refinery fence-line benzene showed no association with reported events. The event reports also ranked plant-years in nearly the same order as the annual emissions inventory the plants file separately (rank correlation 0.78), although both are company reports. The independent checks, the satellite and the monitors, show that the reports mostly record real releases.

3.2 Serious accidents

Under the RMP rule, a plant’s five-year accident history must include every accidental release from a covered process that caused deaths, injuries or significant property damage on site, or known deaths, injuries, evacuations, sheltering in place, property damage or environmental damage off site [11]. We used the RMP database as released by the Data Liberation Project, current to December 31, 2025 [35]. It lists 2,408 Texas facilities and 1,020 Texas accidents after merging records of the same accident filed in different submissions, 651 of them on or after January 31, 2003. A written rule merged 3 further pairs out of 20 candidate duplicates. Because plants report accidents in filings up to five years later, the main window for accidents ends on December 31, 2021, a choice made before the plan was registered. Results to 2023 are shown alongside. It starts on June 1, 2003, so that every day in the test has a full 90-day look-back inside the event data.

The primary outcome is any RMP accident. Three narrower outcomes were registered as secondary tests: fires or explosions; harm to people (any death or injury); and harm outside the fence (off-site deaths, hospital or medical treatment, evacuation, sheltering in place, property damage, or environmental damage).

3.3 Linking plants

A plant is a connected group of TCEQ site numbers and RMP facility identifiers that refer to the same place. Links came from EPA’s Facility Registry Service (1,373 links) and from identifiers that plants listed in their own RMP filings (296), accepted by rule [36]. We reviewed 375 further candidate pairs suggested by location and name by hand, on a page that showed only names, addresses and identifiers and never any event or accident data. The review added 106 links, and a separate check of each plant’s member sites removed 33 links that were wrong. The final table, frozen before any outcome analysis, has 1,140 plants that join RMP and TCEQ records, 596 of them with event reports. The main test uses the 545 accidents at linked plants between June 1, 2003 and December 31, 2021, at 216 plants. Another 44 accidents in that window are at plants with no TCEQ link, mostly water treatment, food and cold-storage sites that rarely file event reports.

3.4 Other records

Four other sources serve as checks (Table 1). Serious reports to the National Response Center (fires, injuries, deaths or evacuations at fixed facilities) were linked to TCEQ sites by hand, blind to all event and accident data: of 604 reports with a candidate site, 262 at 114 sites were linked and 330 answered “can’t tell” were left out [37]. OSHA fatality and catastrophe investigations and severe injury reports were linked the same way: of 959 records with a candidate, 320 at 104 plants were linked and 620 were left out as “can’t tell” [38]. Night-time heat detections from the VIIRS instrument on the Suomi NPP satellite give a measure of flaring that plants do not report themselves [34]. Federal inspection, violation and enforcement records and EPA emissions inventories serve as covariates, and storm, disaster and hurricane records serve in robustness checks.

4 Method

4.1 The comparison

For every day a plant was at risk, we counted the unplanned emission events that started in the 90 days ending one week before that day (days 97 to 8 before). We then compared accident days with the other days of the same plant in the same calendar year. Comparing a plant with itself

Table 1: Data sources. All are public; snapshot dates and SHA-256 hashes of every file are in the repository.

Source	Records	Years	Use
TCEQ Air Emission Event Report Database	103,614 reports, 71,027 unplanned	2003 to 2026	Exposure
EPA Risk Management Program	2,408 facilities, 1,020 accidents	1999 to 2025	Outcome, plant size and sector
EPA Facility Registry Service	Texas file	2026	Linking plants
National Response Center	262 serious reports at 114 sites	2003 to 2023	Independent outcome
OSHA investigations and severe injury reports	320 records at 104 plants	2003 to 2023	Independent outcome
NASA FIRMS, VIIRS (Suomi NPP)	Night-time heat detections	2012 to 2021	Check on reports; independent exposure
EPA ECHO and emissions inventories	Inspections, violations, enforcement, emissions	2003 to 2021	Covariates
NOAA, FEMA, National Hurricane Center	Storms, disasters, hurricane tracks	2003 to 2021	Robustness checks

within one year means that its size, its equipment, its habits in reporting and any slow change over the years cancel out. The one-week gap keeps the accident’s own reports, and the upset that leads straight into it, out of the count, so that an accident cannot count as its own warning. It also means the test does not look at the last week before an accident.

4.2 The model

The model is a conditional Poisson regression with one stratum for each plant-year:

$$\log E[y_{pd}] = \alpha_{p,\text{year}(d)} + \gamma_{\text{month}(d)} + \beta \log(1 + E_{pd}) + \delta \log(1 + S_{pd}),$$

where y_{pd} is the number of accidents at plant p on day d , E_{pd} is the plant’s count of recent events, S_{pd} is the count of events at every other plant in Texas in the same window, and γ allows for the season. A stratum is a group of days that are compared only with each other, here the days of one plant in one year, so each accident is compared only with the days around it. The statewide count adjusts for things that changed for every plant at once, such as storms, freezes, changes in reporting rules and swings in the economy. This model gives the same estimates as the time-stratified case-crossover design [28]. We report 2^β , the rate ratio per doubling of $(1 + E)$: a value of 1.2 means that when a plant’s count of recent events doubles, its daily risk of an accident is 20% higher. Because the count enters as $1 + E$, going from 1 event to 3, or from 3 to 7, counts as one doubling. Confidence intervals come from standard errors clustered by plant, which allow a plant’s days to be related to one another. A shuffle test moved each accident to a random day of the same plant-year 10,000 times. The model was fitted with PyFixest [39], and the main result was recomputed with a separately written NumPy program (the two coefficients differ by 7×10^{-11}) and with statsmodels.

4.3 Preregistration, and what came after

The analysis plan was frozen on October 2, 2026 and registered on the Open Science Framework on October 3 (<https://osf.io/nf426>), before any accident data were compared with events [40].

Table 2: Main result: rate ratio of a serious accident per doubling of a plant’s unplanned emission events in days 97 to 8 before, comparing days within the same plant-year. “Own reports removed” removes each plant’s event reports from the accident day and the two days after; every day stays in the comparison. This correction was adopted after the registered result was seen. All rows end in 2021 unless stated; the OSHA row starts in 2015, when severe injury reports began.

Version	Cases (plants)	As registered	Own reports removed
RMP accidents, 2003 to 2021 (primary)	545 (216)	0.88 (0.75 to 1.02)	1.05 (0.85 to 1.29)
RMP accidents, 2003 to 2023	571 (223)	0.87 (0.75 to 1.01)	
removing the accident day only	545 (216)		1.03 (0.84 to 1.26)
removing the accident day and 7 days after	545 (216)		1.08 (0.88 to 1.33)
National Response Center serious reports	218 days (105 sites)	0.82 (0.66 to 1.01)	1.08 (0.88 to 1.33)
OSHA deaths and severe injuries	210 days (81 plants)	1.10 (0.79 to 1.54)	1.13 (0.80 to 1.60)

The registered plan fixed the outcome, exposure, window, gap and model, 14 secondary tests with a Holm correction [41], 26 robustness checks, a negative-control test, the full set of 1,152 combinations of reasonable choices (a multiverse [42]), tests between plants and on later years, and a replication with National Response Center records. Additions made before any outcome analysis brought the secondary tests to 15 and the robustness checks to 30 (Appendix B). A power check on simulated accident dates, run before the real test, estimated that the main test would detect a 20% rise per doubling about half the time and a 30% rise about 8 times in 10. That check was too hopeful. Its simulated accidents generated no reports of their own, and once they do, the registered design has little power (Section 6). The corrected intervals are also about a third wider than the registered one.

Every change after registration is logged with its reason in the plan’s deviation log, summarized in Appendix B. Analyses added after the main result was seen are labelled as such throughout. The most important are the correction for the accident’s own reports (Section 6), the equivalence bounds, the simulation, the tests with the plant’s own chemicals and with satellite heat, the time-scale rerun (Section 7) and the analysis of planned work (Section 9). Each is recorded in the deviation log, and for the later checks (D14 to D16) the rules were written into the plan before they were run.

5 No sign of a rise before accidents

Table 2 gives the main result in its registered form and in the versions that correct for the accident’s own reports (Section 6). As registered, the rate ratio per doubling of recent events was 0.88 (95% interval 0.75 to 1.02, $p = 0.10$; shuffle test $p = 0.12$). With accidents to 2023 it was 0.87. Neither is a rise. Removing each plant’s reports from the accident day and the two days after, the correction we prefer, gives 1.05 (0.85 to 1.29).

Figure 1 shows the same thing without a model. In each month before an accident, the plant’s events run at about the level seen around random days in the same plant-years. Single months differ by chance, and in the last six months the accident curve is, if anything, slightly lower. The only jump is in the month that starts on the accident day, about 1.9 times the plant’s average against 1.3 around random days. That jump is the accident and its aftermath, and the one-week

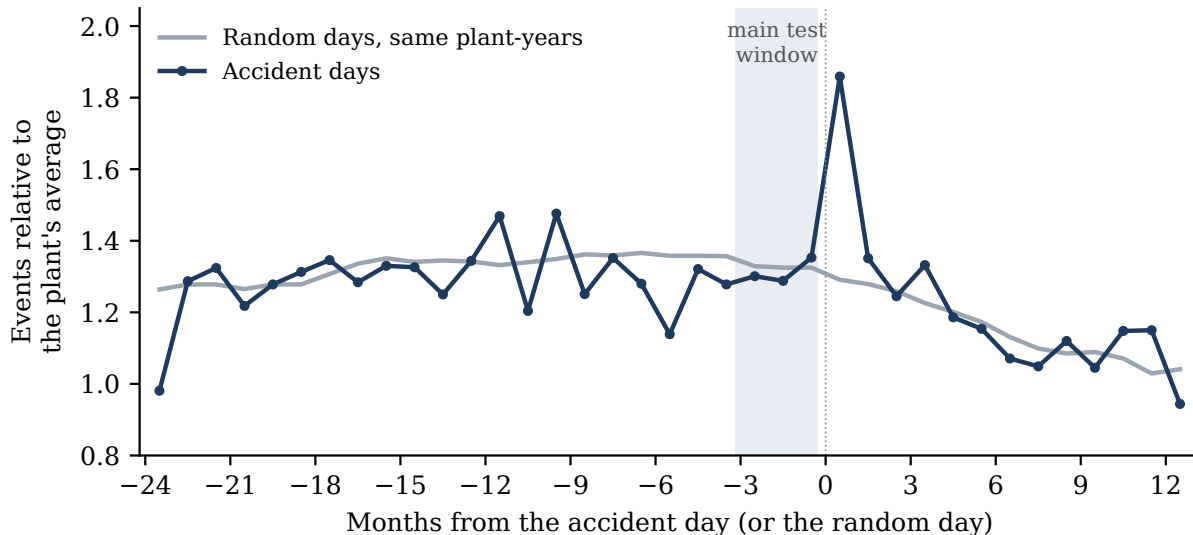


Figure 1: Unplanned emission events in each month around accidents (dark line), against random days in the same plant-years (100 draws, grey line). Each value is the number of events divided by the number expected from the plant’s average daily rate over its study period. The 472 accidents shown are those at plants that reported at least one event in their study period. The shaded band is the window the main test counts (days 97 to 8 before). Both curves sit above 1 because years with accidents are busier than a plant’s average year.

gap keeps it out of the test.

How large a rise is ruled out. The plan called for a null result to be reported as the upper end of its confidence interval, the largest effect the data still allow. A first version of our report used the registered interval and stated a rise of under 2.4% per doubling. We withdrew it when the bias described in Section 6 was found, because that bias pulls the registered estimate down. The corrected 95% interval reaches 1.29, so rises larger than about 30% per doubling are unlikely. A formal equivalence test, two one-sided tests at the 5% level [43], uses the 90% interval and rejects rises larger than 25% (22% to 29% across the corrected versions in Table 2, including the National Response Center). In absolute terms, linked plants have 33 serious accidents per 1,000 plant-years, and a 25% rise would take that to about 42. Smaller rises cannot be ruled out.

Secondary tests. None of the 15 secondary tests points toward a warning sign. Fires and explosions gave 1.01 (0.79 to 1.28) and harm to people 1.06 (0.81 to 1.38). Counting events by the date they were reported, so that only what a regulator could have seen at the time is used, gave 0.86 (0.74 to 1.01). Pounds released gave 0.99 (0.97 to 1.02) and repeat events at the same emission point 0.91 (0.78 to 1.07). Five tests were significant after the Holm correction, and all five point the other way: harm outside the fence (0.56), the 180-day and 365-day windows (0.68 and 0.47), and events 3 to 6 and 6 to 12 months before (0.71 and 0.67). The next section explains why. With the accident’s own reports removed, harm outside the fence gives 0.74 (0.51 to 1.07), and none of the four lag bins is significant. Two more registered exposures were run late, after we noticed they had been missed (plan D16), and are also null: the total hours of recent events, 0.97 (0.91 to 1.03), or 1.03 (0.96 to 1.10) with the accident’s own reports removed; and planned maintenance, startup and shutdown reports counted over the same 90 days, 0.90 (0.74 to 1.10).

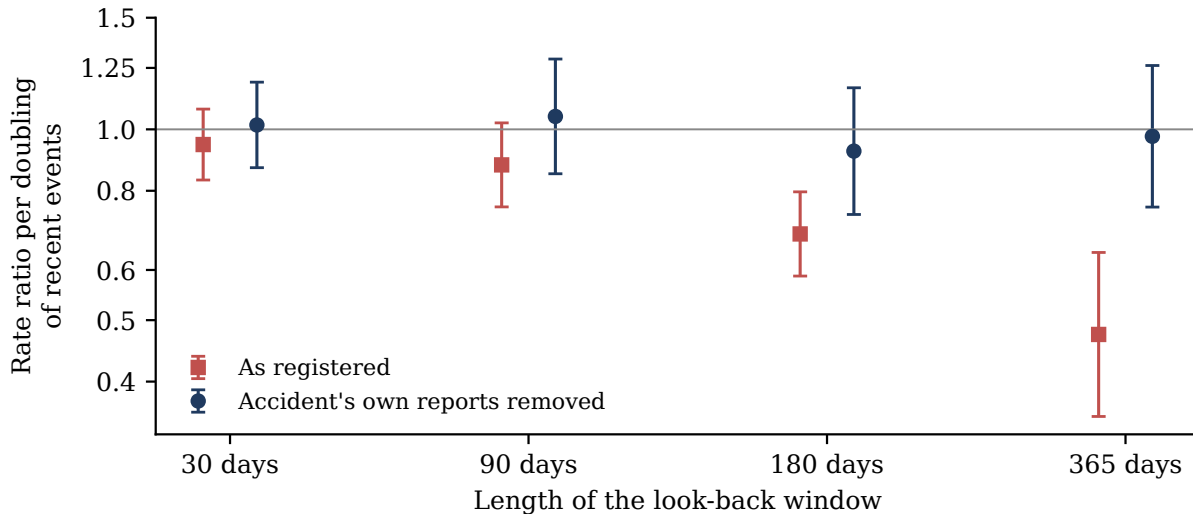


Figure 2: The main test by window length, as registered (squares) and with each plant’s reports from the accident day and the two days after removed (circles), with 95% intervals. As registered, longer windows reach more of the weeks after accidents and give lower estimates. With the accident’s own reports removed, every window is close to 1.

6 Accidents report themselves

On the 541 days with an RMP accident at a linked plant, the plants filed 212 unplanned event reports, against 9.7 expected from their average daily rates: about 22 times as many. On the two days after, the excess is small (26 reports against 19.4 expected). In all, 43% of RMP accidents have a TCEQ report at the same plant within two days. That is expected, since a fire or a large release is itself an emission event, but it changes what the main comparison measures.

Suppose a plant has an accident on March 1 and files three event reports that day. With a 90-day window and a one-week gap, those three reports fall in the look-back count of every day from March 9 to June 6. Those days are comparison days in the same plant-year. They have more recent events and no accident, and the model reads them as evidence that more events go with less risk. A longer window spreads the reports over more comparison days, so the pull grows with the window: the registered estimates are 0.95 at 30 days, 0.88 at 90, 0.68 at 180 and 0.47 at 365 (Figure 2). The pre-run power check could not reveal this, because its simulated accidents generated no reports of their own.

The correction removes each plant’s reports from the accident day and the two days after (269 of the 34,371 events at linked plants, 0.8%) and keeps every day in the comparison. Only earlier accidents remove events from a day’s look-back count, so no information from the future enters. With it, the estimates are 1.02, 1.05, 0.92 and 0.98 for the four windows (Figure 2). Removal of this kind can itself shift the estimate, because it lowers counts only after accidents. To measure that, we applied the same removal around simulated accident dates, where no real accident is involved: the median estimate for the 90-day window is 1.02 when only the accident day is removed, 1.03 when the two days after are removed as well, and 1.06 when seven days after are removed. The corrected estimates in Table 2 rise in the same way with the number of days removed (1.03, 1.05 and 1.08), so that small rise comes mostly from the removal itself.

Table 3: Simulation of the bias: median rate ratio per doubling over 20 simulations per scenario, with real event counts and simulated accidents that generate their own reports at the observed rate.

True effect	90-day window		365-day window	
	Ordinary design	Corrected	Ordinary design	Corrected
None (1.0)	0.84	1.03	0.56	1.08
A real rise of 30% per doubling (1.3)	1.02	1.28	0.63	1.21

A simulation. To check whether the bias is large enough to matter and whether the correction fixes it, we used the real event counts and plant-years, placed accidents on simulated days, and let each simulated accident add event reports at the plant at the excess seen in the real data (0.37 extra reports on the accident day and 0.006 on each of the next two days). In one scenario accidents ignore events. In the other, accident risk truly rises 30% per doubling of recent events. Table 3 gives the median of 20 simulations per scenario. The ordinary design hides nearly all of a real 30% rise at 90 days (1.02) and turns it into an apparent protective effect at 365 days (0.63). The correction recovers it (1.28 and 1.21). With no real effect, the ordinary design gives 0.84 and 0.56, consistent with the 0.88 and 0.47 seen in the real data. With no effect the correction leans slightly upward (1.03 and 1.08), and with a real effect it came out close to the truth. Twenty simulations cannot resolve differences of a few percent.

Fixes that do not work. Two other ways of handling the days after an accident are biased upward. Dropping those days from the comparison altogether (robustness check R16, 1.15) overstates a real effect: in the power simulation, a planted 1.3 came out as 1.39. An earlier version of the removal took out reports from the days after an accident but kept the accident day’s own reports. Its estimates kept climbing as more days were removed, past 1.25, and we stopped it and do not use it.

Why this matters beyond Texas. The comparison of each event day with other days of the same unit and year is standard and is safe for exposures the event cannot affect, such as the weather [29, 30]. It is not safe for self-reported events when the accident itself produces reports. Plants that test their own leading indicators against their own incidents, and researchers who use near-miss logs or regulatory reports, face the same risk: if incidents generate indicator reports, the data can seem to say that more warnings mean less danger. Before reading such a result, count the indicator reports that fall on and just after the incident days.

7 Hints of a long-run link, no short-term warning

The main test asks a short-term question. Widening the comparison to longer spans asks whether a plant’s busier periods are also its more dangerous ones. Figure 3 shows the answer as the comparison widens from the same year to the plant’s whole history. As registered, the estimate rose from 0.88 within a year to 1.32 (1.14 to 1.53) across the whole history. Part of that rise came from the bias of Section 6, which pulls narrow comparisons down more than wide ones. In an exploratory rerun with the accident’s own reports removed (plan D15), the estimates are 1.05 within the same year, 1.07 within the same 2 years, 1.10 within 3, 1.11 within 5, and 1.24 (1.07 to 1.44) across the whole history, the only one clearly above 1. Two cautions apply. No test compares the whole-history and same-year estimates directly, and their intervals overlap. And a comparison across a plant’s whole

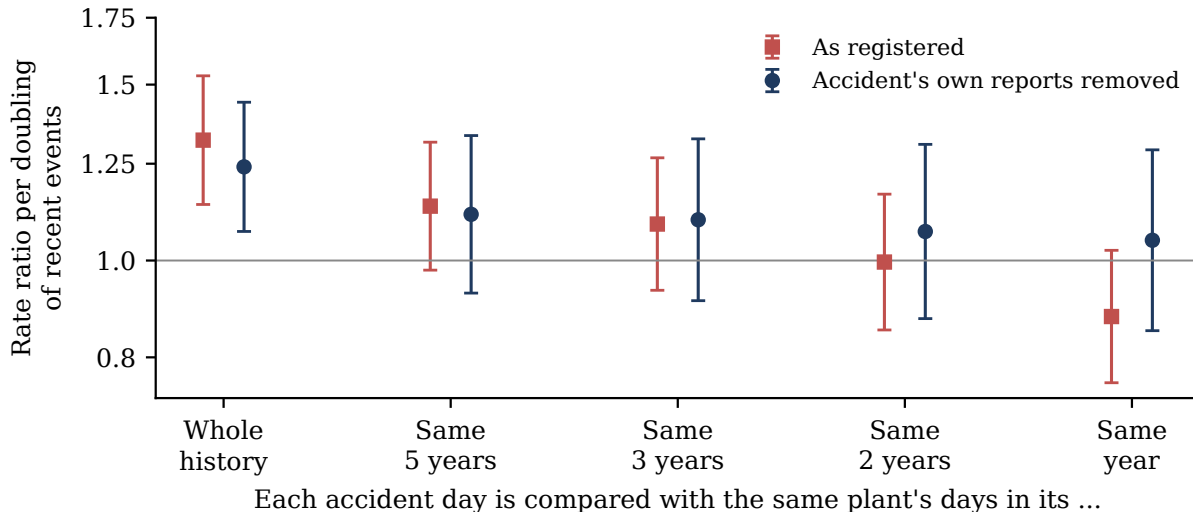


Figure 3: The association between recent events and accidents as the comparison widens, from the same year (the main test) to the plant’s whole history, as registered (squares) and with the accident’s own reports removed (circles), with 95% intervals.

history does not remove slow changes within the plant, such as growth or aging, which the main design was built to remove. The association may reflect such changes rather than anything the events say about risk.

Between plants. The registered between-plant test followed 1,065 plants over 15,506 plant-years with 525 accidents. It related accidents in a year to unplanned events in the year before, adjusting for staff numbers, the largest quantity of a regulated chemical, the number of processes, emissions, sector, accidents in the previous five years and the year. The result was 1.12 per doubling (0.99 to 1.26, $p = 0.08$). Adding federal inspections, violations and enforcement gave 1.15 (1.02 to 1.28, $p = 0.02$). Some of this is accident aftermath. Leaving out the events reported in the 30 days after each accident lowers the second estimate to 1.09 (0.97 to 1.22, $p = 0.15$), but leaving out a month of events at a random time already lowers it to about 1.12, so aftermath explains only part of the drop. In the same model, accidents in the previous five years were by far the strongest predictor (2.34 per doubling), followed by staff numbers (1.23). Inspections, violations and enforcement showed no clear association with accidents. In the accident records themselves, about one accident in five (107 of 545) came within 120 days after an earlier accident at the same plant.

Prediction in later years. A model fitted on 2004 to 2015 and tested unchanged on 2016 to 2021 (20,214 plant-quarters, 135 with an accident) ranked plant-quarters slightly better with events than without: the area under the ROC curve, the chance that a random accident quarter is ranked above a random quarter without one, rose from 0.798 to 0.808, a gain of 0.010 (0.003 to 0.018). The area under the precision-recall curve barely moved (0.047 to 0.049). Both versions put 52% to 53% of the accident quarters in their top 10%, and both were well calibrated: their predicted numbers of accidents matched the observed ones (calibration slopes 0.97 and 0.98, where 1 is ideal). About 18% of test quarters used a plant covariate from a filing received after the quarter began. Refitting with strictly earlier filings gives 0.800 and 0.810. Table 4 shows what simple flags would mean in practice. A flag for any reported event in the previous 90 days catches 42% of accident quarters,

Table 4: Simple flags and accidents, per 1,000 plant-quarters, linked plants, 2004 to 2021. Rates are not adjusted for plant size, so they largely reflect that large plants have more of both events and accidents.

Flag	Accident quarters per 1,000 flagged	Per 1,000 not flagged	Share of accident quarters flagged	False alarms per true alarm
Any event in the prior 90 days	25.6	5.0	42%	38
3 or more events	37.8	6.0	25%	25
10 or more events	46.0	7.2	5%	21

but for every quarter it flags correctly, 38 flagged quarters have no accident.

Events therefore say a little about which plants carry more risk, though less than a plant’s size and past accidents already say, and nothing detectable about when an accident is coming.

8 Robustness checks

Registered robustness checks. Appendix A lists all 30. Each changes one element of the design: the gap, the event types counted, the plant links, the end of the window, storms and disasters, the model family, and others. Apart from the longer windows of R1, every check that keeps the registered design gives an estimate between 0.66 and 1.00, as the bias predicts. The two that handle the days after accidents in other ways give 1.15 (0.88 to 1.50, dropping those days) and 1.01 (0.65 to 1.58, adding a marker for an accident in the past year). Only the check that compares each plant across its whole history is clearly above 1, which is the long-run association of Section 7.

Negative control. Accidents at other plants in the same county and sector should not be predicted by this plant’s events once the statewide count is in the model. They are not: 1.00 (0.96 to 1.04). Shocks shared by nearby plants of the same kind do not drive the results.

All 1,152 variants. The plan required the main test to be run over every combination of the end of the window (four choices), the gap (four), the window length (four), the event types counted (three), the plant links (three) and whether duplicate accidents are merged (two). The median of all 1,152 variants is 0.76, and 58% are significantly below 1. Window length drives almost all of the variation (medians 1.01, 0.87, 0.68 and 0.48 for 30, 90, 180 and 365 days), as the bias predicts, and the other choices barely move the estimate. Only 10 variants (0.9%) are significantly above 1. Merging duplicates makes no difference to them, so they are 5 distinct settings, all in one corner: a 30-day window ending one or two days before the accident, all event types, and links from the federal registry only (1.18 to 1.19). A one-week gap removes them, which pointed to something in the last week before accidents. Section 9 follows it up.

The plant’s own chemicals. A release of a chemical the plant handles in its RMP processes might be a more specific warning than any emission event. Of the event occurrences with listed contaminants, 26% (17,518 of 67,484) released such a chemical, most often propane, ethylene, isobutane, hydrogen sulfide, propylene, butane and pentane. Counting only those events gives 0.91 (0.77 to 1.07) as registered and 1.06 (0.84 to 1.34) with the accident’s own reports removed.

Satellite heat. Every exposure above is reported by the company. Counting instead the nights on which the VIIRS satellite detected heat within 1 km of the plant gives 1.09 (0.90 to 1.32) as

registered and 1.13 (0.93 to 1.38) with detections on the accident day and the two days after removed (238 accidents at 132 plants, 2012 to 2021). An exposure that does not depend on self-reporting shows no clear rise either, although its interval is wide.

Breakdowns. By sector, the registered estimates are 1.19 for refining (0.87 to 1.65, 22 plants), 0.83 for chemicals (0.67 to 1.03), 1.16 for gas processing (0.81 to 1.64) and 0.38 for other plants (0.21 to 0.69). These are estimates only, with the bias still in them, and no claim is made for any sector. No cause category or unit type stands out. The cause categories come from keyword rules written on a development sample. On a separate random sample of 400 events, the rules agreed with our hand coding, done without seeing the rules' output, on the primary cause 72% of the time (Cohen's kappa 0.67). Events with process-control causes (equipment, instruments, upsets, operator error) were no more closely tied to accidents than events with outside or planned causes (ratio of the two, 1.10, 0.82 to 1.46).

9 A lead: planned work in the week before an accident

Planned work reports were registered as a secondary exposure over the 90-day window, where they show nothing (Section 5). Counting them over the last week before an accident was not in the plan. It was prompted by the corner of the multiverse described above, and we report it as exploratory, a lead to be tested on new data.

The positive corner came from one kind of report. When planned work (maintenance, startup or shutdown) had been reported at the plant as starting in the 7 days before, the accident risk was higher: 1.66 per doubling of such reports (1.27 to 2.19). Unplanned events in the same week showed nothing (1.02, 0.79 to 1.31). Of the 541 accident days, 32 (5.9%) had planned work reported at the same plant in the week before, against 3.6% of all days in the comparison. Put as a yes-or-no question, accidents were about 1.6 times as likely when any planned work had been reported in the past week (1.61, 1.08 to 2.39). Applied to the average of 33 accidents per 1,000 plant-years, that ratio would mean about 53 in the weeks after planned work. The 32 accidents occurred at 17 plants.

The model's p -value (0.0003) is too optimistic. When the dates of planned work were shuffled within each plant-year 500 times, with accidents held fixed, the shuffled estimates varied about 1.3 times more than the model's standard error allows. Judged against that spread, $p = 0.018$ (two-sided). Table 5 gives the checks for other explanations. The lead holds in most of them. It is strongest for startups and weakest for routine maintenance, it holds for work that was notified to TCEQ before it started (so the report cannot be a report of the accident), and it is not explained by storms, by restarts after earlier accidents, or by any one plant or year. It is weaker in 2006 to 2010 alone (1.33, not significant).

What the accident records say. RMP has no field for the operating state of the unit, so the evidence here is indirect and weak. With 32 accidents and 14 factors compared, none of the differences below survives a correction for multiple comparisons. For each accident, plants mark contributing factors. Compared with the other 513 accidents, the 32 that followed planned work were more often marked as initiated by human error (53% against 32%), improper procedure (41% against 23%), maintenance (34% against 20%) and management error (19% against 5%), and no more often as equipment failure (41% against 45%). That is the pattern one would expect from non-routine work. On the other hand, none of the 32 records mentions a startup, shutdown, turnaround or restart in its free-text fields (0 of 32, against 0.8% of the other accidents).

Table 5: Checks of the planned-work lead. Rate ratio per doubling of planned work reports starting in the 7 days before, unless stated.

Check	Result
All planned work (maintenance, startup, shutdown)	1.66 (1.27 to 2.19)
Any planned work in the past week, yes or no	1.61 (1.08 to 2.39)
Startups only / shutdowns only / maintenance only	2.00 / 1.60 / 1.08
Work notified to TCEQ before it started	1.68
Days 1 to 3 before / days 4 to 7 before	1.71 / 1.61
Storm days removed (three storm records, each and together)	1.63 to 1.80
Accidents within 120 days after an earlier accident removed	1.66
Each plant left out in turn	1.59 to 1.73
The two plants with the most cases (15 of 32) left out	1.71
Each calendar year left out in turn	1.57 to 1.77
2003 to 2005 / 2006 to 2010 / 2011 to 2021	2.10 / 1.33 / 1.74
Simulated accident dates, 200 draws	median 0.97; none as high as 1.66

How it fits what is known. That startups and shutdowns are riskier than normal operation is long established, although the figures usually quoted rest on thinner evidence than they suggest. The Chemical Safety Board states that process safety incidents occur five times more often during startup than during normal operation [44]. Its source, the Center for Chemical Process Safety [45], cites an insurer’s compilation of large property losses for that figure, and elsewhere in the same book describes the insurer’s finding as covering all non-normal operation, startups and shutdowns together. Duguid’s analysis of 1,000 major incidents in the process industries coded 45% as occurring when the plant was not in normal operation [46], and two industry engineers have written that a typical refining or petrochemical plant spends less than 10% of its time in such transient operations [47]. EPA has warned that startup is a particularly vulnerable time [48]. Our point estimates, highest for startups and lowest for routine maintenance, fit this literature, although their intervals overlap (startups 2.00, 1.14 to 3.51; maintenance 1.08, 0.72 to 1.62). What may be new is that a regulator’s public records show it. We found no published study that uses regulator records of planned work to estimate accident risk day by day.

Another reading. A unit that starts to misbehave may be shut down for planned work, and the accident may then follow on restart. In that case planned work marks a unit already in trouble, as when an early symptom of an illness prompts the treatment that is later blamed for it [49]. That 80% of these reports were filed before the work began, and that the lead holds for such work, narrows this reading but does not rule it out. Either way, in these data the days after planned work were associated with more accidents.

Further tests. Before any further test was run, we wrote down how the lead would be tested and when a test would be too weak to run. The first further test used National Response Center serious reports at TCEQ sites that had no RMP accident within three days, 205 report days at 104 sites. It is not fully independent: it uses the same exposure and mostly the same plants and years. Only 12 of these days had planned work in the preceding week, and a check run first showed that a true effect of 1.5 would be detected only 30% of the time. The result, 1.37 (0.65 to 2.89), is inconclusive: too wide to confirm or rule out the lead. A second test in Louisiana was not feasible: Louisiana publishes no general planned-work reports, and the few dated notices at RMP-linked sites (read

from scanned documents, partly with an AI transcription model; plan D13) gave a power of 7% to 10%, below the threshold set in advance. A test on future accidents in Texas alone would also be weak, because planned work reports at linked plants fell from 1,000 to 1,400 a year in 2003 to 2009 to 65 to 110 a year in 2023 to 2025. The reason for the decline is not established. One possibility is that planned emissions came to be covered by permits and no longer needed a report. Such reports have nearly stopped, so the lead is hard to test.

10 What this means

For plant process safety teams. On this evidence, a rising count of reported emission events is not a short-term alarm for a serious accident at the same plant. The events still matter. Each is a loss of containment with its own cause, and investigating it may prevent a worse release whether or not the count predicts anything. A small-event program can be worth running for what each investigation teaches, and the CCPS guide notes that better reporting can raise counts while serious incidents fall [3]. Startups and shutdowns remain where the process safety literature says the risk is, and the lead in Section 9 is consistent with it. Teams that test their own indicators against their own incident records should count how many indicator reports the incidents themselves generate before reading the result (Section 6).

For regulators. Public event counts go with modestly more accidents across plants and over long stretches of time, but a plant's past accidents and its size rank risk better, and events add little on top. Federal inspections, violations and enforcement showed no clear association with accidents in our models. Because inspections are aimed at plants with problems, that does not mean they have no value. Notices of planned startups, shutdowns and maintenance were the one public record that pointed at days of higher risk, and at linked plants they have fallen from more than 1,000 a year before 2010 to 65 in 2025. It would be worth knowing why they declined and whether the same information is still collected in another form. Comparing RMP with other federal records also suggests that many serious events at RMP plants never appear in RMP, although many of them may not have been reportable under the rule (Section 11).

For people who live near these plants. Emission events matter for air quality and health on their own: they raise nearby pollution and have been linked to deaths among older residents [24]. This paper does not question that. It asks one narrower question, whether the count of events warns in the weeks before that a serious accident is coming, and we found no sign that it does. On the related claim that plants with many events carry more accident risk [4], the data give weak support: between plants, a doubling of events in one year goes with 12% to 15% more accidents the next (significant only once inspections and enforcement are in the model), and about 9% once events in the month after an accident are left out.

For researchers. Any study that asks whether reported incidents predict reported accidents should check whether the accidents generate incident reports of their own. Here they did, and without the correction the data seemed to say that events protect against accidents. Registering the plan in advance did not prevent this, but it kept it visible: the registered result, its withdrawal and the correction are all on the record. The linked data are public, and the plan for testing the planned-work lead is written down for anyone with new data.

11 Scope and limitations

We state the boundaries of the result plainly.

Which plants. The study covers Texas plants that file RMPs and could be linked to TCEQ sites. The main test uses the 216 of them with an accident between 2003 and 2021. Plants without an RMP, and the 44 accidents at plants without a TCEQ link, are outside it. Of the 545 accidents, 73 are at plants that reported no unplanned event in their study period, so they add almost nothing to the comparison. Other states with comparable public records might give a different answer.

Which accidents. RMP records only accidental releases of listed chemicals from covered processes with the consequences the rule names, and plants report them themselves. Other records suggest that many serious events at RMP plants have no RMP record. Of 212 serious National Response Center reports at RMP-linked plants, only 35 (17%) match an RMP accident within three days, and only 14 of the 74 that name a chemical on the RMP list; of 63 OSHA fatality and catastrophe investigations, 5 (8%); of 255 OSHA severe injury reports, 9 (4%); and of 25 incidents at RMP facilities found in a hand-checked random sample of 500 news links, 9. So between 64% and 96% of these serious events, depending on the source, have no RMP record. That is an upper bound on what RMP misses, not a count, because many of these events were not releases of listed chemicals from covered processes. For the most severe accidents, a review of the Chemical Safety Board’s Texas investigations found 1 of 20 that appears reportable but is missing from RMP. The National Response Center replication speaks to this. Its sites are mostly the same plants (88% of its report days are at RMP-linked plants), but its events are mostly different: only 33 of its 218 report days fall within three days of an RMP accident. It gives the same answer, so the main result is unlikely to be an artifact of gaps in RMP alone.

How large an effect. Rises larger than about 30% per doubling of recent events are unlikely. Smaller rises cannot be ruled out.

What is counted. The exposure is what plants report. The checks in Section 3 show that the reports mostly record real releases, and the satellite test in Section 8 gives the same answer with an exposure the plants do not report, but neither covers every kind of upset. Counts are per plant, not per unit, so a warning confined to one unit of a large plant would be diluted. In RP 754 terms, many TCEQ events are flaring, a demand on a safety system that RP 754 would count in Tier 3, and larger releases may reach Tier 1 or 2. Our test is of the public count, not of any one tier [2].

The correction was added after the fact. The removal of the accident’s own reports was adopted after the registered result was seen, when a review of the model traced the below-1 estimates to the accident’s own reports. The simulation in Section 6 supports it, and the calibration on simulated dates shows how much the removal itself moves the estimate. That calibration places accidents at random, while real accidents cluster: about one in five follows another at the same plant within 120 days. We did not adapt the case-series method for exposures changed by the event [32], which was built for a different setting and would need reworking for a count of events. The registered result is reported beside the corrected one in every table.

Prediction, not cause. The question is whether events warn of accidents, not whether they cause them. A modest long-run association can reflect common causes, such as age of equipment or the quality of management, as easily as any direct link.

The planned-work lead. It was found after looking at 1,152 variants and then splitting the result, it rests on 32 accidents at 17 plants, mostly from years when planned work reports were common, and its first further test was inconclusive.

Other limits. Cause categories come from keyword rules (agreement with hand coding 72%, kappa 0.67), which rarely assign operator error. Links answered “can’t tell” were treated as no link, which shrinks the National Response Center and OSHA checks. One registered robustness check was not run: relinking plants with TCEQ’s own site coordinates, for which the records request was withdrawn. It could have added at most about 10 of the 589 accidents in the window. Accidents after 2021 are left out of the main test because plants report late.

12 Ethics and reporting

This study uses public records only. Following a rule fixed in the registered plan, no plant or company in the study is named or ranked in this paper, and results are reported by sector, region, cause and unit type, with any group of fewer than five plants merged or left out. The plant-link review, the review of National Response Center and OSHA links, and the coding of causes were done by hand without access to event counts or accident data. All registered analyses are reported here or in the results reports in the repository, whatever they showed. The author declares no competing financial interests.

13 Reproducibility

Every table and figure in this paper is produced by a script in the repository, and one command, `python src/run_all.py`, reruns them all in order from the tables in the repository (about 10 hours on a four-core laptop). Every downloaded file and every derived table is listed with its SHA-256 hash, the date it was retrieved, and, for the RMP data, the source commit. Random draws use fixed seeds. The main estimate was computed by two separately written programs that agree to ten decimal places, and the exposure counts were rebuilt independently for 1,400 plant-days with no differences. The registered plan, its deviation log and the code for every analysis in this paper, including those that were stopped, are in the repository.

14 Conclusion

Small emission events in Texas are real, frequent and public. We found no sign that they rise before serious accidents at the same plant: across 545 accidents, a plant’s count of recent events was no higher before an accident than at other times in the same year, and rises in risk larger than about 30% per doubling are unlikely. Between plants and over long stretches of time, more events go with modestly more accidents, which says something about where risk is but little about when an accident will happen. Reaching this answer depended on a correction found after the registered result was in. Accidents generate their own event reports, and until those were removed the data seemed to say that events protect against accidents. Notices of planned work, mainly startups and

shutdowns, were the one kind of report that pointed at days of higher risk, and that lead still needs a proper test.

Data and code availability

The code, the derived tables, the hand-review answers, the manifests of SHA-256 hashes, the registered analysis plan and its deviation log are at <https://github.com/OrygnsCode/emission-events-warning> and archived with this preprint at <https://doi.org/10.5281/zenodo.23348378>, together with a copy of the TCEQ report pages as downloaded in October 2026. The tables identify sites by the public ID numbers that TCEQ and EPA assign, so that every link between the two databases can be checked. The registration is at <https://osf.io/nf426>. The source data are public: TCEQ’s event reports [9], the RMP database as released by the Data Liberation Project under CC BY-SA 4.0 [35], and the other sources in Table 1. Derived tables built from RMP data carry the same license. Contact: daniel@orygn.tech.

References

- [1] H. W. Heinrich. *Industrial Accident Prevention: A Scientific Approach*. McGraw-Hill, New York, 1931.
- [2] American Petroleum Institute. API RP 754 Fact Sheet. September 2021. <https://www.api.org/~media/files/oil-and-natural-gas/refining/process%20safety/rp-754-fact-sheet.pdf>
- [3] Center for Chemical Process Safety. *Process Safety Metrics: Guide for Selecting Leading and Lagging Indicators*, Version 4.1. American Institute of Chemical Engineers, 2022. https://www.aiche.org/sites/default/files/docs/pages/ccps_process_safety_metrics_-_v4.1.pdf
- [4] Air Alliance Houston. *Emission Events in Harris County 2017–2024*. February 2026. <https://airalliancehouston.org/wp-content/uploads/2026/02/AAH-Emission-Events-Final.pdf>
- [5] P. Swuste, J. Theunissen, P. Schmitz, G. Reniers, and P. Blokland. Process safety indicators, a review of literature. *Journal of Loss Prevention in the Process Industries* 40: 162–173, 2016.
- [6] J. Mendeloff. Testimony on process safety indicators, drawing on J. Mendeloff, B. Han, L. A. Fleishman-Mayer and J. V. Vesely, *Evaluation of Process Safety Indicators Collected in Conformance with ANSI/API Recommended Practice 754* (RAND, 2012). RAND Corporation, CT-392, June 2013, p. 5. https://www.rand.org/content/dam/rand/pubs/testimonies/CT300/CT392/RAND_CT392.pdf
- [7] U.S. Chemical Safety and Hazard Investigation Board. Public hearing on safety performance indicators, Houston, Texas, July 23–24, 2012. Transcript, pp. 24–26. https://www.csb.gov/assets/1/20/csb_20public_20hearing.pdf
- [8] European Commission Joint Research Centre. eMARS, the Major Accident Reporting System. <https://emars.jrc.ec.europa.eu/> (accessed October 2026).
- [9] Texas Commission on Environmental Quality. Air Emission Event Report Database. <https://www2.tceq.texas.gov/oce/eer/> (all reports retrieved October 2–3, 2026).
- [10] Texas Administrative Code, Title 30, §101.201, Emissions Event Reporting and Recordkeeping Requirements. <https://www.law.cornell.edu/regulations/texas/30-Tex-Admin-Code-SS-101-201>
- [11] Code of Federal Regulations, Title 40, Part 68, Chemical Accident Prevention Provisions, §68.42, Five-year accident history.
- [12] A. Hopkins. Thinking about process safety indicators. Working Paper 53, National Research Centre for OHS Regulation, Australian National University, May 2007. Published in *Safety Science* 47(4): 460–465, 2009. <https://doi.org/10.1016/j.ssci.2007.12.006>

- [13] The BP U.S. Refineries Independent Safety Review Panel. *The Report of the BP U.S. Refineries Independent Safety Review Panel*. January 2007. https://www.csb.gov/assets/1/20/baker_panel_report1.pdf
- [14] P. L. Yorio and S. M. Moore. Examining factors that influence the existence of Heinrich’s safety triangle using site-specific H&S data from more than 25,000 establishments. *Risk Analysis* 38(4): 839–852, 2018. <https://doi.org/10.1111/risa.12869>
- [15] P. Marshall, A. Hirmas, and M. Singer. Heinrich’s pyramid and occupational safety: a statistical validation methodology. *Safety Science* 101: 180–189, 2018. <https://doi.org/10.1016/j.ssci.2017.09.005>
- [16] M. R. Elliott, P. R. Kleindorfer, J. J. DuBois, Y. Wang, and I. Rosenthal. Linking OII and RMP data: does everyday safety prevent catastrophic loss? *International Journal of Risk Assessment and Management* 10(1/2): 130, 2008. <https://doi.org/10.1504/IJRAM.2008.021058>
- [17] P. R. Kleindorfer, J. C. Belke, M. R. Elliott, K. Lee, R. A. Lowe, and H. I. Feldman. Accident epidemiology and the U.S. chemical industry: accident history and worst-case data from RMP*Info. *Risk Analysis* 23(5): 865–881, 2003. <https://doi.org/10.1111/1539-6924.00365>
- [18] P. R. Kleindorfer, R. A. Lowe, I. Rosenthal, R. Fu, and J. C. Belke, with the collaboration of M. R. Elliott, A. L. Santiago, and Y. Wang. *Accident Epidemiology and the RMP Rule: Learning from a Decade of Accident History Data for the U.S. Chemical Industry*. Final report, Risk Management and Decision Processes Center, The Wharton School, University of Pennsylvania, and U.S. EPA Office of Emergency Management, December 2007.
- [19] U.S. Environmental Protection Agency. Accidental Release Prevention Requirements: Risk Management Programs Under the Clean Air Act; Safer Communities by Chemical Accident Prevention. Final rule. *Federal Register* 89: 17622, March 11, 2024.
- [20] U.S. Environmental Protection Agency. Accident History Document: Common Sense Approach to Chemical Accident Prevention (workbook “Analysis of RMP Reportable Accident Data”). Docket EPA-HQ-OLEM-2025-0313-0060, posted February 24, 2026. <https://www.regulations.gov/document/EPA-HQ-OLEM-2025-0313-0060>
- [21] R. Kowalewski and P. Young. Data-driven risk models could help target pipeline safety inspections. Bureau of Transportation Statistics Special Report SR-010, July 2008. <https://doi.org/10.21949/1501636>
- [22] C. F. Murphy and D. T. Allen. Hydrocarbon emissions from industrial release events in the Houston-Galveston area and their impact on ozone formation. *Atmospheric Environment* 39: 3785–3798, 2005. <https://doi.org/10.1016/j.atmosenv.2005.02.051>
- [23] N. Zirogiannis, A. J. Hollingsworth, and D. M. Konisky. Understanding excess emissions from industrial facilities: evidence from Texas. *Environmental Science & Technology* 52: 2482–2490, 2018. <https://doi.org/10.1021/acs.est.7b04887>
- [24] A. J. Hollingsworth, D. M. Konisky, and N. Zirogiannis. The health consequences of excess emissions: evidence from Texas. *Journal of Environmental Economics and Management* 108: 102449, 2021. <https://doi.org/10.1016/j.jeem.2021.102449>
- [25] Environmental Integrity Project. *The Polluter’s Playbook*. March 2023. <https://environmentalintegrity.org/wp-content/uploads/2023/03/TX-Polluters-Playbook-final-report-3.23.23.pdf>
- [26] M. Maclure. The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* 133(2): 144–153, 1991. <https://doi.org/10.1093/oxfordjournals.aje.a115853>
- [27] H. Janes, L. Sheppard, and T. Lumley. Case-crossover analyses of air pollution exposure data: referent selection strategies and their implications for bias. *Epidemiology* 16(6): 717–726, 2005. <https://doi.org/10.1097/01.ede.0000181315.18836.9d>

- [28] B. G. Armstrong, A. Gasparrini, and A. Tobias. Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis. *BMC Medical Research Methodology* 14: 122, 2014. <https://doi.org/10.1186/1471-2288-14-122>
- [29] W. Navidi. Bidirectional case-crossover designs for exposures with time trends. *Biometrics* 54(2): 596–605, 1998. <https://doi.org/10.2307/3109766>
- [30] M. A. Mittleman and E. Mostofsky. Exchangeability in the case-crossover design. *International Journal of Epidemiology* 43(5): 1645–1655, 2014. <https://doi.org/10.1093/ije/dyu081>
- [31] Y. G. Weldeselassie, H. J. Whitaker, and C. P. Farrington. Use of the self-controlled case-series method in vaccine safety studies: review and recommendations for best practice. *Epidemiology and Infection* 139(12): 1805–1817, 2011. <https://doi.org/10.1017/S0950268811001531>
- [32] C. P. Farrington, H. J. Whitaker, and M. N. Hocine. Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostatistics* 10(1): 3–16, 2009. <https://doi.org/10.1093/biostatistics/kxn013>
- [33] J. M. Wooldridge. Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory* 13: 667–678, 1997. <https://doi.org/10.1017/S0266466600006125>
- [34] W. Schroeder, P. Oliva, L. Giglio, and I. A. Csiszar. The New VIIRS 375 m active fire detection data product: algorithm description and initial assessment. *Remote Sensing of Environment* 143: 85–96, 2014. <https://doi.org/10.1016/j.rse.2013.12.008>
- [35] Data Liberation Project. EPA Risk Management Program database, released as spreadsheets (data current to December 31, 2025). <https://github.com/data-liberation-project/epa-rmp-spreadsheets>
- [36] U.S. Environmental Protection Agency. Facility Registry Service, Texas state combined file (retrieved October 2, 2026). https://ordsext.epa.gov/FLA/www3/state_files/state_combined_tx.zip
- [37] National Response Center. Incident reports. <https://nrc.uscg.mil/>
- [38] U.S. Department of Labor, Occupational Safety and Health Administration. Enforcement data and severe injury reports. <https://data.dol.gov/data-catalog/OSHA/>; <https://www.osha.gov/severe-injury-reports>
- [39] PyFixest: fast high-dimensional fixed effects regression in Python, version 0.60.0. <https://github.com/py-econometrics/pyfixest>
- [40] D. Okwor. Do small unplanned emission events rise before serious chemical accidents at Texas industrial plants? A within-plant study using TCEQ emission event reports and EPA Risk Management Program accident histories. Preregistration, Open Science Framework, October 3, 2026. <https://osf.io/nf426>
- [41] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6(2): 65–70, 1979.
- [42] S. Steegen, F. Tuerlinckx, A. Gelman, and W. Vanpaemel. Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science* 11(5): 702–712, 2016. <https://doi.org/10.1177/1745691616658637>
- [43] D. J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* 15(6): 657–680, 1987.
- [44] U.S. Chemical Safety and Hazard Investigation Board. Safety Digest: CSB Investigations of Incidents during Startups and Shutdowns. 2018. https://www.csb.gov/assets/1/6/csb_digest_-_startup_shutdown.pdf
- [45] Center for Chemical Process Safety. *Guidelines for Safe Process Operations and Maintenance*. American Institute of Chemical Engineers, New York, 1995, pp. 113 and 245.

- [46] I. M. Duguid. Analysis of past incidents in the process industries. *Hazards XX*, IChemE Symposium Series No. 154, paper 87, 2008. <https://www.icheme.org/media/9781/xx-paper-87.pdf>
- [47] S. W. Ostrowski and K. K. Keim. Tame your transient operations. *Chemical Processing*, June 2010. <https://www.chemicalprocessing.com/articles/2010/123/>
- [48] U.S. Environmental Protection Agency. Enforcement Alert: Risk of Chemical Accidents During Process Startup. EPA 300N21002, February 2021. <https://www.epa.gov/sites/default/files/2021-02/documents/ncistartupsafety-enforcementalert.pdf>
- [49] R. I. Horwitz and A. R. Feinstein. The problem of “protopathic bias” in case-control studies. *American Journal of Medicine* 68: 255–258, 1980. [https://doi.org/10.1016/0002-9343\(80\)90363-0](https://doi.org/10.1016/0002-9343(80)90363-0)

A Registered robustness checks

Each check changes one element of the main test (registered estimate 0.88, 0.75 to 1.02). Rate ratio per doubling of recent events, accidents to 2021, as registered; the trap of Section 6 is still in these estimates.

Check	Change	Result (95% interval)
R1	Windows of 30, 180 and 365 days	0.95 / 0.68 / 0.47 (Section 6)
R2	Gap of 0 / 1 / 30 days	0.93 / 0.91 / 0.84
R3	Plus opacity events / all event types	0.87 (0.75 to 1.02) / 0.87 (0.74 to 1.00)
R4	Events counted by report date	0.86 (0.74 to 1.01)
R5	Storm periods removed (NOAA)	0.83 (0.70 to 0.99)
R6	Regional instead of statewide count	0.88 (0.76 to 1.03)
R7	One-to-one plant links only	0.79 (0.66 to 0.94)
R8	Window ending 2022 / 2023 / 2025	0.87 / 0.87 / 0.85
R8b	Ending 2025, weighted for late reporting	0.83 (0.72 to 0.96)
R9	Accidents within 3 days merged	0.88 (0.75 to 1.03)
R10	Each plant left out in turn (216 fits)	0.86 to 0.90
R11	Conditional logistic / negative binomial	0.87 (0.75 to 1.02) / 0.88 (0.75 to 1.02)
R12	Strata by plant across all years	1.32 (1.14 to 1.53)
R13	Every report counted	0.87 (0.75 to 1.01)
R14	One effect per calendar month	0.87 (0.74 to 1.02)
R15	Accident chains dropped	0.82 (0.67 to 1.00)
R16	Days after accidents dropped	1.15 (0.88 to 1.50)
R17	Not-reportable and filed-in-error reports dropped	0.86 (0.74 to 1.01)
R18	Original accident dates	0.88 (0.76 to 1.03)
R19	No duplicate merges	0.87 (0.75 to 1.02)
R20	Federal registry links only	1.00 (0.81 to 1.24)
R21	Reporting-rule periods 2003–05 / 2006–14 / 2015–21	0.66 / 0.98 / 0.88
R22	FEMA disaster periods removed	0.84 (0.68 to 1.02)
R23	Hurricane and tropical storm passages removed	0.88 (0.75 to 1.02)
R24	2020 COVID months removed	0.87 (0.75 to 1.02)
R25	Freeze, lightning and wind controls	0.87 (0.74 to 1.02)
R26	Refinery use and industry output controls	0.88 (0.76 to 1.03)
R27	Marker for an accident in the past year	1.01 (0.65 to 1.58)
R28	Days after an ownership change removed	0.88 (0.74 to 1.04)
R29	Relinking with TCEQ’s own site coordinates	not run; request withdrawn (Section 11)
R30	Merged occurrences	0.88 (0.76 to 1.03)

B Changes to the registered plan

Every change made after the plan was registered is recorded, with its reason, in the plan’s deviation log in the repository (entries D1 to D16), and the registration is at <https://osf.io/nf426>. In brief:

Before any outcome analysis.

- Decisions registered with the plan itself: the main window ends in 2021 because plants report accidents late, and written rules for dates, plant links, duplicate accidents, reporting-rule periods and disasters.
- Plant-linking rules made specific and tested, and the link table frozen after a blind hand review.

- Additions from the literature review, including the process-control contrast and robustness checks R27 to R30.
- Rules for checking the event reports against satellites, air monitors and other sources, and fixes from a review of those checks.
- Code and inputs locked before the main test was run.

After the main result was seen. Each of these is labelled where it appears in the paper.

- Rules for running the remaining registered analyses.
- The correction for the accident’s own reports, with the first bound withdrawn (Sections 5 and 6).
- Rules for the National Response Center and OSHA replications, written after the blind link review and before either was run, and a news-article check of how complete RMP is (Section 11).
- A written plan for testing the planned-work lead, its first further test, and the finding that a test in Louisiana was not feasible (Section 9).
- The equivalence bounds, the simulation, the tests with the plant’s own chemicals and with satellite heat, and the time-scale rerun (Sections 5 to 8).
- Two registered exposures that had been missed, run late; the simulation rerun with the observed after-accident rate; the calibration of the correction; and the overlap of the National Response Center replication counted (Sections 5, 6 and 11).