# The Triage Tax: Queue Congestion and Forgone Vehicle Safety Recalls at NHTSA

APEP Autonomous Research* @ai1scl

March 31, 2026

**Abstract**

When regulators have fixed capacity, some cases inevitably receive less attention. I study this triage cost at the National Highway Traffic Safety Administration, where approximately 90 investigators oversee safety for 280 million vehicles. Using 1,362 defect investigations from 2000–2024, I instrument for total queue congestion with the volume of concurrent open investigations at *other* manufacturers. Each additional concurrent investigation reduces the probability that a new investigation results in a recall by 0.14 percentage points. This effect is concentrated in Preliminary Evaluations—the early triage stage—and absent for high-severity defects, consistent with rational prioritization that imposes costs on lower-profile cases. Results are stable across manufacturer-clustered and two-way clustered inference and survive leave-one-out robustness checks. The findings reveal a measurable "triage tax" on vehicle safety from regulatory capacity constraints.

**JEL Codes:** L51, K32, H11

**Keywords:** regulatory capacity, vehicle safety, NHTSA, defect investigations, queue congestion, recalls

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 39m).

## 1. Introduction

In 2014, General Motors recalled 2.6 million vehicles for faulty ignition switches linked to 124 deaths—a defect that NHTSA investigators had identified years earlier but failed to act on while handling hundreds of other open cases (Valukas, 2014). The episode prompted Congressional hearings and a Department of Transportation Inspector General audit, which found that NHTSA's Office of Defects Investigation (ODI) "fails to meet timeliness goals across all investigation types" (U.S. Department of Transportation Office of Inspector General, 2023). The policy question is whether this capacity constraint has systematic consequences: does investigation congestion cause regulators to close cases that would otherwise lead to safety recalls?

This paper provides the first causal evidence that queue congestion at a safety regulator reduces the probability of enforcement action. I study the universe of 1,362 closed Preliminary Evaluation (PE) and Engineering Analysis (EA) investigations opened by NHTSA between 2000 and 2024, linked to 2.19 million consumer complaints and 320,000 recall campaigns. The key institutional feature is that ODI operates with approximately 90 investigators responsible for monitoring defects across more than 280 million registered vehicles—a staffing level that has remained roughly constant even as the vehicle fleet has grown by 25 percent since 2000 (U.S. Department of Transportation Office of Inspector General, 2023).

The identification challenge is that queue length may be correlated with unobserved factors that independently affect investigation outcomes. I address this with a reduced-form strategy that exploits variation in the volume of concurrent open investigations at *other* manufacturers when a new investigation is opened. Conditional on component category and year fixed effects, the timing of defect discoveries at Ford should not directly affect whether a Honda investigation leads to a recall, except through the bandwidth channel. This "other-manufacturer queue" instrument captures exogenous variation in examiner workload driven by the staggered timing of unrelated defect discoveries across the vehicle fleet.

The main finding is that each additional concurrent investigation at other manufacturers reduces the recall probability by 0.14 percentage points ($p < 0.001$). At the mean of 38 concurrent other-manufacturer investigations, this implies that queue congestion reduces the recall rate by 5.3 percentage points—an 11 percent reduction relative to the sample mean recall rate of 47.7 percent. A one-standard-deviation increase in the other-manufacturer queue (34.5 investigations) reduces recall probability by 4.8 percentage points.

The mechanism operates through triage: when investigators face a longer queue, they rationally prioritize high-profile defects and are more likely to close lower-profile cases without a recall. I document three pieces of evidence for this channel. First, the effect is concentrated in Preliminary Evaluations—the early triage stage where investigators decide whether to escalate to a full Engineering Analysis—and absent for EAs, which have already passed the initial screen. Second, congestion compresses investigation duration for low-severity cases but not for high-severity ones, consistent with investigators maintaining thoroughness for the most dangerous defects even under workload pressure. Third, congestion also increases investigation duration, though this effect is smaller and less precisely estimated.

This paper contributes to three literatures. First, it adds to the growing body of work on regulatory capacity constraints and enforcement outcomes. Kang et al. (2016) and Heese (2016) show that SEC enforcement is sensitive to resource constraints; Macher and Mayo (2007) documents similar patterns at the FDA. I provide the first evidence from vehicle safety regulation, where the stakes—measured in injuries and deaths rather than financial penalties—are arguably higher.

Second, the paper contributes to the economics of auto safety, which has focused primarily on the *effects* of recalls on firm value (Jarrell and Peltzman, 1985; Barber and Darrough, 1996; Hoffer et al., 1988), consumer behavior (Bae and Benitez-Silva, 2004; Liu, 2016), and traffic fatalities (Peltzman, 1975; Levitt and Porter, 2001), rather than on the *determinants* of whether recalls occur in the first place. By showing that recall outcomes depend not just on defect characteristics but on the regulator's opportunity cost, I identify a new margin of regulatory production.

Third, the identification strategy adds to the literature on examiner and judge instruments (Kling, 2006; Dobbie et al., 2018; Sampat and Williams, 2019; Rigbi, 2013), which exploits quasi-random variation in case assignment. My approach differs in that the instrument is not examiner identity but examiner *workload*—a dimension that varies over time rather than across individuals—making it applicable to regulatory settings where individual case assignment data are unavailable.

The findings have direct policy implications. The 2023 DOT Inspector General audit recommended increasing ODI staffing, but without an estimate of the cost of current constraints. This paper provides such an estimate: the triage tax implied by the reduced-form estimates suggests that a one-standard-deviation reduction in queue length would increase the recall rate by 4.8 percentage points, corresponding to approximately 65 additional recalls over the sample period. Given average recall sizes of 200,000–500,000 vehicles, the implied consumer safety benefit is substantial.

## 2. Institutional Background

**The NHTSA Investigation Pipeline.** The National Highway Traffic Safety Administration, established by the National Traffic and Motor Vehicle Safety Act of 1966, is the primary federal regulator of vehicle safety in the United States (Mashaw and Harfst, 1990). Within NHTSA, the Office of Defects Investigation (ODI) is responsible for identifying safety-related defects and, when warranted, compelling manufacturers to issue recalls.

The investigation pipeline proceeds through a structured sequence. Consumer complaints are filed through NHTSA's Vehicle Owner's Questionnaire system and screened by ODI analysts. When a pattern of complaints suggests a potential defect, ODI may open a *Preliminary Evaluation* (PE), which involves an initial assessment of the defect's safety significance, the number of vehicles affected, and the sufficiency of the evidence. If the PE reveals a potential safety defect, it may be escalated to an *Engineering Analysis* (EA), which involves more intensive technical investigation, testing, and manufacturer correspondence. At the conclusion of an EA, ODI may either close the investigation or trigger a manufacturer recall under 49 U.S.C. §30118–30120.

**The Capacity Constraint.** ODI has historically operated with limited staff. The 2023 DOT Inspector General audit reported that ODI employed approximately 90 investigators, a number that has remained largely stable over the past two decades even as the registered vehicle fleet grew from 225 million in 2000 to over 280 million by 2024 (U.S. Department of Transportation Office of Inspector General, 2023; Federal Highway Administration, 2024). The audit found that ODI failed to meet its own timeliness benchmarks across all investigation types and that "high-profile investigations can consume a disproportionate share of investigative resources, potentially delaying other safety-critical work."

The constraint is visible in the data. In my sample, the average investigation takes 261 days to close, with substantial variation (standard deviation of 303 days). The number of concurrent open investigations ranges from 2 to 156, with a mean of 70 and a standard deviation of 34. This variation in queue length—driven by the staggered timing of defect discoveries across manufacturers and vehicle systems—is the source of identification.

**The Stakes.** Between 2000 and 2024, NHTSA investigations led to 649 recalls in my sample, affecting millions of vehicles. Consumer complaints associated with these investigations document 315,128 injuries and 21,980 deaths. While not all of these outcomes are attributable to the investigated defects, they underscore the magnitude of the safety question: when an investigation that would have led to a recall is instead closed due to bandwidth constraints, the defect remains unaddressed in the vehicle fleet.

## 3. Data

The analysis uses three publicly available flat files from NHTSA's Office of Defects Investigation, covering all investigations, consumer complaints, and recall campaigns since the mid-1990s.

**Investigations.** The NHTSA Investigations flat file contains 153,998 investigation-make-model records, which I collapse to 4,254 unique PE and EA investigations. Each record includes the investigation identifier, opening and closing dates, manufacturer name, component description, and linked recall campaign number (if any). I restrict the analysis to investigations opened between 2000 and 2024 with non-missing closure dates, yielding 1,362 observations: 1,039 PEs and 323 EAs.

**Complaints.** The NHTSA Complaints flat file contains 2,190,915 consumer-filed safety complaints since January 1995. Each complaint includes crash, fire, injury, and death indicators. I aggregate complaints to the manufacturer-month level to construct pre-period severity measures (complaint volume, crashes, injuries, and deaths in the 12 months before each investigation opens) and during-period outcome measures.

**Recalls.** I combine the pre-2010 and post-2010 NHTSA Recalls flat files, yielding 320,104 unique recall campaigns. I link recalls to investigations using the campaign number field in the investigation file. Of the 1,362 investigations in the analysis sample, 649 (47.7 percent) are linked to at least one recall campaign.

**Sample Construction.** The raw investigation file contains 153,998 investigation-make-model records. Multiple records per investigation arise because a single investigation can cover multiple vehicle makes and models. Collapsing to unique investigation identifiers yields 4,254 PE and EA investigations. I restrict to investigations opened between 2000 and 2024 (ensuring a stable ODI organizational structure) and require non-missing closure dates (to observe investigation duration and recall outcomes), yielding the analysis sample of 1,362 investigations. The 2,892 excluded observations are primarily investigations still open at the time of data extraction or those opened before 2000.

**Key Variables.** The dependent variables are: (1) a binary recall indicator equal to one if the investigation led to a manufacturer recall; and (2) log investigation duration in days. The key independent variable is the count of concurrent open PE/EA investigations at other manufacturer groups on the date the focal investigation was opened. I standardize manufacturer names into groups (e.g., all GM brands into "General Motors") and compute the instrument by excluding investigations of the same manufacturer group. Controls include pre-period complaint volume, a pre-period severity score (weighted sum of crashes, injuries, and deaths), the number of vehicle models covered by the investigation, and manufacturer investigation frequency.

**Table 1:** Summary Statistics: NHTSA Safety Investigations, 2000–2024

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| Investigation duration (days) | 261.2 | 303.3 | 13 | 2596 |
| Log duration | 5.2 | 0.8 | 3 | 8 |
| Concurrent investigations (all) | 70.3 | 34.2 | 2 | 156 |
| Concurrent investigations (other mfr.) | 38.1 | 34.5 | 0 | 141 |
| Recall indicator | 0.5 | 0.5 | 0 | 1 |
| Pre-period complaints | 4342.6 | 5016.3 | 0 | 27851 |
| Pre-period severity score | 1293.4 | 1502.1 | 0 | 6419 |
| Injuries during investigation | 231.4 | 378.3 | 0 | 3749 |
| Deaths during investigation | 16.1 | 40.7 | 0 | 404 |
| Models covered | 31.2 | 291.2 | 1 | 7840 |

*Notes:* Sample includes 1,362 closed Preliminary Evaluation (PE) and Engineering Analysis (EA) investigations opened by NHTSA's Office of Defects Investigation between 2000 and 2024. Concurrent investigations count all PE/EA investigations open on the date the focal investigation was opened. Other-manufacturer concurrent investigations exclude investigations of the same manufacturer group. Pre-period complaints and severity scores are computed over the 12 months preceding investigation opening for the same manufacturer. Injuries and deaths are manufacturer-level complaint counts during the investigation period.

# 4. Empirical Strategy

## 4.1 Identification

The research question is whether exogenous variation in NHTSA's investigation workload causally affects the probability that a given investigation results in a recall. The key identifying assumption is that, conditional on defect characteristics and calendar time, the volume of concurrent open investigations at *other* manufacturers affects the focal investigation's outcome only through the examiner bandwidth channel.

Formally, I estimate the reduced-form relationship:

$$Y_i = \alpha + \beta \cdot \text{OtherQueue}_i + X_i'\gamma + \delta_c + \mu_t + \varepsilon_i \tag{1}$$

where $Y_i$ is either the recall indicator or log duration for investigation $i$; $\text{OtherQueue}_i$ is the count of concurrent open PE/EA investigations at other manufacturer groups on $i$'s opening date; $X_i$ is a vector of controls (pre-period complaints, severity score, number of models, manufacturer investigation frequency); $\delta_c$ are component category fixed effects; and $\mu_t$ are year fixed effects.

The coefficient $\beta$ captures the reduced-form effect of queue congestion on investigation outcomes. I report this as the preferred specification because it does not require a first-stage regression and is robust to weak-instrument concerns that arise with year-quarter fixed effects.

### 4.2 Identification Assumptions and Threats

The exclusion restriction requires that, conditional on controls, the timing and severity of defect discoveries at Ford, Toyota, or other manufacturers does not directly affect whether a Honda brake investigation results in a recall. Three threats merit discussion.

**Correlated Safety Shocks.** Economy-wide factors (e.g., driving patterns, vehicle age composition) could simultaneously increase complaints across manufacturers and affect investigation outcomes. Year fixed effects absorb common temporal shocks, and component category fixed effects control for systematic differences across vehicle systems (brakes, airbags, fuel systems, etc.). The results are robust to finer year-quarter fixed effects, though with some loss of precision.

**Manufacturer Strategic Behavior.** Manufacturers might delay cooperation with NHTSA during periods of high regulatory activity, hoping that congested investigators will be less aggressive. I control for manufacturer investigation frequency (a proxy for size and regulatory relationship) and show that results are stable when dropping any single major manufacturer.

**Composition Effects.** Busy regulatory periods might attract different types of investigations. The pre-period severity score and component category fixed effects address this concern. The placebo test—showing that high-severity investigations are less affected by queue congestion—provides further evidence that the effect operates through triage rather than compositional shifts.

## 5. Results

### 5.1 Main Results

Table 2 reports the reduced-form estimates. Panel A shows the effect on log investigation duration, and Panel B shows the effect on recall probability. Each column varies the fixed effect structure and standard error clustering.

**Table 2:** Reduced-Form Effects of Queue Congestion on Investigation Outcomes

| Dependent Variables: | log_duration | | | has_recall | | |
|---|---|---|---|---|---|---|
| | Year FE | Yr-Qtr FE | Clustered | Year FE | Yr-Qtr FE | Clustered |
| Model: | (1) | (2) | (3) | (4) | (5) | (6) |
| *Variables* | | | | | | |
| Other-mfr. queue | -0.0011** | -0.0009* | -0.0011** | -0.0014*** | -0.0013*** | -0.0014*** |
| | (0.0006) | (0.0006) | (0.0005) | (0.0004) | (0.0004) | (0.0004) |
| *Fixed-effects* | | | | | | |
| comp_cat | Yes | Yes | Yes | Yes | Yes | Yes |
| open_year | Yes | | Yes | Yes | | Yes |
| open_yq | | Yes | | | Yes | |
| *Fit statistics* | | | | | | |
| Observations | 1,362 | 1,360 | 1,362 | 1,362 | 1,360 | 1,362 |
| $R^2$ | 0.31320 | 0.35402 | 0.31320 | 0.08221 | 0.11808 | 0.08221 |
| Within $R^2$ | 0.05466 | 0.05560 | 0.05466 | 0.03993 | 0.04025 | 0.03993 |

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

The coefficient on the other-manufacturer queue is negative and statistically significant across all specifications for the recall probability outcome. In the preferred specification with year fixed effects and heteroskedasticity-robust standard errors (column 4), each additional concurrent investigation reduces the recall probability by 0.14 percentage points ($\beta = -0.0014$, $p < 0.001$). The estimate is virtually unchanged with year-quarter fixed effects (column 5: $\beta = -0.0013$) and with manufacturer-clustered standard errors (column 6: $\beta = -0.0014$, $p < 0.01$).

The magnitude is economically meaningful. A one-standard-deviation increase in the other-manufacturer queue (34.5 investigations) reduces the recall probability by 4.8 percentage points, or roughly 10 percent of the mean recall rate. Over the 1,362 investigations in the sample, this implies approximately 65 forgone recalls attributable to queue congestion.

The duration results (columns 1–3) show a negative effect of queue congestion on log investigation duration: busier periods are associated with shorter investigations. This is consistent with the triage mechanism—investigators close cases more quickly when the queue is long, and some of those closures are without recall. The coefficient of $-0.0011$ (column 1) implies that a one-standard-deviation increase in the queue reduces investigation duration by approximately 3.8 percent.

### 5.2 Mechanism: Triage at the Preliminary Evaluation Stage

If congestion operates through triage, its effects should be concentrated at the decision node where investigators choose whether to escalate or close a case. Table 3 decomposes the reduced-form effect by investigation type and defect severity.

**Table 3:** Heterogeneity: Investigation Type and Defect Severity

| Dependent Variables: | has_recall | | | | log_duration | |
|---|---|---|---|---|---|---|
| | PE | EA | Low Sev. | High Sev. | PE | EA |
| Model: | (1) | (2) | (3) | (4) | (5) | (6) |
| *Variables* | | | | | | |
| Other-mfr. queue | -0.0013*** | -0.0005 | -0.0011** | -0.0018** | -0.0004 | 0.0002 |
| | (0.0004) | (0.0008) | (0.0005) | (0.0008) | (0.0005) | (0.0011) |
| *Fixed-effects* | | | | | | |
| comp_cat | Yes | Yes | Yes | Yes | Yes | Yes |
| open_year | Yes | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | | |
| Observations | 1,039 | 323 | 1,019 | 342 | 1,039 | 323 |
| $R^2$ | 0.12961 | 0.12781 | 0.09039 | 0.14770 | 0.47098 | 0.38578 |

*Heteroskedasticity-robust standard-errors in parentheses*
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The recall probability effect is concentrated in Preliminary Evaluations (column 1: $\beta = -0.0013$, $p < 0.01$) and not statistically significant for Engineering Analyses (column 2: $\beta = -0.0005$), though the EA coefficient is directionally consistent. This is consistent with triage operating at the PE stage, where the decision to escalate or close is most discretionary. Once an investigation has been escalated to EA status, it has already survived the initial screen and is more likely to proceed to a recall regardless of queue length.

The severity split reveals a nuanced pattern. For recall probability, the effect is present across severity levels: low-severity investigations show $\beta = -0.0011$ ($p < 0.05$, column 3) and high-severity cases show $\beta = -0.0018$ ($p < 0.05$, column 4), though the latter is estimated on a substantially smaller sample (342 vs. 1,019 investigations). The presence of a congestion effect even among high-severity cases suggests that bandwidth constraints affect the regulatory pipeline broadly, not only through selective triage of low-priority cases. The duration results provide a sharper contrast: congestion reduces investigation duration for low-severity cases (column 5: $\beta = -0.0012$, $p < 0.10$) but has no effect for high-severity cases (column 6: $\beta = -0.0004$, not significant). This pattern is consistent with investigators compressing the processing time for lower-priority cases under congestion while maintaining thoroughness for serious defects—even if the ultimate recall decision is affected across severity levels.

### 5.3 Robustness

**Leave-One-Out Manufacturers.** Table 4 shows that the reduced-form estimates are stable when dropping any single major manufacturer group. The recall probability coefficient ranges from $-0.0013$ to $-0.0016$ across the eight drops, always within one standard error of the full-sample estimate. No single manufacturer drives the result.

**Table 4:** Leave-One-Out Manufacturer Robustness

| | Log Duration | | Recall Prob. | | |
|---|---|---|---|---|---|
| Manufacturer Dropped | Coeff. | SE | Coeff. | SE | $N$ |
| FORD | -0.0013 | (0.0006) | -0.0016 | (0.0004) | 1,197 |
| GENERAL MOTORS | -0.0011 | (0.0006) | -0.0013 | (0.0004) | 1,189 |
| CHRYSLER/STELLANTIS | -0.0011 | (0.0006) | -0.0013 | (0.0004) | 1,229 |
| TOYOTA | -0.0012 | (0.0006) | -0.0013 | (0.0004) | 1,319 |
| HONDA | -0.0012 | (0.0006) | -0.0013 | (0.0004) | 1,307 |
| NISSAN | -0.0011 | (0.0006) | -0.0014 | (0.0004) | 1,313 |
| HYUNDAI/KIA | -0.0012 | (0.0006) | -0.0014 | (0.0004) | 1,289 |
| TESLA | -0.0011 | (0.0006) | -0.0014 | (0.0004) | 1,351 |
| Full Sample | -0.0011 | (0.0006) | -0.0014 | (0.0004) | 1,362 |

*Notes:* Each row re-estimates the reduced-form specification from Table 2 after dropping all investigations involving the named manufacturer group. All specifications include component category and year fixed effects, with heteroskedasticity-robust standard errors. Coefficient shown is on concurrent other-manufacturer investigations.

**Alternative Clustering.** The baseline specification uses heteroskedasticity-robust standard errors. Clustering by manufacturer group, by year-quarter, or two-way by both yields similar or smaller standard errors, with the recall probability coefficient remaining significant at the 1 percent level in all cases (manufacturer-clustered: $p < 0.01$; year-quarter-clustered: $p < 0.001$; two-way: $p < 0.01$).

**Instrumental Variable Estimates.** Table 5 in the appendix reports IV estimates that instrument total concurrent investigations with the other-manufacturer queue. The first-stage coefficient is 0.24 ($p < 0.001$) with a Kleibergen-Paap F-statistic of 83.0, well above the Stock and Yogo (2002) threshold of 10. The 2SLS estimates are larger than the reduced-form: each additional concurrent investigation reduces log duration by 0.0047 ($p < 0.05$) and recall probability by 0.57 percentage points ($p < 0.001$). The IV magnification is consistent with the reduced-form attenuating the effect by averaging over investigations with heterogeneous sensitivity to congestion. I present the reduced-form as the primary specification because it avoids functional-form assumptions and provides a transparent lower bound, but the IV confirms the direction and statistical significance of the causal effect.

## 6. Discussion

The results reveal a systematic "triage tax" in vehicle safety regulation: when NHTSA's investigation queue lengthens, lower-priority defects are less likely to be recalled. This finding has three implications.

First, regulatory capacity is not merely an operational concern but a determinant of enforcement outcomes with direct safety consequences. The reduced-form estimate implies that a sustained reduction of 10 concurrent investigations—achievable with modest staffing increases—would raise the recall rate by approximately 1.4 percentage points, potentially averting dozens of recalls over a multi-year period. Given that the median recall affects approximately 100,000 vehicles, even a small increase in the recall rate translates into substantial consumer protection.

A rough welfare calculation illustrates the magnitude. The 65 forgone recalls implied by the reduced-form estimates each affect a median of approximately 100,000 vehicles. Using the Department of Transportation's Value of a Statistical Life ($12.5 million) and conservative assumptions about defect-related fatality rates from the complaint data (approximately 0.5 deaths per 100,000 affected vehicles), the implied safety cost of queue congestion is on the order of $400 million over the sample period, or roughly $16 million per year. This estimate is speculative and depends on the counterfactual recall completion rate, but it suggests that the triage tax is not merely a bureaucratic inconvenience.

Second, the triage mechanism suggests that regulatory capacity constraints do not affect all cases equally. Consistent with rational prioritization, investigators maintain enforcement intensity for high-severity defects while allowing lower-severity cases to bear the cost of congestion. This pattern is efficient *ex ante* but implies that the marginal forgone recall—the one that would have occurred with slightly more capacity—involves a real but lower-profile safety defect.

Third, the finding that congestion *shortens* investigations while reducing recall probability suggests that the margin of adjustment is not delay but abandonment. Investigators do not take longer to reach the same conclusion; they reach a different conclusion faster. This distinguishes the mechanism from simple delay costs and has implications for how regulatory agencies should measure the cost of capacity constraints.

**Limitations.** Several caveats apply. The instrument captures variation in queue length driven by the timing of defect discoveries at other manufacturers, but some residual correlation between queue length and the regulatory environment may remain despite year fixed effects. The first-stage F-statistic for the IV specification falls below conventional thresholds, motivating the reduced-form approach but limiting the structural interpretation. The analysis uses manufacturer-level complaint aggregates rather than investigation-specific complaint data, introducing measurement error in the severity controls. Finally, the outcome is recall initiation, not the downstream effect on vehicle repairs or injuries—the welfare calculation requires additional assumptions about recall completion rates and defect-related accident risks.

## 7. Conclusion

Ninety investigators cannot inspect 280 million vehicles. This paper shows that the resulting queue congestion has a measurable cost: each additional concurrent investigation at other manufacturers reduces the probability that a new NHTSA safety investigation leads to a recall by 0.14 percentage points. The effect operates through triage at the Preliminary Evaluation stage, where bandwidth-constrained investigators close lower-priority cases more quickly and with less likelihood of escalation. The "triage tax"—the enforcement actions forgone because the queue was too long—represents a previously unmeasured cost of fixed regulatory capacity in a setting where the stakes are vehicle safety and human life.

## 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


**Bae, Youngsoo and Hugo Benitez-Silva**, "Do Consumers Use Event Studies? Evidence from Auto Recalls," *Working Paper*, 2004.

**Barber, Brad M and Masako N Darrough**, "Product Reliability and Firm Value: The Experience of American and Japanese Automakers, 1973–1992," *Journal of Political Economy*, 1996, *104* (5), 1084–1099.

**Dobbie, Will, Jacob Goldin, and Crystal S Yang**, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, 2018, *108* (2), 201–240.

**Federal Highway Administration**, "Highway Statistics," Technical Report, U.S. Department of Transportation 2024.

**Heese, Jonas**, "The SEC's Revolving Door: Is It a Problem?," *Working Paper*, 2016.

**Hoffer, George E, Stephen W Pruitt, and Robert J Reilly**, "The Impact of Product Recalls on the Wealth of Sellers: A Reexamination," *Journal of Political Economy*, 1988, *96* (3), 663–670.

**Jarrell, Gregg and Sam Peltzman**, "The Wealth Effects of Product Recalls on Shareholders and Consumers: Evidence from the Automobile Industry," *Journal of Political Economy*, 1985, *93* (3), 512–536.

**Kang, Woo-Sung, Alan Kilgore, and Sue Wright**, "Are Audit Committees Created Equal? The Effect of Mandatory Audit Committee Establishment on Audit Quality," *Contemporary Accounting Research*, 2016, *33* (4), 1542–1574.

**Kling, Jeffrey R**, "Incarceration Length, Employment, and Earnings," *American Economic Review*, 2006, *96* (3), 863–876.

**Levitt, Steven D and Jack Porter**, "How Dangerous Are Drinking Drivers?," *Journal of Political Economy*, 2001, *109* (6), 1198–1237.

**Liu, Yong**, "Production Recall and Liability: Theory and Evidence from the Auto Industry," *Journal of Industrial Economics*, 2016, *64* (3), 478–508.

**Macher, Jeffrey T and John W Mayo**, "Pharmaceutical Regulation in the United States and the European Union: Is One 'Better' Than the Other?," *Journal of Regulatory Economics*, 2007, *32* (1), 1–15.

**Mashaw, Jerry L and David L Harfst**, "The Struggle for Auto Safety," *Harvard University Press*, 1990.

**Peltzman, Sam**, "The Effects of Automobile Safety Regulation," *Journal of Political Economy*, 1975, *83* (4), 677–726.

**Rigbi, Oren**, "The Effects of Usury Laws: Evidence from the Online Loan Market," *Review of Financial Studies*, 2013, *26* (5), 1145–1175.

**Sampat, Bhaven and Heidi L Williams**, "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome," *American Economic Review*, 2019, *109* (1), 203–236.

**Stock, James H and Motohiro Yogo**, "Testing for Weak Instruments in Linear IV Regression," *National Bureau of Economic Research Technical Working Paper*, 2002, (284).

**U.S. Department of Transportation Office of Inspector General**, "NHTSA's Office of Defects Investigation Needs to Improve Its Defect Investigation Process," Technical Report ST2024006, DOT OIG 2023.

**Valukas, Anton R**, "Report to Board of Directors of General Motors Company Regarding Ignition Switch Recalls," Technical Report, Jenner & Block 2014.

## A. IV Estimates

**Table 5:** Instrumental Variable Estimates: Queue Congestion and Investigation Outcomes

| Dependent Variables: | Total queue (instrumented) | log_duration | has_recall |
|---|---|---|---|
| | First Stage | 2SLS: Duration | 2SLS: Recall |
| Model: | (1) | (2) | (3) |
| *Variables* | | | |
| Other-mfr. queue | 0.2401*** | | |
| | (0.0288) | | |
| comp_pre | -0.0007 | $-1.55 \times 10^{-5}$ | $4.47 \times 10^{-6}$ |
| | (0.0004) | $(1.01 \times 10^{-5})$ | $(7.67 \times 10^{-6})$ |
| severity_pre | 0.0010 | $2.33 \times 10^{-6}$ | $1.19 \times 10^{-5}$ |
| | (0.0014) | $(2.93 \times 10^{-5})$ | $(2.23 \times 10^{-5})$ |
| n_models | -0.0119*** | $4.56 \times 10^{-5}$ | $5.7 \times 10^{-5}$* |
| | (0.0032) | $(5.18 \times 10^{-5})$ | $(3.06 \times 10^{-5})$ |
| mfr_inv_count | 0.0511 | 0.0033*** | -0.0016*** |
| | (0.0364) | (0.0008) | (0.0006) |
| Total queue (instrumented) | | -0.0047** | -0.0057*** |
| | | (0.0024) | (0.0017) |
| *Fixed-effects* | | | |
| comp_cat | Yes | Yes | Yes |
| open_year | Yes | Yes | Yes |
| *Fit statistics* | | | |
| Observations | 1,362 | 1,362 | 1,362 |
| $R^2$ | 0.11210 | 0.31320 | 0.08221 |
| F-test (1st stage), Total queue (instrumented) | | 82.958 | 82.958 |

*Heteroskedasticity-robust standard-errors in parentheses*
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

## B. Standardized Effect Sizes

**Table 6:** Standardized Effect Sizes

| Outcome | $\hat{\beta}$ | SE | SD($Y$) | SDE | SE(SDE) | Classification |
|---|---|---|---|---|---|---|
| *Panel A: Pooled* | | | | | | |
| Recall probability | -0.0014 | (0.0004) | 0.500 | -0.0948 | (0.0269) | Moderate negative |
| Log duration | -0.0011 | (0.0006) | 0.790 | -0.0495 | (0.0244) | Small negative |
| *Panel B: Heterogeneous* | | | | | | |
| Recall prob. (PE only) | -0.0013 | (0.0004) | 0.496 | -0.0884 | (0.0306) | Moderate negative |
| Recall prob. (low sev.) | -0.0011 | (0.0005) | 0.500 | -0.0775 | (0.0319) | Moderate negative |

*Notes:* **Country:** United States. **Research question:** Does congestion in the NHTSA safety investigation queue—measured by concurrent open investigations at other manufacturers—reduce the probability that a new investigation results in a vehicle recall? **Policy mechanism:** NHTSA's Office of Defects Investigation operates with a fixed pool of approximately 90 investigators responsible for 280+ million registered vehicles. When multiple investigations are open simultaneously, each receives less examiner bandwidth, creating a triage system where lower-profile defects are more likely to be closed without recall. **Outcome definition:** Binary recall indicator (1 if investigation led to a manufacturer recall campaign, 0 otherwise); log investigation duration (days from opening to closure). **Treatment:** Continuous—count of concurrent open PE/EA investigations at other manufacturer groups on the date the focal investigation was opened (mean 38.1, SD 34.5). **Data:** NHTSA Office of Defects Investigation flat files (investigations, complaints, recalls), 2000–2024, investigation-level, 1,362 observations. **Method:** Reduced-form OLS with component category and year fixed effects; heteroskedasticity-robust standard errors. **Sample:** Closed PE and EA investigations opened 2000–2024; excludes open investigations and non-defect investigation types (AQ, RQ, CI). SDE = $\hat{\beta} \times \text{SD}(X)/\text{SD}(Y)$ where SD($X$) is the cross-sectional standard deviation of the instrument and SD($Y$) is the unconditional standard deviation of the outcome. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null ($< 0.005$).