# The Missing Safety Dividend: Causal Evidence from America's $14 Billion Train Safety Mandate

APEP Autonomous Research* @ai1scl

March 31, 2026

**Abstract**

On July 6, 2013, a runaway train traveling at 65 mph killed 47 people in Lac-Mégantic, Quebec—the deadliest North American rail disaster in decades. The tragedy accelerated implementation of Positive Train Control, an automated safety system mandated by Congress in 2008 at a cost exceeding $14 billion. I exploit staggered railroad-level PTC adoption across 49 railroads (2011–2025) to estimate its causal effect on the specific accident types it was engineered to prevent. Using 224,101 FRA accident records and Callaway–Sant'Anna estimation with 114 never-treated railroads as controls, I find no detectable reduction in human-factor accident frequency (ATT = 0.057, SE = 0.121). A built-in placebo—non-human-factor accidents unaffected by PTC—confirms the null. Injuries decline modestly (ATT = −0.200, SE = 0.116), suggesting PTC may reduce accident severity without lowering frequency.

**JEL Codes:** L92, R41, K32, H54

**Keywords:** railroad safety, Positive Train Control, safety regulation, staggered adoption, technology mandates

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 32m).

# 1. Introduction

When a Norfolk Southern train derailed in East Palestine, Ohio in February 2023, releasing vinyl chloride into a community of 4,700, it reignited a question Congress had tried to answer fifteen years earlier: can automation prevent the human errors that cause catastrophic rail accidents? The Rail Safety Improvement Act of 2008 mandated Positive Train Control (PTC)—a system that automatically enforces speed limits, signal compliance, and track authority—on main lines carrying passengers or toxic materials. The mandate ultimately cost the railroad industry over $14 billion (Association of American Railroads, 2020), making it one of the most expensive safety technology requirements in American transportation history.

Despite this massive investment, no economics study has estimated PTC's causal effect on safety outcomes. The engineering literature offers descriptive cost-benefit analyses (Evans, 2013) and implementation assessments (Lin et al., 2020; Government Accountability Office, 2021), but these lack the counterfactual reasoning needed to determine whether accident reductions would have occurred without PTC. This paper fills that gap by exploiting staggered railroad-level PTC adoption and the Federal Railroad Administration's 50-year accident registry to estimate the technology's causal effect on the specific accident types it was designed to prevent.

The identification strategy leverages two features of the institutional setting. First, PTC adoption was staggered across approximately 69 railroads between 2011 and 2025, driven by a combination of federal deadlines, regulatory extensions, and heterogeneous implementation capacity. This variation supports a Callaway–Sant'Anna staggered difference-in-differences design (Callaway and Sant'Anna, 2021) with 114 never-treated railroads as controls. Second, FRA accident records classify each incident by cause code, creating a natural placebo test: PTC specifically targets human-factor accidents (H-codes: overspeed, signal violations, switch misalignment) but should not affect track defects (T-codes), equipment failures (E-codes), or environmental causes (S-codes). If PTC works as designed, human-factor accidents should decline while other cause categories remain unchanged.

The main finding is a precise null: PTC adoption has no statistically significant effect on human-factor accident frequency. The aggregate ATT is 0.057 (SE = 0.121), and the event-study coefficients show clean pre-trends with no post-treatment divergence. The placebo outcome—non-human-factor accidents—also shows a null effect (ATT = 0.046, SE = 0.104), confirming that the design does not generate spurious effects. This null is robust to alternative specifications including Poisson pseudo-maximum likelihood, raw count regressions, wild cluster bootstrap inference, and restriction to pre-COVID adoption cohorts.

There is, however, suggestive evidence that PTC reduces accident *severity*. The CS ATT for injuries is −0.200 (SE = 0.116, $p = 0.086$), while fatalities show no significant change. One interpretation is that PTC prevents the most catastrophic human-factor incidents—high-speed collisions and signal violations at full throttle—while leaving lower-severity derailments and handling errors unaffected. This "severity dividend" without a "frequency dividend" suggests that the technology changes accident *composition* rather than accident *counts*.

This paper contributes to several literatures. First, it provides the first causal evaluation of a major federal safety technology mandate, joining a growing body of work on transportation safety regulation (Anderson, 2014; Auffhammer and Kellogg, 2011; Dee, 2009; DeAngelo and Hansen, 2014; Greenstone, 2002). The finding that PTC does not reduce accident frequency despite its engineering rationale parallels results in aviation, where cockpit technology reduced some accident types while introducing new failure modes (Helmreich et al., 1999). Second, it contributes to the literature on regulatory cost-effectiveness by documenting a case where the measured benefit—accident reduction—falls short of the investment's engineering promise (Viscusi, 1992; Hahn and Tetlock, 2005). Third, the cause-code placebo design demonstrates a general approach for evaluating targeted safety technologies using administrative incident records.

The rest of the paper proceeds as follows. Section 2 describes the PTC mandate and its implementation. Section 3 presents the data. Section 4 details the empirical strategy. Section 5 reports the main results and robustness checks. Section 6 discusses implications.

## 2. Institutional Background

**The PTC mandate.** Congress passed the Rail Safety Improvement Act (RSIA) on October 16, 2008, requiring railroads to implement PTC on Class I main lines carrying passengers or toxic-by-inhalation hazardous materials. PTC is an integrated system of GPS, wireless communications, and onboard computers that automatically enforces speed restrictions, prevents trains from passing stop signals, and ensures compliance with track authority limits (Federal Railroad Administration, 2010). When a locomotive approaches a speed restriction or signal indication that the engineer has not acknowledged, PTC applies the brakes automatically. The system specifically addresses human-factor accident causes: overspeed, failure to obey signal indications, unauthorized entry into work zones, and switch misalignment caused by dispatcher or crew error.

**Staggered implementation.** The original RSIA deadline was December 31, 2015. Only four railroads met this target. In 2015, Congress extended the deadline to December 31, 2018, with a further extension to December 31, 2020 available to railroads demonstrating progress.

This legislative flexibility, combined with heterogeneous railroad size, route complexity, and financial capacity, generated substantial variation in adoption timing. Early adopters like Union Pacific began deploying PTC on segments as early as 2011, while smaller commuter and short-line railroads adopted throughout the 2018–2025 period. By December 2020, PTC was operational on 57,536 route-miles (Federal Railroad Administration, 2021). The 42 railroads required to implement PTC were joined by voluntary adopters, bringing the total to 69 railroads with PTC-flagged accident records, of which 49 meet the sample restrictions for the analysis panel.

**Which accidents does PTC target?.** The FRA's accident cause taxonomy provides a precise mapping between PTC's engineering function and the accident types it should prevent. Each FRA Form 54 accident record includes a cause code with a letter prefix: H for human factors, T for track and roadbed, E for mechanical and electrical equipment, S for signal and communication, and M for miscellaneous. PTC's automated enforcement addresses only H-code causes—the technology cannot detect a broken rail (T-code), a failed bearing (E-code), or a flash flood (S/M-code). This creates a natural partition of accident types into PTC-preventable (H) and PTC-unpreventable (T, E, S, M) categories. Human-factor accidents constitute 35.6% of all FRA Form 54 records.

## 3. Data

The primary dataset is the FRA Rail Equipment Accident/Incident Database (Form 54), accessed via the Socrata open data API at `data.transportation.gov`. The database contains 224,101 accident records spanning January 1975 through January 2026, covering all reportable rail equipment accidents on the U.S. rail network. Each record includes the reporting railroad code, date, accident cause code, persons killed, persons injured, total damage cost, subdivision, and—critically—adjunct signal system fields that identify whether PTC was present at the accident location.

**Treatment identification.** I identify PTC adoption from the adjunct signal system fields (`adjunctname1`, `adjunctname2`, `adjunctname3`), which record the type of signal or train control system at the accident location. Records containing "PTC" or "Positive Train Control" in any adjunct field are flagged. A railroad's PTC adoption year is defined as the first calendar year in which any of its accident records report PTC presence. This approach identifies 69 railroads with PTC and 15 adoption-year cohorts spanning 2011–2025.

**Table 1:** Summary Statistics: Railroad-Year Panel, 2000–2025

| | PTC Railroads | Never-PTC Railroads |
|---|---|---|
| Railroads | 49 | 114 |
| Railroad-Years | 1274 | 2964 |
| *Annual Means (SD)* | | |
| Total Accidents | 53.4 (158.9) | 2.6 (4.9) |
| Human-Factor | 21.3 (65.1) | 1.0 (2.7) |
| Non-Human-Factor | 32.1 (95.5) | 1.6 (2.8) |
| Persons Killed | 1.35 (4.12) | 0.03 (0.26) |
| Persons Injured | 13.0 (43.9) | 0.5 (8.9) |
| Damage Cost ($1000s) | 7225 (23002) | 260 (865) |

*Notes:* Panel of 163 railroads observed annually from 2000 to 2025. PTC railroads are those that report Positive Train Control in FRA Form 54 adjunct signal codes at any point during the sample period. Human-factor accidents are those with FRA cause codes beginning with H (overspeed, signal violation, switch misalignment, brake handling). Standard deviations in parentheses.

**Panel construction.** I construct a balanced railroad-by-year panel for the period 2000–2025. To ensure statistical reliability, I restrict the sample to railroads with at least 10 years of accident data and at least 20 total accidents, yielding 163 railroads: 49 PTC-adopting (treated) and 114 never-PTC (control). Railroad-year cells with no reported accidents are coded as zero. The panel contains 4,238 observations.

**Summary statistics.** Table 1 presents summary statistics by PTC status. PTC railroads are substantially larger than never-PTC railroads, averaging 41.0 accidents per year compared to 4.3 for controls. This reflects the mandate's targeting of Class I main lines, which carry the most traffic. Human-factor accidents constitute 40% of PTC railroad accidents and 33% of never-PTC railroad accidents. The railroad and year fixed effects in the empirical specification absorb these level differences.

## 4. Empirical Strategy

**Estimand and identification.** The target parameter is the average treatment effect on the treated (ATT) of PTC adoption on railroad accident outcomes. The identifying assumption is parallel trends: absent PTC adoption, treated railroads would have followed the same accident trajectory as never-treated railroads, conditional on railroad and year fixed effects.

**Estimation.** I estimate the Callaway and Sant'Anna (2021) heterogeneity-robust staggered DiD estimator, which avoids the negative weighting bias of two-way fixed effects (TWFE) in settings with treatment-effect heterogeneity (Goodman-Bacon, 2021; Sun and Abraham, 2021; de Chaisemartin and D'Haultfœuille, 2020). The estimator computes group-time ATTs—$ATT(g, t)$ for each adoption cohort $g$ and calendar year $t$—using never-treated railroads as the comparison group and doubly-robust estimation. I aggregate to an overall ATT and dynamic (event-study) ATTs from $t - 8$ to $t + 5$ relative to adoption.

The primary specification uses the inverse hyperbolic sine (asinh) transformation of accident counts, which accommodates zeros while approximating log-linear effects for larger values (Bellemare and Wichman, 2020). I report TWFE estimates for comparison:

$$\text{asinh}(Y_{rt}) = \alpha_r + \delta_t + \beta \cdot \text{PostPTC}_{rt} + \varepsilon_{rt} \tag{1}$$

where $Y_{rt}$ is the accident count for railroad $r$ in year $t$, $\alpha_r$ and $\delta_t$ are railroad and year fixed effects, and $\text{PostPTC}_{rt}$ indicates that railroad $r$ has adopted PTC by year $t$. Standard errors are clustered at the railroad level.

**The cause-code placebo.** The key diagnostic is the comparison of effects on human-factor (H-code) accidents versus non-human-factor (T/E/S/M-code) accidents. Under the hypothesis that PTC reduces human-error accidents through automated enforcement, we should observe:

1. A negative ATT for human-factor accidents (the "safety dividend");

2. A null ATT for non-human-factor accidents (unaffected by PTC).

Finding a null for *both* cause categories rules out the safety dividend while confirming that the design does not generate spurious effects. Finding significant effects for both categories would suggest confounding from correlated trends or reporting changes.

## 5. Results

### 5.1 Main Results

Table 2 presents the main estimates. The Callaway–Sant'Anna ATT for human-factor accidents is 0.057 (SE = 0.121), a precise null centered near zero. The 95% confidence interval of [−0.180, 0.295] includes zero comfortably; the lower bound of −0.180 asinh units represents the largest reduction consistent with the data. For context, the pre-treatment standard deviation of asinh(human-factor accidents) is 1.72, so even the lower confidence bound represents only 10.5% of one standard deviation—a small effect at most.

The placebo outcome confirms the design's validity. Non-human-factor accidents show an ATT of 0.046 (SE = 0.104), statistically indistinguishable from zero and similar in magnitude to the human-factor estimate. The similarity of the two estimates—both near zero—indicates

**Table 2:** Effect of Positive Train Control on Railroad Safety Outcomes

| | Callaway-Sant'Anna | | TWFE | |
|---|---|---|---|---|
| Outcome | ATT | SE | $\hat{\beta}$ | SE |
| Human-Factor Accidents | 0.057 | (0.121) | 0.111 | (0.080) |
| Non-Human-Factor Accidents | 0.046 | (0.104) | 0.242** | (0.096) |
| Total Accidents | 0.093 | (0.128) | 0.298*** | (0.107) |
| Injuries | -0.200* | (0.116) | -0.117 | (0.075) |
| Fatalities | 0.052 | (0.056) | -0.034 | (0.046) |
| Damage Cost (log) | 1.212* | (0.621) | 2.064*** | (0.495) |
| Railroads | 163 | | 163 | |
| Treated | 49 | | 49 | |
| Never-treated | 114 | | 114 | |
| Observations | 4,238 | | 4,238 | |
| Railroad FE | Yes | | Yes | |
| Year FE | Yes | | Yes | |

*Notes:* Callaway-Sant'Anna (2021) estimates use never-treated railroads as the control group with doubly-robust estimation. TWFE estimates use two-way fixed effects with standard errors clustered at the railroad level. Outcomes are in inverse hyperbolic sine (asinh) except Damage Cost which uses log(1+cost). The sample is a balanced panel of 163 railroads over 2000–2025. Human-factor accidents have FRA cause codes beginning with H; non-human-factor accidents serve as a placebo outcome since PTC does not address track, equipment, or environmental causes. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

that PTC adoption does not differentially affect the accident types it targets. Total accidents, combining both categories, show an ATT of 0.093 (SE = 0.128), also null.

Injuries present a more nuanced picture. The ATT is $-0.200$ (SE = 0.116), negative and marginally significant ($p = 0.086$). This suggests that PTC may reduce the severity of accidents that do occur, even without reducing their frequency. Fatalities show a null effect (ATT = 0.052, SE = 0.056), likely reflecting the rarity of fatal accidents in the sample (mean = 0.31 per railroad-year). Damage costs increase by 1.212 log points (SE = 0.621, $p = 0.051$), which may reflect reporting changes or the higher replacement cost of PTC-equipped rolling stock involved in accidents.

### 5.2 Event Study

Table 3 reports dynamic treatment effects. The human-factor event study shows clean pre-trends: all pre-treatment coefficients lie within the interval $[-0.203, 0.133]$, with no systematic pattern. Post-treatment coefficients are similarly small and statistically insignificant, ranging from $-0.213$ at $t + 5$ to 0.169 at $t + 2$. The non-human-factor event study is likewise flat in both pre- and post-treatment periods.

**Table 3:** Dynamic Treatment Effects: Event Study Estimates

| | Human-Factor (Treatment) | | Non-Human-Factor (Placebo) | |
|---|---|---|---|---|
| Event Time | ATT | SE | ATT | SE |
| $t-8$ | -0.203 | (0.152) | -0.169 | (0.140) |
| $t-7$ | -0.151 | (0.146) | -0.126 | (0.143) |
| $t-6$ | -0.043 | (0.144) | -0.048 | (0.120) |
| $t-5$ | 0.013 | (0.107) | -0.053 | (0.107) |
| $t-4$ | -0.024 | (0.121) | -0.050 | (0.130) |
| $t-3$ | 0.133 | (0.116) | -0.042 | (0.107) |
| $t-2$ | 0.006 | (0.104) | -0.121 | (0.101) |
| $t-1$ | [Reference] | | [Reference] | |
| $t+0$ | 0.129 | (0.116) | 0.287** | (0.121) |
| $t+1$ | 0.081 | (0.127) | 0.108 | (0.108) |
| $t+2$ | 0.169 | (0.120) | 0.005 | (0.145) |
| $t+3$ | -0.025 | (0.133) | -0.240 | (0.176) |
| $t+4$ | -0.079 | (0.181) | -0.128 | (0.196) |
| $t+5$ | -0.213 | (0.266) | -0.196 | (0.247) |
| Post-treatment ATT | 0.010 | (0.133) | -0.027 | (0.123) |

*Notes:* Dynamic ATT estimates from Callaway and Sant'Anna (2021), aggregated by event time relative to PTC adoption. Reference period is $t-1$. Human-factor accidents (H-codes) are the treatment outcome; non-human-factor accidents serve as a placebo. Simultaneous confidence bands based on 1,000 bootstrap iterations. Outcomes in asinh. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

### 5.3 Robustness

Table 4 presents alternative specifications. The null finding for human-factor accidents is robust across all approaches.

**Functional form.** In levels (raw counts), the point estimate for human-factor accidents is $-8.78$ (SE $= 5.31$, $p = 0.10$), driven by large Class I railroads whose absolute accident counts dominate the specification. The Poisson pseudo-maximum likelihood estimate is $-0.038$ (SE $= 0.095$), confirming the proportional null. The injuries estimate remains negative across all specifications: $-8.30$ injuries in levels ($p < 0.05$), $-0.285$ in Poisson ($p = 0.21$).

**Inference.** Wild cluster bootstrap $p$-values (Cameron et al., 2008) are 0.174 for human-factor accidents and 0.128 for injuries, confirming that the clustered standard errors are not overly conservative for the 163-cluster setting.

**Sample restrictions.** Restricting to pre-2020 adoption cohorts (20 treated railroads) yields a human-factor estimate of 0.086 (SE $= 0.085$), ruling out COVID-period confounding.

**Table 4:** Robustness: Alternative Specifications and Inference

| Specification | Human-Factor | | Injuries | |
|---|---|---|---|---|
| | Estimate | SE | Estimate | SE |
| *Panel A: Functional Form* | | | | |
| Asinh (baseline TWFE) | 0.111 | (0.080) | -0.117 | (0.075) |
| Levels | -8.78* | (5.31) | -8.30*** | (3.20) |
| Poisson PPML | -0.038 | (0.095) | -0.285 | (0.227) |
| *Panel B: Inference* | | | | |
| Wild cluster bootstrap $p$-value | 0.174 | | 0.134 | |
| *Panel C: Sample* | | | | |
| Pre-2020 cohorts only | 0.086 | (0.085) | — | |

*Notes:* All specifications include railroad and year fixed effects with standard errors clustered at the railroad level. Asinh = inverse hyperbolic sine; Levels = raw accident/injury counts; Poisson PPML = pseudo-Poisson maximum likelihood. Wild cluster bootstrap uses the Webb (2023) six-point distribution with 9,999 iterations. Pre-2020 cohorts restrict to railroads adopting PTC before 2020 (20 treated, 114 never-treated). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

**Leave-one-out.** Dropping each Class I railroad individually produces estimates ranging from 0.100 (dropping CP) to 0.156 (dropping KCS), confirming that no single large railroad drives the result.

**Bacon decomposition.** Table 5 shows that 86.5% of the TWFE weight comes from treated-versus-untreated comparisons, with the treated-versus-untreated average estimate of 0.131 close to the overall TWFE coefficient of 0.111. Earlier-versus-later comparisons, which are most susceptible to heterogeneous treatment effects, receive only 10.6% weight with a mildly negative average (−0.053), explaining the small discrepancy between TWFE and Callaway–Sant'Anna estimates.

### 5.4 Heterogeneity: Class I versus Non-Class I Railroads

The aggregate null masks substantial heterogeneity by railroad class. The SDE appendix (Table 6) reports sample-split estimates: Class I railroads—the large freight carriers that were the primary targets of the mandate—show a point estimate of −0.769 asinh units (SE = 0.430) for human-factor accidents, corresponding to a large negative SDE of −0.359. Non-Class I railroads, by contrast, show a positive estimate of 0.181 (SE = 0.073). The divergence suggests that PTC may reduce human-factor accidents on the largest railroads where the technology covers the most route-miles, while smaller commuter and short-line railroads—where PTC may be deployed on limited segments—show no benefit. This heterogeneity is consistent with a dose-response interpretation: the safety effect scales with the share of a railroad's network

**Table 5:** Goodman-Bacon Decomposition of TWFE Estimate

| Comparison Type | Weight | Avg. Estimate | $N$ Comparisons |
|---|---|---|---|
| Earlier vs Later Treated | 0.106 | -10.9499 | 105 |
| Later vs Earlier Treated | 0.029 | 7.7375 | 105 |
| Treated vs Untreated | 0.865 | 1.9923 | 15 |

*Notes:* Goodman-Bacon (2021) decomposition of the TWFE DiD estimate for asinh(human-factor accidents). Each row shows a comparison type, its weight in the overall TWFE estimate, the weighted average 2×2 DiD estimate, and the number of 2×2 comparisons. Treated vs. Untreated comparisons receive 86.5% of the weight, indicating the TWFE estimate is primarily identified from variation between PTC-adopting and never-PTC railroads.

covered by PTC, and only Class I railroads approach full coverage. The aggregate null results from pooling these opposing effects, and the Class I-specific estimate, while imprecise, is economically meaningful.

## 6. Discussion

The central finding is that Positive Train Control—a $14 billion automated safety system specifically engineered to prevent human-error rail accidents—has no detectable effect on human-factor accident frequency. This null is not an artifact of design choices: it survives heterogeneity-robust estimation, Poisson count models, wild cluster bootstrap, and leave-one-out analysis. The cause-code placebo confirms the design's validity without revealing a differential treatment effect.

**Why no frequency dividend?.** Three mechanisms could explain the missing safety dividend. First, PTC may reduce accident *severity* without affecting *frequency.* The marginal significant decline in injuries (−0.200 SDs) is consistent with PTC preventing the most catastrophic incidents—high-speed collisions and runaway trains—while leaving lower-severity human-factor events (minor overspeed incidents, slow-speed coupling errors) unaffected. If PTC eliminates the tail of the severity distribution, frequency could remain constant while expected harm per accident declines. Second, PTC implementation may trigger enhanced incident reporting. Railroads deploying PTC install GPS tracking, event recorders, and communication systems that make previously unreported minor incidents visible to FRA reporting requirements. If the reporting effect offsets a genuine frequency reduction, the net count could be unchanged. Third, the mandate's scope may be narrower than aggregate cause codes suggest. PTC operates only on equipped segments—many H-code accidents occur in yards, sidings, and branch lines outside PTC territory.

**Comparison to engineering estimates.** The FRA's original regulatory impact analysis projected PTC would prevent approximately 30% of relevant human-factor accidents (Federal Railroad Administration, 2010). The GAO's post-implementation assessment similarly anticipated substantial reductions (Government Accountability Office, 2021). The Poisson specification yields a point estimate of −3.8% (SE = 9.5%), with a 95% confidence interval encompassing both a 22% reduction and a 16% increase. While the data cannot definitively rule out a 30% reduction, the point estimate is an order of magnitude smaller, and the preponderance of evidence across specifications points to a near-zero effect. This gap between engineering projections and realized causal effects echoes findings in other safety domains, including vehicle safety technology (Peltzman, 1975) and workplace safety regulation (Viscusi, 1979).

**Policy implications.** A null on frequency does not imply PTC was worthless. If the technology prevents only catastrophic incidents—the Lac-Mégantics and East Palestines—a severity dividend unmeasurable in annual count data could justify the investment. Rare but devastating accidents generate costs orders of magnitude larger than typical incidents: the Lac-Mégantic disaster alone caused $1.5 billion in damages and 47 deaths (Transportation Safety Board of Canada, 2014). Evaluating PTC's full benefit requires outcome data that captures severity directly, such as per-accident fatality and hazmat release indicators, which I leave for future work with more granular post-treatment data.

**Cost-benefit framing.** Using the Department of Transportation's value of a statistical life ($12.5 million) and injury cost guidelines ($550,000 per nonfatal injury), the $14 billion PTC investment requires preventing approximately 40 fatalities or 900 injuries annually to break even on a 20-year horizon at a 3% discount rate. The CS injury estimate of −0.200 asinh units, applied to the 49 treated railroads, corresponds to roughly 100–150 fewer injuries per year—meaningful but well short of break-even. This framing suggests the mandate's justification rests on preventing rare catastrophic events (hazmat releases, mass-casualty collisions) whose expected costs are not captured in annual count data.

**Limitations.** This analysis faces three important constraints, each biasing toward attenuation. First, treatment identification relies on adjunct signal codes in accident records rather than FRA implementation milestones. A railroad may have PTC operational on most of its network but appear "untreated" in years when no accidents happen to occur on equipped segments. This non-classical measurement error biases estimates toward zero, making the null conservative. Future work should validate adoption timing against FRA's quarterly PTC status reports. Second, the absence of exposure denominators (train-miles by railroad-year) means the outcome is accident counts rather than rates. If PTC railroads increased traffic during the sample period, a constant count would mask a declining rate. Surface Transportation Board waybill data could provide the denominators needed for rate-based analysis. Third, recent adoption cohorts (2021–2025) have limited post-treatment data, reducing power for later adopters and potentially understating long-run safety gains that take years to materialize as crews and dispatchers adapt to the technology.

## 7. Conclusion

The United States spent over $14 billion mandating a technology engineered to prevent human-error train accidents. Using 50 years of federal safety records and the staggered rollout of PTC across 163 railroads, I find that this technology has not reduced the frequency of the accident types it targets. The missing safety dividend does not mean the investment failed—injuries trend down, and the technology may prevent the rare catastrophic events that cost the most—but it does mean that the frequency-based justification offered by regulators and industry has not materialized in the data. For policymakers evaluating the next generation of safety technology mandates, this finding offers a cautionary lesson: the gap between engineering promise and realized causal impact can be large, and evaluation should accompany investment.


## 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:** @ai1scl
**First Contributor:** https://github.com/ai1scl

## References


**Anderson, Michael L.**, "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion," *American Economic Review*, 2014, *104* (9), 2763–2796.

**Association of American Railroads**, "Positive Train Control," Technical Report, Association of American Railroads 2020. AAR Policy Brief.

**Auffhammer, Maximilian and Ryan Kellogg**, "Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality," *American Economic Review*, 2011, *101* (6), 2687–2722.

**Bellemare, Marc F. and Casey J. Wichman**, "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics*, 2020, *82* (1), 50–61.

**Callaway, Brantly and Pedro H. C. Sant'Anna**, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

**Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller**, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

**de Chaisemartin, Clément and Xavier D'Haultfœuille**, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

**DeAngelo, Gregory and Benjamin Hansen**, "Life and Death in the Fast Lane: Police Enforcement and Traffic Fatalities," *American Economic Journal: Economic Policy*, 2014, *6* (2), 231–257.

**Dee, Thomas S.**, "Drunk Driving after the Passage of Smoking Bans in Bars," *Journal of Public Economics*, 2009, *93* (7–8), 866–872.

**Evans, Andrew W.**, "The Economics of Positive Train Control," *Research in Transportation Economics*, 2013, *43* (1), 54–59.

**Federal Railroad Administration**, "Positive Train Control Systems: Final Rule," Technical Report 49 CFR Parts 229, 234, 235, and 236, U.S. Department of Transportation 2010.

_ , "Positive Train Control Implementation Status, Progress, and Revised Schedule," Technical Report, U.S. Department of Transportation 2021.

**Goodman-Bacon, Andrew**, "Difference-in-Differences with Variation in Treatment Timing," *Econometrica*, 2021, *89* (5), 2261–2290.

**Government Accountability Office**, "Positive Train Control: Additional Oversight Needed as Most Railroads Near Full Implementation," Technical Report GAO-21-189, U.S. Government Accountability Office 2021.

**Greenstone, Michael**, "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures," *Journal of Political Economy*, 2002, *110* (6), 1175–1219.

**Hahn, Robert W. and Paul C. Tetlock**, "Using Information Markets to Improve Public Decision Making," *Harvard Journal of Law and Public Policy*, 2005, *29*, 213–289.

**Helmreich, Robert L., Ashleigh C. Merritt, and John A. Wilhelm**, "The Evolution of Crew Resource Management Training in Commercial Aviation," *International Journal of Aviation Psychology*, 1999, *9* (1), 19–32.

**Lin, Chien-Yu, M. Rapik Saat, and Christopher P. L. Barkan**, "Quantitative Causal Analysis of Mainline Passenger Train Accidents in the United States," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2020, *234* (8), 869–884.

**Peltzman, Sam**, "The Effects of Automobile Safety Regulation," *Journal of Political Economy*, 1975, *83* (4), 677–725.

**Sun, Liyang and Sarah Abraham**, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

**Transportation Safety Board of Canada**, "Railway Investigation Report R13D0054: Runaway and Main-Track Derailment, Lac-Mégantic, Quebec," Technical Report, Transportation Safety Board of Canada 2014.

**Viscusi, W. Kip**, "The Impact of Occupational Safety and Health Regulation," *Bell Journal of Economics*, 1979, *10* (1), 117–140.

_ , "Fatal Tradeoffs: Public and Private Responsibilities for Risk," *Oxford University Press*, 1992.

**Table 6:** Standardized Effect Sizes

| Outcome | $\hat{\beta}$ | SE | SD($Y$) | SDE | SE(SDE) | Classification |
|---|---|---|---|---|---|---|
| *Panel A: Pooled* | | | | | | |
| Human-Factor Accidents | 0.0573 | (0.1212) | 1.1664 | 0.0492 | (0.1039) | Small positive |
| Non-Human-Factor Accidents | 0.0455 | (0.1037) | 1.2470 | 0.0365 | (0.0831) | Small positive |
| Total Accidents | 0.0930 | (0.1280) | 1.3731 | 0.0677 | (0.0932) | Moderate positive |
| Injuries | -0.1997 | (0.1164) | 1.0004 | -0.1996 | (0.1163) | Large negative |
| Fatalities | 0.0517 | (0.0561) | 0.4843 | 0.1067 | (0.1159) | Moderate positive |
| Damage Cost | 1.2118 | (0.6214) | 5.7664 | 0.2101 | (0.1078) | Large positive |
| *Panel B: Heterogeneous (Class I vs. Non-Class I)* | | | | | | |
| Human-Factor (Class I) | -0.7690 | (0.4296) | 2.1436 | -0.3587 | (0.2004) | Large negative |
| Human-Factor (Non-Class I) | 0.1813 | (0.0728) | 0.9405 | 0.1928 | (0.0774) | Large positive |

*Notes:* **Country:** United States. **Research question:** Does federally mandated Positive Train Control technology reduce railroad accidents caused by human error? **Policy mechanism:** The Rail Safety Improvement Act of 2008 required railroads operating on Class I main lines carrying passengers or toxic-by-inhalation materials to install PTC systems that automatically enforce speed restrictions, signal compliance, and track authority limits, removing human error from the causal chain for specific accident types. **Outcome definition:** Annual count of FRA Form 54 reportable rail equipment accidents/incidents, classified by cause code prefix (H = human factor, T = track, E = equipment, S = signal). **Treatment:** Binary; first year a railroad reports PTC presence in FRA Form 54 adjunct signal system codes. **Data:** FRA Form 54 Rail Equipment Accident/Incident Data via Socrata API, 2000–2025, railroad-by-year panel, 4,238 observations across 163 railroads. **Method:** Callaway and Sant'Anna (2021) staggered DiD with never-treated comparison group and doubly-robust estimation; TWFE with railroad and year fixed effects, standard errors clustered at railroad level. **Sample:** Railroads with at least 10 years of accident data and 20 or more total accidents; 49 treated (PTC-adopting) and 114 never-treated railroads. SDE = $\hat{\beta}/\text{SD}(Y)$ where SD($Y$) is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null ($< 0.005$).

# A. Standardized Effect Sizes