A CHECK ENGINE LIGHT FOR THE AUTOMOTIVE INDUSTRY

Detecting Automotive Failures from Consumer Complaints and Online Forums

Tommy Jones[1,2], Likhitha Kempegowda[1], Spencer Krivo[1], Adrian Ng[2]
[1]DrivePulse
[2]Georgetown University

ABSTRACT

We develop an unsupervised statistical framework for early detection of vehicle safety issues from consumer complaint data. Our approach applies a Cohen's d-based indicator with logarithmic scaling to NHTSA complaints and online forum discussions, identifying when complaint rates for specific vehicle-component combinations exceed historical norms through hypothesis testing on rolling 180-day windows. This method has generated 15,811 detected events: 15,490 from NHTSA complaints and 321 from Mechanics Stack Exchange and CarTalk forums. Automated summarization using large language models produces human-interpretable event descriptions from aggregated complaint narratives.

We validate the approach using the GM ignition switch recalls as a case study. Our system detected pre-recall events for 19 of 24 recalled vehicles (79%) with a mean 7.43-year lead time. Critically, 62% of these events occurred during 2001-2007 when GM was documented to be internally aware of the defect but had not acted, demonstrating that statistical complaint monitoring can extract genuine early warning signals with temporal alignment to manufacturer knowledge rather than merely correlating with recall announcements.

Descriptive analysis of events across component categories from 2000-2025 reveals patterns consistent with known industry shifts including ADAS adoption, vehicle electrification, and software integration. Forum data contributed minimally (321 events) due to limited post volume in the selected forums, though expanded manufacturer-specific forum coverage is in progress.

The GM case provides proof-of-concept for complaint-based early detection. However, comprehensive validation faces fundamental challenges: our unsupervised approach lacks ground truth labels for most events, many recalls occur without complaint spikes, and many complaint events never result in recalls. This work establishes feasibility and a methodological foundation while acknowledging that broader validation and integration with additional data sources remain critical next steps.

Keywords: Consumer complaint analysis, defect detection, early warning systems, statistical anomaly detection, post-market surveillance, NHTSA data mining, unsupervised machine learning, automotive recalls, real-world safety data

INTRODUCTION

Motivation

The composition of the U.S. vehicle fleet has been undergoing a sustained shift toward older vehicles, with important implications for safety monitoring and defect detection. Over the past several decades, the average age of passenger vehicles on U.S. roads has steadily increased, reflecting both improvements in vehicle durability and changing economic conditions [1]. As vehicles remain in service longer, a growing share of the fleet consists of cars that are well beyond their original warranty periods and no longer subject to routine manufacturer oversight.

At the same time, the cost of acquiring new vehicles has risen substantially. Average transaction prices for new automobiles have increased in real terms over recent years, driven by higher production costs, increased vehicle complexity, and evolving regulatory requirements [2], [3]. These trends have led many consumers to retain vehicles for longer periods or turn to the used vehicle market, further increasing the prevalence of aging vehicles in everyday operation.

Recent trade and industrial policies have contributed additional upward pressure on vehicle prices. Tariffs on imported vehicles and automotive components can raise production costs for domestically assembled vehicles or increase the retail prices of imported models [3], [4]. While the magnitude of these effects continues to be debated, their net impact reinforces incentives for consumers to delay vehicle replacement, extending vehicle lifespans across the fleet.

These structural shifts matter because many vehicle safety and quality defects do not manifest immediately after production. Instead, defects often emerge years after vehicles enter service, as components age, accumulate wear, or are exposed to diverse operating environments. Analysis of NHTSA's recall database shows that many recalls involve aging vehicles rather than newly produced models. Since 1996, 30.6% of recalled vehicles were at least four years old at the time of recall initiation, and 18.8% were six years old or older [5]. In many cases, vehicles of this age fall outside manufacturer warranty periods, reducing opportunities for systematic warranty-based monitoring.

NHTSA employs a structured, expert-driven investigation process informed by complaint data, technical analysis, and case-specific judgment. However, this process is not designed as a formal, statistical early-warning system operating continuously across all makes, models, and model years in real time [6]. As a result, subtle or slowly emerging safety and quality issues—particularly those affecting aging vehicles—may be difficult to detect until they reach a scale that warrants regulatory or manufacturer action.

In parallel with NHTSA's formal processes, a variety of non-regulatory mechanisms exist for monitoring emerging vehicle defects. Consumer-facing platforms such as *CarComplaints.com* aggregate and visualize owner-reported issues by make, model, model year, and component, often highlighting recurring problems based on complaint volumes and user severity ratings [7]. While these platforms can provide timely and intuitive signals to consumers, they primarily rely on raw or heuristically normalized complaint counts and are not designed as statistically grounded early-warning systems operating across heterogeneous data sources.

In the academic literature, a growing body of work has explored predictive modeling approaches for defect detection using consumer complaints, warranty proxies, and online discussion forums, applying methods ranging from topic models and support vector machines to gradient boosting and neural language models (e.g., [8], [9], [10]). Although these studies demonstrate that safety and quality outcomes can often be identified retrospectively, they are typically evaluated offline, focus on limited subsets of the vehicle market, and are not deployed as continuous, time-aware monitoring systems.

Post-market vehicle safety surveillance in the United States is therefore inherently fragmented. It relies on a combination of consumer complaints, manufacturer communications such as technical service bulletins, regulatory investigations, and recall actions—many of which are initiated only after potential issues have become widespread or severe. While these mechanisms play a critical role in ensuring vehicle safety, they are largely reactive and are not designed to provide continuous, quantitative early warning signals across the full vehicle fleet.

Together, these trends highlight a growing gap between the evolving structure of the vehicle fleet and the tools currently available for post-market safety monitoring. Addressing this gap requires approaches that can systematically analyze heterogeneous data sources over time and identify emerging risks before they culminate in large-scale harm or costly recalls.

Contributions

This work makes three primary contributions. First, we develop and demonstrate an unsupervised statistical framework for detecting vehicle safety events from consumer complaints that operates continuously across the full vehicle fleet without requiring labeled training data or manual configuration per vehicle model. Second, we provide empirical validation through the GM ignition switch case study, demonstrating that complaint-based surveillance can detect safety issues years before regulatory action (mean 7.43-year lead time) with temporal alignment to manufacturer knowledge timelines rather than mere correlation with recall announcements. Third, we establish a scalable methodological pipeline integrating NHTSA complaints with online forum data, generating interpretable event descriptions from 15,811 detected anomalies spanning nearly three decades of vehicle safety data. While comprehensive system validation remains an open challenge due to the unsupervised nature of the approach and asymmetric relationships between complaints and recalls, these contributions establish the feasibility of statistical complaint monitoring as a complement to existing expert-driven surveillance processes.

LITERATURE REVIEW

Prior research on post-market vehicle safety and defect detection spans multiple disciplines, including transportation safety, operations management, text mining, and machine learning. This literature can be broadly grouped into four overlapping areas: (1) societal and economic impacts of vehicle defects, (2) regulatory recall analysis and monitoring approaches, (3) studies using social media and other consumer-generated text, and (4) predictive modeling approaches for defect detection and recall prediction for automotive and non-automotive products.

Societal and Economic Impacts of Vehicle Defects

Vehicle defects impose substantial societal and economic costs. Defects in key systems can lead to severe injury or fatalities [11]. Vehicle recalls are associated with statistically significant negative stock market reactions, with average losses on the order of tens of millions of dollars per recall event [12]. These impacts are amplified when investigations are prolonged or when defects involve complex supplier relationships. From a consumer perspective, recall inconvenience, delayed repairs, and low completion rates—particularly for older vehicles—further underscore the importance of earlier and more effective defect detection.

Regulatory Recall Analysis

NHTSA's Office of Defects Investigation (ODI) employs a structured, risk-based process for evaluating potential vehicle safety defects [6]. This process integrates consumer complaints, manufacturer information, and technical analysis to determine whether formal investigations or recalls are warranted. ODI investigations may proceed through multiple stages and often result in manufacturer-initiated recalls rather than regulator-initiated action.

Research has also highlighted that a large fraction of recalls are administrative or compliance-related rather than driven by newly discovered safety defects [13]. Other work has analyzed large automakers' recall timing decisions. One paper finds that after an initial report of suspected quality failures and found that discovery-to-recall is longer when recall reports are external to the automaker, fault attribution is directed to the supplier rather than the automaker itself, the source of defect is design-related rather than due to manufacturing, and the defect in question affects more models, rather than fewer [14]. Astvansh et al. developed curated data sets from unstructured recall documentation to further study recall decision-making and timing [15]. A simulation study finds that, among other things, a large volume of complaints actually prolongs the timing of discovery-to-recall [16].

Notably, although not all defects result in investigations or recalls, there does not appear to be much research into defects where there is no regulatory action.

Research Using Social Media and Consumer-Generated Text

Parallel to work using official data, a growing body of research has explored the use of social media, online forums, and consumer reviews to detect product defects. In the automotive domain, several studies have shown that forum posts and online discussions can surface defect-related information prior to formal recalls [10].

Abrahams et al. demonstrated that text classification methods can successfully isolate automotive components discussed in online forums, enabling downstream defect analysis [9]. Subsequent work extended these ideas to defect discovery across multiple industries, emphasizing the limited utility of sentiment analysis and the importance of domain-specific language. Chakraborty and Gupta analyze non-regulatory complaint data from *carcomplaints.com* [17] alongside NHTSA complaint data to predict recalls. Maniyur et al. combine scraped forum data with NHTSA complaints, as does this study, with a focus only on one automotive brand [18].

Despite these advances, the literature using social media data remains fragmented. Most studies focus on a single manufacturer or platform, rely on static data sets, or treat forum signals independently from regulatory data. Integrated analyses combining social and official data sources remain rare.

Predictive Modeling and Machine Learning Approaches

Much work uses exploratory text mining and clustering methods to identify latent themes in complaint narratives and to characterize common failure modes [19] [11] [9] [20] [21] [22] [23]. Others have applied machine learning models to predict vehicle recalls or defect risk using consumer complaints and related data. Examples include using support vector machines [8], random forest [20], and gradient boosting [24] [25]. More recent work has employed the use of sophisticated language models for these tasks [17] [26].

Some studies report high predictive power to forecast recalls. However, we are skeptical of these results. Evaluation methodologies often rely on random train–test splits that fail to respect the strong temporal dependencies inherent in post-market safety data. Failing to account for temporal dependencies can result in data from the "future" being used to predict data from the "past". Attempts to contact the corresponding authors of these studies were not successful as we did not receive a response.

Gaps in the Existing Literature

Across these bodies of work, several gaps are evident:

1. **Limited integration across data sources**: Most studies analyze either regulatory data or social media data in isolation.
2. **Lack of statistical event detection**: Few approaches focus on identifying sustained abnormal behavior in complaint activity as opposed to directly predicting regulatory action.
3. **Insufficient treatment of temporal structure**: Many predictive models ignore the time-series nature of complaints, investigations, and recalls, resulting in overly-optimistic estimates of predictive accuracy.
4. **Neglect of the data-generating process**: The structured progression from complaints to investigations to recalls is rarely modeled explicitly.
5. **Absence of continuous monitoring frameworks**: Existing studies are typically retrospective and not designed for ongoing surveillance.

The present study directly addresses gaps in statistical event detection, absence of continuous monitoring frameworks, and treatment of temporal structure while partially addressing integration across data sources, and neglect for the data-generating process. We propose a statistically grounded, event-based framework for early detection of emerging vehicle safety and quality issues which respects the temporal nature of the data. By integrating regulatory data with consumer-generated text and focusing on sustained deviations from historical norms, this work complements existing predictive and descriptive approaches while aligning more closely with the realities of post-market safety monitoring.

DATA AND METHODS

Data Sources

Our analysis draws on multiple complementary data sources capturing post-market vehicle safety signals from both regulatory and consumer perspectives. We use data from NHTSA's Office of Defects Investigation (ODI) and forum data from Mechnanics Stack Exchange and the CarTalk Community Forums.

NHTSA Office of Defects Investigation Databases We use data from the National Highway Traffic Safety Administration's (NHTSA) Office of Defects Investigation (ODI), which maintains detailed records related to post-market vehicle safety. While this analysis primarily uses complaint and recall data, ODI has four relevant databases to this work: consumer complaints, investigations, recalls, and manufacturer communications (commonly referred to as technical service bulletins, or TSBs). Together, these data sources capture different stages of the post-market safety life cycle, from early consumer-reported concerns to formal regulatory and manufacturer responses. All ODI data were accessed via NHTSA's public data sets and APIs portal [5].

The **consumer complaints** database consists of free-text narratives submitted voluntarily by vehicle owners, along with structured metadata describing the affected vehicle (e.g., make, model, model year, and component) and reported outcomes such as injuries or fatalities. Complaints serve as an early, but noisy, signal of potential safety or quality issues.

The **investigations** database documents formal defect investigations initiated by NHTSA. These records include investigation start and close dates, affected components, investigation type (e.g., preliminary evaluation, engineering analysis), and narrative summaries describing the suspected defect and scope of concern.

The **recalls** database records safety-related defect and noncompliance recalls. Each recall includes information on the initiating party (manufacturer or regulator), affected vehicles, component categories, recall scope, and timing. Importantly, not all recalls are associated with prior investigations, and many recalls are initiated voluntarily by manufacturers.

The **manufacturer communications (TSBs)** database captures technical service bulletins and related communications submitted to NHTSA. These documents often address known issues, repair procedures, or diagnostic guidance and may precede, accompany, or never result in formal regulatory action.

Across the ODI tables, we summarize scale, temporal coverage, and coverage of manufacturers, models, and components in Table 1.

Table 1.
Summary statistics of NHTSA ODI data

Metric	**Complaints**	**Recalls**	**Technical Service Bulletins**	**Investigations**
Records	2,052,325	264,122	5,170,041	153,556
Min date	1995-01-03	1979-10-12	1995-01-05	1972-03-10
Max date	2025-10-21	2025-10-22	2025-10-22	2025-10-21
Makes	1,221	1,741	789	645
Models	6,248	12,961	10,363	3,670
Model years	74	80	75	64
Components	751	623	757	416
Component groups (L1)	43	33	39	33

Online Automotive Repair Forums To complement regulatory data, we incorporate consumer-generated content from online automotive repair communities. These sources capture informal, real-time discussions of vehicle problems, diagnostics, and repair experiences that may not be reported through official channels. Specifically, we draw on two forums: *Mechanics Stack Exchange*, a question-and-answer platform focused on automotive diagnostics and repair and *The CarTalk Community Forum*, restricted to posts tagged under *Maintenance/Repair* and *Safety*.

Unlike ODI data, forum posts are unstructured and do not natively include standardized vehicle metadata. To enable integration with the regulatory data pipeline, forum posts are processed using supervised natural language processing classifiers that identify vehicle make, model, model year, component, and relevance to potential defects. Once classified, forum posts are treated analogously to complaints for the purposes of indicator construction and event detection. Descriptive statistics for forum data are reported in Table 2.

Table 2.
Summary statistics of social forum data

Metric	**Mechanics Stack Exchange**	**Cartalk Forums**
Posts	21,144	3,528
Min. date	2010-11-26	2007-05-31
Max. date	2025-07-01	2025-07-03
Makes	92	31
Models	2,571	1,037
Components	30	17

<u>Derived Complaint-Based Event Data Set</u> In addition to the raw ODI tables and forum posts, we analyze derived event data sets constructed from consumer complaint data and forum posts related to mechnanical questions or complaints. Events represent periods during which complaint volumes for a specific make–model–component configuration exceed historical norms, as identified by a statistical event-detection procedure described in the next section. Each event is characterized by a start date, end date, duration, and associated complaint volume, and is linked to a narrative summary synthesized from the underlying complaint text.

For each event, we compute the total number of complaints observed from one year prior to event initiation through the event's end date. This window reflects the trailing history used to construct the indicator time series on which event detection is performed and provides a consistent measure of complaint intensity associated with each detected event.

The event data set spans a wide range of manufacturers, models, and components, and captures both short-lived anomalies and more persistent safety or quality concerns. Summaries of the event data report the total number of detected events, their temporal distribution, typical durations, and the distribution of complaint volumes per event. These descriptive statistics provide important context for interpreting downstream analyses relating events to investigations, recalls, and case studies. Descriptive statistics for event data are reported in Table 3.

Table 3.
Summary statistics of detected events from NHTSA ODI complaints and social forum data

Metric	NHTSA Complaint Events	Forum Complaint Events
Events	15,490	321
min_start_date	1996-06-29	2012-05-23
max_start_date	2025-10-19	2025-06-22
Makes	176	23
Models	1,882	88
Components	82	19
Avg. event length (days)	798	503
Median event length (days)	448	406
Mean complaints	57	12
Median complaints	8	5

Event Detection Framework

<u>Overview</u> The objective of the proposed framework is to identify **statistically abnormal increases in complaints** for specific vehicle–component combinations relative to historical norms. These abnormalities are interpreted as potential early warning signals of emerging safety or reliability issues.

The framework consists of four main stages:

1. Normalization of complaint activity across vehicles and components
2. Construction of a continuous complaint indicator over time
3. Statistical detection of sustained abnormal behavior
4. Event-level summarization and filtering

This structure allows heterogeneous complaint data—spanning vehicles produced in vastly different volumes—to be placed on a common, interpretable scale.

<u>Normalizing Complaint Rates Across Vehicles</u> Raw complaint counts are not directly comparable across vehicles due to large differences in fleet size and usage. For example, high-volume vehicles will naturally generate more complaints even in the absence of elevated risk. To address this, we compare proportional complaint rates rather than counts.

Let:

- $C_{i,j,t}$ denote the number of complaints for vehicle i and component j in the one-year period ending at time t
- $P_{i,j,t}$ denote the proportion of complaints for component j attributable to vehicle i:

$$P_{i,j,t} = \frac{C_{i,j,t}}{\sum_i C_{i,j,t}}$$

- $P_{j,t}$ denote the overall proportion of complaints related to component j across all vehicles:

$$P_{j,t} = \frac{\sum_i C_{i,j,t}}{\sum_j \sum_i C_{i,j,t}}$$

These quantities allow us to compare a vehicle's component-specific complaint rate to a global baseline for that component. For example, Figure 1 compares the rolling 12-month number of complaints for engine and engine cooling issues of the Honda Accord and Land Rover Discovery from the NHTSA complaints database. Taken at face value, it would appear that the engine of the Honda Accord may be more prone to failure. However, Figure 2 compares engine complaints for these two vehicles using our complaint indicator, described below. Once normalized, the Honda Accord's engine seems fairly reliable whereas there is clearly a jump associated with an engine issue on the Land Rover Discovery.

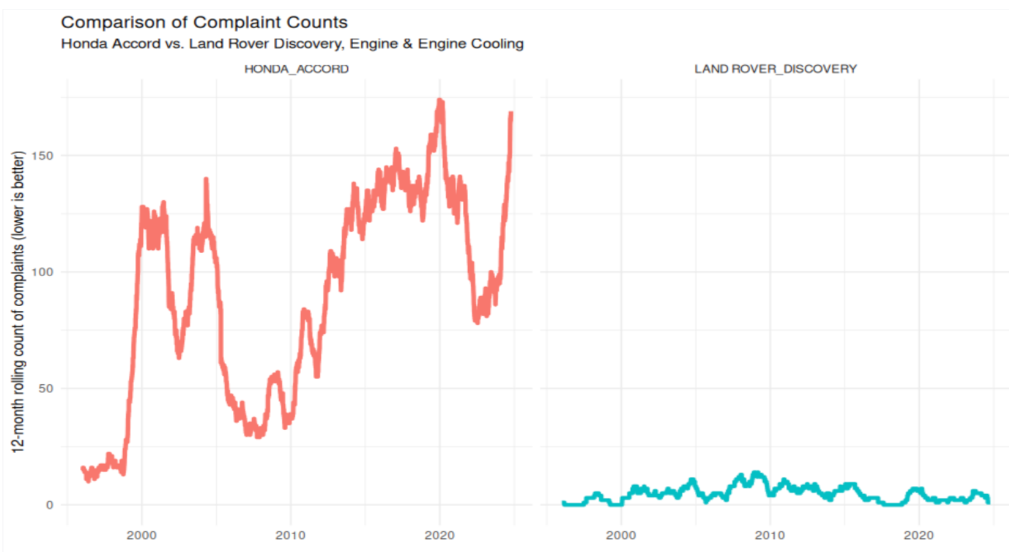

***Figure 1. The Honda Accord has more engine complaints than the Land Rover Discovery*.**

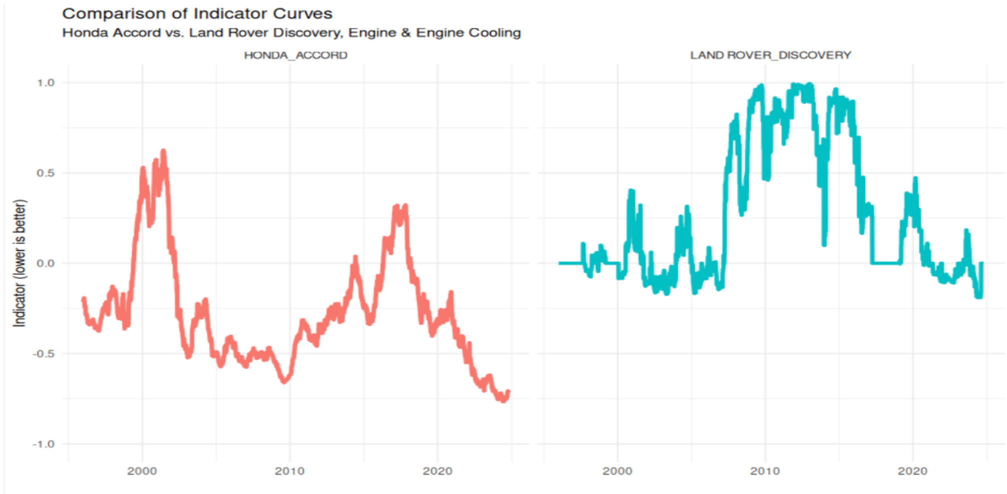

Figure 2. The D-Indicator normalizes for production volume, enabling defect detection.

Construction of the Complaint Indicator *D-Indicator***:** We define a continuous indicator, denoted $\delta_{i,j,t}$, which measures the degree to which complaints for vehicle i and component j exceed historical norms at time t.

The indicator is based on a scaled effect-size formulation:

$$\delta_{i,j,t} = \log(C_{i,j,t}) \cdot D(P_{i,j,t}, P_{j,t})$$

where $D(\cdot,\cdot)$ denotes a Cohen's d-style standardized difference between proportions.

This formulation balances two competing considerations:

- **Effect magnitude**: Large deviations from the baseline proportion should be emphasized
- **Statistical support**: Extremely small sample sizes should not generate large signals

Using raw t-statistics would overstate significance in large data sets, while unscaled effect sizes can exaggerate sparse data. The logarithmic scaling of complaint counts provides a practical compromise.

*P-Indicator***:** For interpretability, the D-Indicator is transformed into a bounded quantity:

$$\phi_{i,j,t} = 2\left(\Pr(Z \le \delta_{i,j,t}) - 0.5\right), \quad Z \sim \mathcal{N}(0,1)$$

The resulting **P-Indicator** is centered at zero, with:

- Positive values indicating worse-than-average performance
- Negative values indicating better-than-average performance

This transformation produces a smooth, continuous signal suitable for time-series analysis.

Statistical Event Detection: Rather than flagging individual spikes, we focus on **sustained abnormal behavior**. For each vehicle–component pair, we conduct a rolling hypothesis test over a 180-day window ending at time t:

- **Null hypothesis**: The average complaint signal is consistent with baseline behavior
- **Alternative hypothesis**: The average signal exceeds a predefined threshold

Formally:

$$H_0: \frac{1}{180} \sum_{T=t-180}^{t} \left[0.5(\phi_{i,j,T} + 1)\right] \leq 0.65$$

$$H_A: \frac{1}{180} \sum_{T=t-180}^{t} \left[0.5(\phi_{i,j,T} + 1)\right] > 0.65$$

This procedure yields a time series of p-values. Periods where $1 - p_t \geq 0.99$ are flagged as potential event activity. Adjacent flagged periods separated by less than 365 days are merged into a single contiguous **event window**, reflecting the persistent nature of emerging defects. See Figure 3 for a visualization of event detection for the Ford Escape Engine & Engine Cooling.

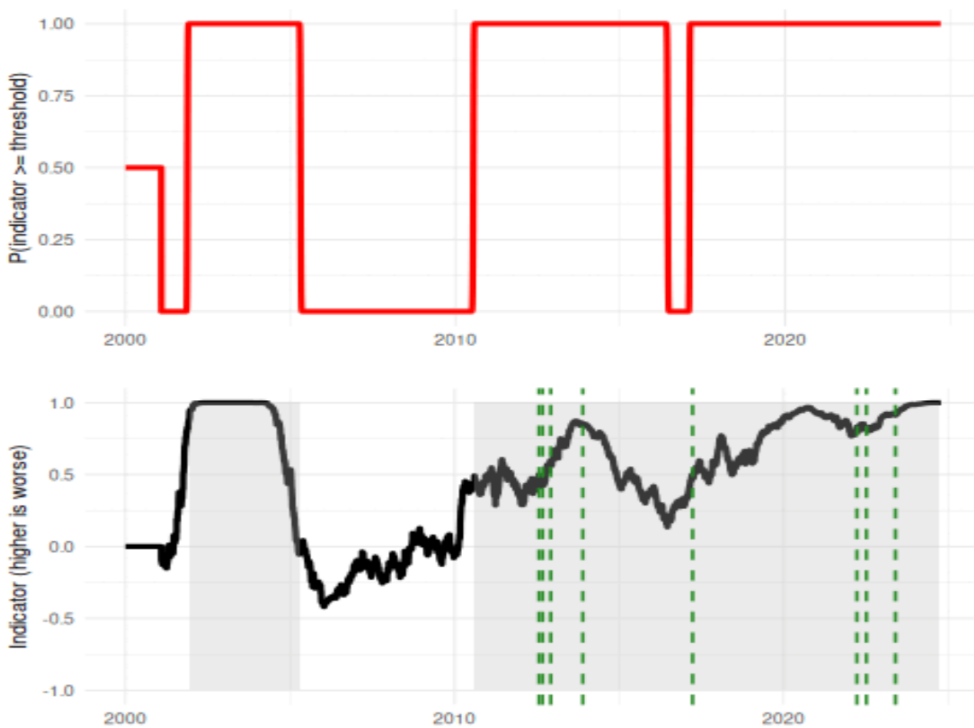

Figure 3. Event detection for the Ford Escape Engine & Engine Cooling.

Event Filtering and Validation: To reduce false positives, detected events are filtered using two criteria:

1. **Minimum complaint support**: The average number of complaints during the event window must exceed a minimum threshold (set to three complaints per period in this study).
2. **Temporal coherence**: Events must persist beyond isolated or transient spikes.

These filters ensure that flagged events reflect meaningful and sustained patterns rather than random noise.

Event-Level Summarization: For each detected event, we compute summary statistics using all complaints within the event window **and the preceding 365 days**, accounting for the trailing nature of the indicator. Event summaries include:

- Event duration (days)
- Total number of complaints
- Affected model years, identified via a binomial test comparing observed complaint proportions to a uniform baseline

In addition, complaint narratives are summarized using Meta's Llama 3.1 8B model [27] to produce concise, human-readable descriptions of the underlying issue. These summaries are used for interpretation only and do not influence event detection. The example below details complaints in a detected event for engine & engine cooling of the 2001 - 2002 Ford Escape.

Top Issues Summary:

- *Vehicles in this model are experiencing engine stalling or power loss at low speeds, often triggered by deceleration or idling, with possible loss of steering and braking control. Possible cause: faulty fuel injection, ignition timing, or related electrical controls.*

Detailed Description:

- *Multiple complaints describe the engine stalling or losing power at low speeds. Representative symptoms include:*
  - *Engine stalls when decelerating or idling*
  - *Possible loss of power steering and braking control*
  - *Power loss accompanied by warning lights*
  - *Stalling events occur most often going down hills or at low speeds (30-40 mph)*
- *These observations indicate a likely issue with fuel injection, ignition timing, or related electrical controls. Various complaints describe the issue recurring after dealership repairs, suggesting incomplete resolution of the problem.*

Integration of Online Forum Data Forum posts are unstructured and lack standardized metadata. To integrate them into the same analytical pipeline, we apply supervised natural language processing classifiers to tag posts by:

- Vehicle make
- Model
- Model year
- Component
- Relevance to defects

Once tagged, forum posts are treated analogously to complaints, allowing the same indicator construction and event detection framework to be applied. This enables direct temporal comparison between regulatory data and consumer-driven discussion.

RESULTS

Data Analysis

Trends in Detected Safety Events Figure 4 displays trends in active events detected across 25 components from 2000 through 2025. Several patterns are immediately apparent. The Air Bags component shows a pronounced spike peaking in late 2017 and early 2018, corresponding to the Takata airbag recall crisis. The Other or Unknown category exhibits a sharp increase beginning around 2015 that continues through the end of the observation period. Three ADAS-related components—Forward Collision Avoidance, Back Over Prevention, and Lane Departure—show rapid growth starting in 2020. Electrical System events demonstrate a long-run upward trend punctuated by spikes in 2014 and 2024. Engine and Engine Cooling events increase steadily until mid-2015, then decline. Fuel System events rise until approximately 2012, then flatten and show modest decline. We do not examine the Air Bags trend in detail, as the Takata crisis has been extensively documented elsewhere. Instead, we focus on the remaining trends to investigate the degree to which they are driven by technological changes in automobiles. To assess this, we conducted a detailed review of event narratives for components showing notable temporal variation. The review combines manual analysis with large language model support to identify recurring themes.

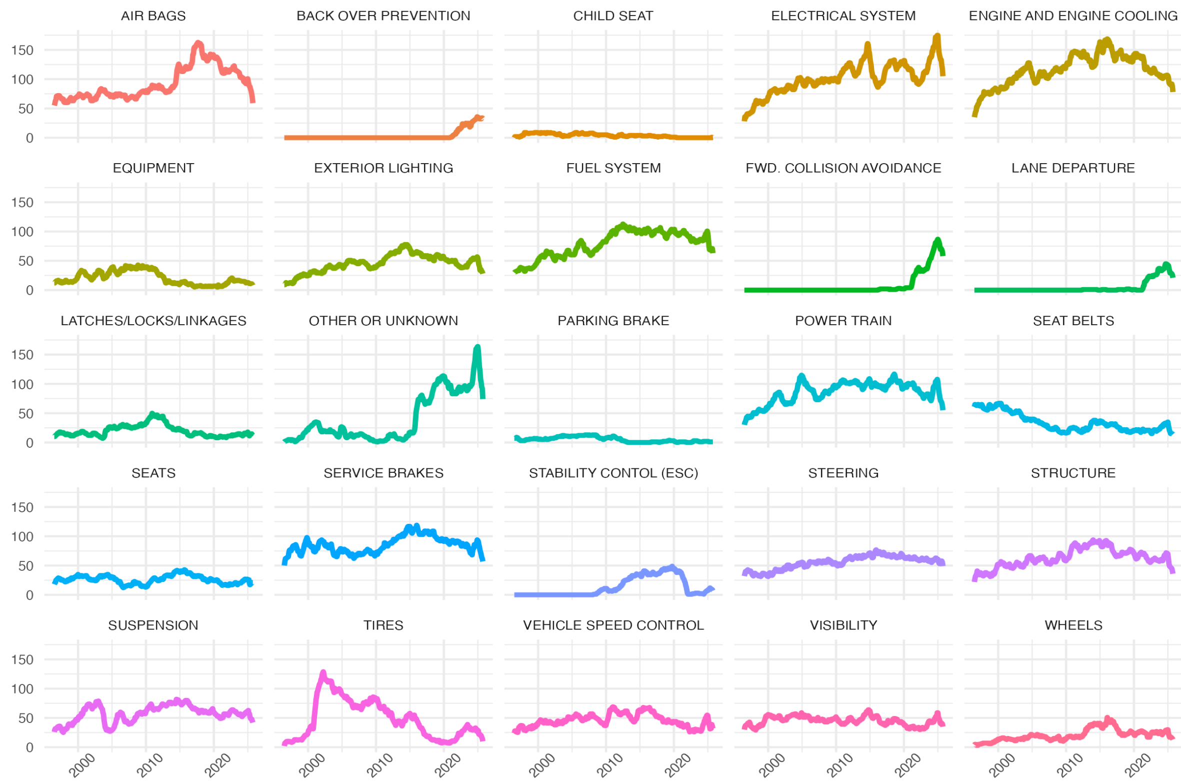

Figure 4. Count of active detected events over time.

<u>Trend in Other or Unknown Events</u> The Other or Unknown category primarily captures system-level failures that span multiple interconnected components rather than isolated defects. Most event summaries describe problems involving electronic modules, software, sensors, displays, and warning systems that produce cascading alerts, intermittent function loss, or system resets without identifiable mechanical causes. These multi-component failures do not align well with existing component taxonomies and are difficult to classify accurately. While some events reference ADAS-related symptoms—sensor malfunctions, camera failures, calibration problems, or unexpected system warnings—explicit references to autonomous functionality are rare, suggesting ADAS contributes only partially to this category. Water intrusion and moisture-related problems also appear frequently, typically manifesting indirectly through electrical or sensor failures rather than as isolated structural defects. These findings indicate that the growth in Other or Unknown events reflects both emerging ADAS-related issues and a broader class of system-level problems that strain traditional component-based classification schemes.

<u>Trend in Forward Collision Avoidance, Backover Prevention, and Lane Departure Events</u> The rapid increase in Forward Collision Avoidance, Back Over Prevention, and Lane Departure events reflects growing ADAS adoption in the vehicle fleet rather than taxonomy changes or misclassification. Review of natural-language summaries confirms these events are appropriately categorized: summaries consistently describe ADAS functionality—cameras, radar, sensors, warnings, and automated interventions—not unrelated mechanical issues. The dominant failure mode is system-level and software-mediated rather than hardware-based. Consumers report perception problems (sensor or camera performance and calibration), decision logic issues (false positives, unexpected interventions, or inconsistent activation), and system availability problems (features disabling themselves, becoming unavailable after updates, or failing intermittently). Explicit references to autonomous or self-driving functionality are uncommon; instead, consumers describe driver-assistance behaviors such as false collision warnings, unexpected braking, lane-keeping corrections, failure to detect obstacles, or intermittent backup alerts. These patterns indicate the observed increase represents genuine ADAS-related complaints at scale as these technologies become widespread, not a reclassification artifact.

Trend in Electrical System Events Electrical system events show a fundamental shift in failure types over time. The long-run increase reflects both greater electrical content in vehicles and expansion toward complex, integrated electronic and software-controlled systems. Traditional electrical problems—battery, alternator, wiring, no-start conditions, and intermittent power loss—dominate early events. Later events increasingly describe software-mediated system failures (persistent warning messages, intermittent faults affecting multiple functions, display/infotainment issues) and electrified powertrain problems (EV/hybrid charging, high-voltage systems, power-management warnings).

The two prominent spikes reflect distinct failure profiles. The 2014 spike is characterized by classic electrical and drivability problems, particularly ignition and starting failures, stalling, and sudden power loss. Manual review reveals frequent mentions of ignition switch issues, consistent with the major GM ignition switch recall that year. The 2024 spike shows a modern profile: summaries frequently reference charging and power-management concerns (EV/hybrid charging failures, high-voltage warnings) alongside concentrated infotainment, display, and software issues (repeated alerts, feature unavailability, intermittent electronic faults). These patterns indicate that spike periods capture time-localized concentrations of specific electrical themes rather than uniform trends, with the 2014 spike reflecting traditional mechanical-electrical failures and the 2024 spike reflecting software and electrification issues.

Trend in Engine and Engine Cooling Events Engine and engine cooling events show a modest decline beginning in mid-2015, with fewer new event onsets. This decline does not reflect fundamental changes in failure types: the dominant issues remain drivability symptoms—stalling, loss of power, engine shutoff, often with overheating or coolant problems—in both pre- and post-2015 periods. However, modest compositional changes are evident. Post-2015 summaries more frequently reference oil consumption, oil burning, and timing system issues (particularly timing chains) compared to earlier periods, suggesting evolving internal combustion failure modes. Explicit references to hybrid or electric propulsion systems remain rare and do not increase meaningfully around 2015. While fleet electrification may contribute to aggregate complaint volume trends by reducing the proportion of traditional internal combustion engine vehicles, the engine event summaries themselves provide no direct evidence of a shift toward electrified powertrains within this category.

Trend in Fuel System Events Fuel system events show a flattening and modest decline beginning in the early 2010s, driven by shorter event windows post-2012 and fewer new event onsets. More significantly, the nature of reported fuel system problems shifts substantially across this period. Early events emphasize physical fuel containment and delivery failures—fuel odors, visible leaks, problems with tanks, filler necks, caps, and associated fire risks. These align with traditional failure modes where fuel escapes the system or hardware degrades in seals, lines, and tank components.

Post-2012 events increasingly describe electronically mediated fuel delivery failures. Fuel pump problems become more prevalent, including drivability issues such as no-start conditions, intermittent power loss, or sudden stalling from fuel pressure loss. References to modern fuel delivery architectures (high-pressure fuel pumps) and emissions control subsystems (EVAP components) also increase. While this dataset cannot directly attribute changes to fleet composition shifts (hybrid or electric vehicle growth), the transition from leakage and tank-centered narratives toward pump, injection, and electronically controlled fuel delivery narratives is consistent with evolving propulsion and fuel management systems that rely more heavily on electronic control.

Consolidated Trend Analysis Consumer complaint events increasingly reflect complex, software-integrated vehicle behavior consistent with industry-wide shifts toward electrified, sensor-rich, and software-mediated architectures. This trend appears across multiple component groups. Electrical system events show long-run frequency increases with later events describing system-level electronic failures (persistent warnings, intermittent multi-function faults, display/infotainment issues, software updates). Other/Unknown events concentrate on cross-system symptoms difficult to assign to single components, featuring software behavior, sensors, cascading alerts, and water intrusion manifesting as electrical or sensor failures. ADAS categories (forward collision avoidance, lane departure, back-over prevention) show sharp post-2020 increases with coherent summaries describing perception problems (camera/sensor performance, calibration), decision logic issues (false warnings, unexpected interventions), and system availability problems (intermittent feature disabling) rather than physical component failures. Even traditionally mechanical domains show this shift: fuel system events transition from tank/leak/fire narratives toward electronically mediated pump and injection issues in the early 2010s, while engine events show modest post-2015 volume declines with stable symptoms but compositional drift toward oil consumption and timing system problems.

The consistency of patterns across multiple independent component categories provides strong descriptive evidence of systematic shifts in vehicle defect reporting. The convergence of themes—rising electrical and ADAS narratives, increasing system-level complexity in Other/Unknown, and drift toward electronically mediated fuel and powertrain descriptions—demonstrates that consumer complaints increasingly reflect software-integrated vehicle behavior and indicates the need for monitoring and taxonomy approaches that accommodate cross-system interactions and emerging functionality.

While the analysis is observational rather than causal, several factors strengthen confidence in these findings. The coherence of narratives within categories—particularly the concentration of ADAS-related symptoms, software behaviors, and cross-system failures in Other/Unknown—indicates substantive phenomena rather than classification artifacts alone. The qualitative shift in complaint content provides independent corroboration of technological change beyond simple volume trends. Component labels do reflect reporter perceptions and administrative categorization, and event count changes may partly reflect reporting behavior shifts, platform mix changes, and detection mechanics. The analysis does not directly measure fleet electrification rates, ADAS penetration, or exposure-adjusted risk, limiting precise attribution of volume changes to specific technology adoption rates. However, the systematic nature of the observed shifts across multiple component domains supports the interpretation that these patterns reflect genuine evolution in vehicle architectures and failure modes.

GM Ignition Switch Recalls: A Case Study in Early Defect Detection

Background On February 7, 2014, General Motors announced the first of what became a series of recalls for faulty ignition switches, ultimately affecting nearly 30 million vehicles worldwide across eight NHTSA campaigns. The defect—ignition switches with insufficient torque that could move out of the "run" position during vehicle operation—caused engine shutdown, loss of power steering and brakes, and airbag system disablement. The recalls were linked to at least 124 deaths and resulted in a $900 million Department of Justice criminal settlement.

Internal GM documents and subsequent investigations revealed that GM engineers were aware of the ignition switch deficiency by 2004-2005 but declined to initiate recalls, leaving the defect unaddressed for nearly a decade. This documented timeline—with manufacturer knowledge in 2004-2005 and recall announcement in February 2014—provides an ideal validation case for complaint-based early warning systems: a severe defect with known timing, substantial affected population, and a critical gap between internal awareness and regulatory action.

Approach: Event Identification and Validation We identified candidate events by extracting all make-model combinations from NHTSA's eight GM ignition switch recall campaigns and filtering our events database to these vehicles with electrical system component classifications. We restricted the temporal window to events ending after September 2001 (preceding GM's documented internal awareness) and starting before September 2015 (GM's Department of Justice settlement). Given the breadth of electrical system categories, we further filtered for event summaries containing "ignition," yielding 37 candidate events spanning 24 unique recalled vehicle models. Notably, all 37 events were derived from the NHTSA complaints database; no forum-derived events appeared.

Manual review of all 37 events found none that definitively should be excluded. Summary relevance varied: some explicitly described characteristic ignition switch symptoms (engine stalling while driving, key-related issues, loss of power steering, airbag warnings), while others described broader electrical issues potentially encompassing ignition problems. This variation reflects consumer complaint data characteristics—complainants report observed symptoms rather than diagnosed root causes.

For each vehicle, we calculated lead times as the difference between event start dates and the earliest NHTSA recall announcement date.

Key Findings Our analysis revealed strong early detection capability across the majority of recalled vehicles. Of the 24 unique vehicle models ultimately included in the GM ignition switch recalls, our system detected pre-recall events for 19 vehicles (79%). These 29 pre-recall events exhibited a mean lead time of 7.43 years (median: 7.59 years), with lead times ranging from 25 days to 16.3 years before the February 2014 recall announcement.

The temporal distribution of detected events aligned remarkably with the known timeline of GM's internal awareness. Eighteen events (62% of pre-recall detections) occurred during the 2001-2007 period—precisely when GM was internally aware of the defect but had not initiated recalls. Notable early detections included the Saturn Ion in May 2004 (9.7 years before recall), Chevrolet Cobalt in June 2006 (7.6 years before recall), and Pontiac Grand

Prix in February 2001 (13.0 years before recall). The Grand Prix case is particularly striking: our system detected three separate pre-recall events spanning 2001, 2006, and 2010, demonstrating persistent signal across the entire period of GM's inaction.

Vehicle detection patterns fell into three categories. Sixteen vehicles (67% of the 24 total) showed only pre-recall events with no post-recall spike, suggesting genuine early warning signals rather than recall-triggered artifacts. Three vehicles—Chevrolet Cobalt, Pontiac G5, and Pontiac Grand Prix—exhibited both early detection and post-recall events, validating that our system tracks evolving defect patterns through their lifecycle. The Cobalt's progression is particularly illustrative: events detected in June 2006, December 2011, and January 13, 2014 (just 25 days before the official recall announcement), showing escalating complaint volume as the defect manifested in aging vehicle populations. Finally, five vehicles (21%) were detected only post-recall, representing sensitivity limitations likely attributable to lower production volumes or complaint rates.

Figure 5 presents the temporal evolution of our statistical indicator for each of the 24 vehicles, with detected event windows and recall dates overlaid, illustrating the correspondence between elevated complaint indicators and official recall timing.

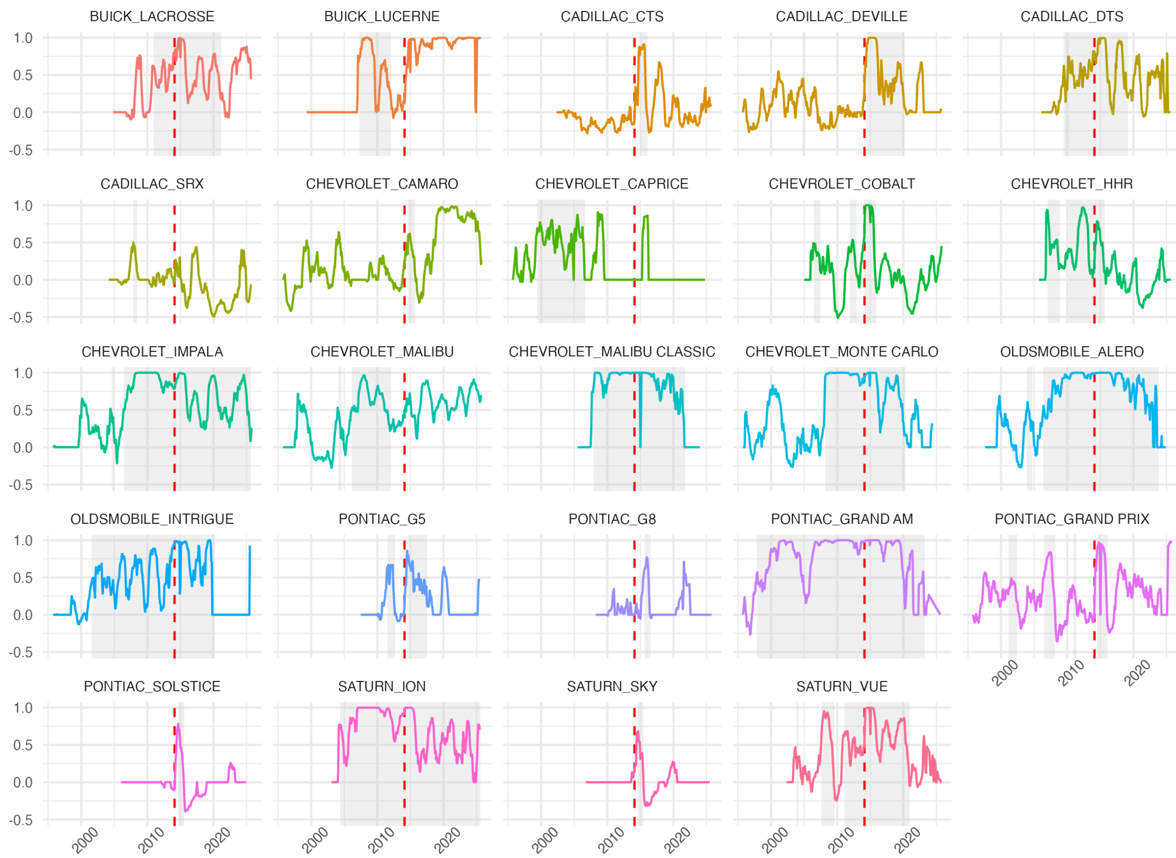

Figure 5. P-Indicator curves for vehicles in the GM ignition switch recall incident. Earliest recall for each vehicle is denoted with a vertical dashed line. Gray boxes denote detected event windows.

Assessment: Strengths and Limitations Our analysis demonstrates proof-of-concept for complaint-based early warning detection while revealing important methodological considerations. The primary strength lies in the temporal alignment: detecting 62% of events during the exact period when GM knew about the defect but had not acted provides compelling evidence that our unsupervised statistical approach extracts genuine safety signals from noisy consumer data. The 79% vehicle coverage and 7.43-year mean lead time indicate actionable warning capability—regulatory authorities or manufacturers monitoring such signals could have investigated years before the actual recalls.

The three vehicles showing both early and recall-triggered events (Cobalt, G5, Grand Prix) serve as particularly strong validation. These demonstrate that our system tracks genuine evolving defect signals rather than random noise or data artifacts. The Cobalt's three-event progression from 2006 through January 2014—culminating in an event starting just weeks before the announcement—captures the natural history of a defect as it manifests increasingly in aging vehicles and gains consumer awareness.

A notable methodological characteristic is the presence of long event windows spanning multiple years. This reflects several underlying patterns: overlapping events that our algorithm merges, ongoing service-related complaints following recalls (customers reporting repair delays or recurring problems), and the gradual manifestation of defects as vehicles age. While long windows can be interpreted as detecting "too much," several events were notably crisp and temporally focused—such as the 2006 Cobalt event (29 complaints over 12 months), the 2007 HHR event (32 complaints over 22 months), and the 2011 G5 event (8 complaints over 14 months). These focused events demonstrate that when complaint patterns are acute, our method generates correspondingly sharp temporal signals.

The conservative nature of our model year estimates represents both a strength and limitation. The binomial test typically identified a subset rather than the complete range of affected model years, which is appropriate for an unsupervised approach focused on detecting statistically elevated complaint rates rather than comprehensively enumerating all affected units. This conservatism prioritizes precision over sensitivity, reducing false positives at the cost of incomplete coverage—a reasonable tradeoff for an early warning system where flagging genuine problems takes precedence over capturing every edge case.

Limitations include the 21% of vehicles detected only post-recall, reflecting sensitivity gaps particularly for lower-volume vehicles. Model year coverage was incomplete, requiring complementary analysis to fully characterize defect scope. Event summaries exhibited variable specificity—some explicitly described ignition switch failures while others referenced broader electrical issues—though this reflects consumer reporting patterns rather than algorithmic failure. We cannot assess false negative rates: vehicles below our detection thresholds or with insufficient complaint volume remain unobservable in this retrospective analysis.

Critically, our system identified signals that NHTSA's own surveillance may have missed. While NHTSA investigated Chevrolet Cobalt power steering issues in 2009, the agency did not identify the ignition switch defect until the 2014 recalls. Our earliest relevant detection—Saturn Ion in May 2004—preceded regulatory action by nearly a decade and occurred concurrent with GM's internal awareness. This suggests that systematic statistical analysis of complaint patterns, even with simple threshold-based methods, can extract signals that escape traditional review processes.

DISCUSSION AND LIMITATIONS

Validation of Early Detection Capability

The GM case provides proof-of-concept for complaint-based early warning. Our system detected 19 of 24 recalled vehicles (79%) pre-recall with a mean 7.43-year lead time. Critically, 62% of pre-recall events occurred during 2001-2007 when GM was internally aware but had not acted—temporal alignment with manufacturer knowledge rather than recall timing demonstrates extraction of genuine safety signals. The Cobalt's three-event progression (2006, 2011, January 2014) illustrates tracking of evolving defect patterns, with the final event starting 25 days before the official announcement.

Our earliest detection (Saturn Ion, May 2004) was concurrent with GM's documented internal awareness. NHTSA investigated Cobalt power steering in 2009 but missed the ignition switch defect until 2014, suggesting systematic statistical monitoring can extract signals that escape expert-driven review of the same data.

Beyond GM validation, detected event patterns align with industry shifts: sharp post-2015 ADAS increases, the 2014 electrical spike (including GM issues), and the 2021 spike reflecting software integration. This coherence across independent component categories strengthens signal validity, though we emphasize these are descriptive observations rather than causal analyses.

Methodological Considerations

Our statistical framework balances sensitivity and specificity through deliberate design choices. The Cohen's d-based indicator with logarithmic scaling enables fair comparison across heterogeneous production volumes, while the 180-day rolling window and $p \geq 0.99$ threshold focus on sustained patterns rather than isolated spikes. Event merging and minimum complaint filtering further reduce false positives, though these conservative choices may miss low-volume or emerging signals.

The binomial test for model years is conservative—for better or worse—identifying statistically elevated complaint years rather than complete affected ranges. In the GM case, this detected subset years (e.g., Cobalt [2006, 2007] versus actual [2005-2010]) adequate for triggering investigation but insufficient for comprehensive scope determination—a reasonable tradeoff for an early warning system prioritizing precision over exhaustive enumeration.

Automated LLM summarization enables interpretable event descriptions at scale despite variable consumer complaint specificity. The GM case demonstrated this value: summaries captured symptom clusters even when individual narratives lacked technical precision, facilitating rapid relevance assessment across 37 candidate events.

Limitations and Validation Gaps

A critical limitation of this work is that the GM ignition switch case represents our sole detailed validation. We cannot infer system performance across the full 15,811 event database from this single example. The GM case may represent best-case, worst-case, or typical performance—we cannot determine which without broader evaluation. This limitation is not merely a matter of insufficient effort but reflects fundamental challenges in validating unsupervised anomaly detection systems for this domain.

Comprehensive validation is difficult for several reasons. First, our approach is unsupervised: we have no ground truth labels for the vast majority of detected events. Second, the data-generating process is asymmetric: many recalls occur without prior complaint spikes (manufacturer-initiated, compliance-driven, or based on warranty data rather than consumer complaints), while many complaint events never result in recalls (quality issues, isolated failures, customer service problems that do not constitute safety defects). Third, temporal mismatches complicate validation: we detected complaint events over a decade before the GM recall announcement, yet the vast majority of events in our database have no regulatory outcome against which to validate. These factors mean that traditional precision and recall metrics cannot be straightforwardly computed. Our results therefore demonstrate feasibility and proof-of-concept rather than validated system performance.

Within the GM case itself, 21% of vehicles (5 of 24) were detected only post-recall, reflecting sensitivity gaps inherent to statistical approaches. Lower production volumes generate fewer complaints, making statistical anomaly detection more challenging. We cannot assess the false negative rate for vehicles or defects that generated complaint signals below our detection thresholds—these remain unobservable in retrospective analysis.

Event windows are often long, spanning multiple years. This reflects several underlying patterns: overlapping events that our algorithm merges, ongoing service-related complaints following recalls (customers reporting repair delays or recurring problems), and the gradual manifestation of defects as vehicles age. While long windows can complicate interpretation, several events in our GM analysis were notably crisp and temporally focused—such as the 2006 Cobalt event (29 complaints over 12 months) and the 2011 G5 event (8 complaints over 14 months)—demonstrating that when complaint patterns are acute, our method generates correspondingly sharp temporal signals.

NHTSA's component taxonomy is hierarchical and can be quite granular, but we use only top-level categories for event detection. Overly-granular categorization disperses statistical signal across narrow subcategories, hampering anomaly detection—particularly for defects that consumers describe variably or that manifest across related subsystems. Consequently, our "Electrical System" category encompasses ignition switches, lighting, sensors, and other electrical components, meaning detected events may reflect multiple co-occurring issues rather than isolated failure modes. This breadth complicates root cause identification but is appropriate for a surveillance system designed to flag abnormal patterns for expert review rather than to diagnose specific technical failures. Hierarchical refinement remains an opportunity for future work once initial events are detected.

Forum Data: Limited Contribution in Present Analysis

In the GM ignition switch case study specifically, all 37 candidate events originated from NHTSA complaints; zero events were detected from forum data for the affected vehicles. The limited contribution of forum data in the present analysis reflects the characteristics of the selected forums. Mechanics Stack Exchange and CarTalk, while valuable resources for DIY troubleshooting and automotive discussion, lacked the systematic defect reporting volume and coverage necessary to generate statistically significant complaint rate elevations for most vehicle-component combinations. These forums skew toward enthusiast and technical discussions rather than broad-based defect reporting from the general vehicle-owning population.

This outcome motivated expanded data collection efforts. We are currently scraping approximately 40 manufacturer-specific automotive forums representing over 90% of U.S. market share by sales volume. These data have not yet been integrated into our analysis pipeline and are therefore not included in the present work. The hypothesis underlying this expansion is that manufacturer-specific forums with larger, more focused user bases may provide earlier or complementary signals relative to NHTSA complaints. The present work establishes the methodological framework for forum integration; comprehensive assessment of forum data value remains a future research direction.

Implications for Post-Market Surveillance

The 7.43-year average lead time demonstrates actionable early warning potential, with 62% of events occurring while the manufacturer knew but had not acted. This system would function as triage—prioritizing vehicle-component combinations for expert review—not replacing engineering analysis or root cause investigation.

Signal persistence (multiple events for the same vehicle) and characteristic aging patterns (complaints escalating over time) could enable escalation protocols in operational systems: repeated events trigger elevated scrutiny, while aging patterns inform predictive timing models. Integration opportunities include combining complaint events with TSB patterns, warranty clustering, and platform-based propagation across vehicles sharing common components.

Our approach complements NHTSA's expert-driven review by providing continuous, automated statistical flagging. This is increasingly important given fleet aging (average 12+ years) and industry transitions toward electrification, software integration, and ADAS. These shifts introduce new failure modes—intermittent software failures, complex system interactions—that statistical monitoring may detect where traditional mechanical defect patterns do not apply.

Future Research Directions

Several refinements would strengthen the approach demonstrated here. Additional validation case studies across different defect types (mechanical failures, electronic malfunctions, ADAS issues) would provide broader evidence of system performance and help characterize the conditions under which complaint-based detection is most effective. Systematic comparison to NHTSA investigation timing beyond publicly documented cases would clarify how detected events relate to internal agency processes. False positive analysis—sampling events that did not result in recalls and assessing their relevance through expert review—would provide critical information about signal specificity currently unavailable.

The parameters governing event detection ($p \geq 0.99$ threshold, 180-day rolling window, minimum 3 complaints average) were chosen conservatively based on initial exploration. Systematic sensitivity analysis across the parameter space could optimize the tradeoff between detection sensitivity and false positive rate, though this would likely require a ground truth dataset of expert-labeled events.

Integration of Technical Service Bulletins and NHTSA investigations, deferred in the present work, would enable analysis of temporal relationships between different information sources. Do manufacturer bulletins precede, follow, or coincide with complaint events? What fraction of detected events correspond to formal NHTSA investigations? Such analyses would illuminate the sequential structure of the complaint-investigation-recall pipeline and help identify where in this process statistical monitoring provides the most value.

The expanded forum data collection currently in progress represents a major opportunity to test whether manufacturer-specific forums with larger user bases provide earlier or different signals compared to NHTSA complaints. Comparative analysis of forum versus complaint event characteristics—including technical depth of discussions, temporal patterns, and affected vehicle populations—would clarify the distinct value proposition of each data source.

CONCLUSIONS

This work demonstrates that unsupervised statistical monitoring of consumer complaints can detect vehicle safety issues years before regulatory action. The GM ignition switch case study provides compelling validation: 79% of recalled vehicles detected pre-recall with a 7.43-year mean lead time, and 62% of events occurring during the documented period when the manufacturer knew but had not acted. This temporal alignment with internal knowledge timelines—not mere correlation with recall announcements—demonstrates extraction of genuine safety signals from noisy consumer data. Beyond this validation, the descriptive patterns we observe across component categories align with known industry shifts (ADAS proliferation, electrification, software integration), strengthening confidence that our framework captures real-world defect trends.

The GM case represents our sole detailed validation, and comprehensive system evaluation faces fundamental challenges inherent to unsupervised anomaly detection: asymmetric relationships between complaints and recalls, lack of ground truth labels, and temporal mismatches that preclude standard precision-recall metrics. Forum data integration, while methodologically demonstrated, contributed minimally due to limited volume in the selected forums, though expanded coverage is in progress. Integration of Technical Service Bulletins and investigation data would enable richer temporal analysis but remains future work.

These limitations notwithstanding, we have established proof-of-concept for a scalable approach to continuous vehicle safety surveillance. The framework operates across the full vehicle fleet without manual configuration, generates interpretable event descriptions through automated summarization, and has produced 15,811 events spanning nearly three decades—a resource that could support systematic early warning monitoring. The 7.43-year lead time and temporal alignment with manufacturer knowledge suggest actionable warning potential, particularly as the U.S. fleet ages (average 12+ years) and new failure modes emerge from electrification and software integration.

The path forward requires expanded validation across defect types, parameter optimization, false positive assessment, and integration with complementary data sources. Yet the core finding stands: systematic statistical monitoring can extract early warning signals from the same complaint data available to regulators today. Whether applied as continuous surveillance, retrospective analysis, or decision support for investigation prioritization, complaint-based event detection represents a feasible complement to existing expert-driven processes. This work establishes the foundation and demonstrates potential; realizing that potential in operational systems represents both an opportunity and a necessity as vehicle technology and fleet demographics continue to evolve.

ACKNOWLEDGEMENTS

The authors used Claude 4.5 Sonnet [28] and ChatGPT [29] to assist with data analysis, structuring arguments, and manuscript preparation. All AI-generated content was reviewed, fact-checked, and revised by the authors, who retain full responsibility for the accuracy and integrity of this work.

REFERENCES

[1] Bureau of Transportation Statistics, "Average age of automobiles in operation (CARS)." https://www.bts.gov/content/average-age-automobiles-and-trucks-operation-united-states, 2024.

[2] Cox Automotive, "Kelley blue book report: New-vehicle prices trend higher, as higher costs hit automakers, 2026 model year product arrives." https://www.coxautoinc.com/insights-hub/https-www-coxautoinc-com-market-insights-july-2025-atp-report/, 2025.

[3] NerdWallet, "Are car prices going up or down?" https://www.nerdwallet.com/auto-loans/learn/car-market-prices, 2025.

[4] JPMorgan Chase & Co. Global Research, "Who will pay the price for US auto tariffs?" https://www.jpmorgan.com/insights/global-research/autos/auto-tariffs, 2025.

[5] National Highway Traffic Safety Administration, “NHTSA datasets and APIs.” https://www.nhtsa.gov/nhtsa-datasets-and-apis, 2025.

[6] National Highway Traffic Safety Administration, “Risk-based processes for safety defect analysis and management of recalls,” U.S. Department of Transportation, 2020.

[7] CarComplaints.com, “CarComplaints.com: What’s wrong with your car?” https://www.carcomplaints.com, 2025.

[8] A. S. Abrahams, J. Jiao, W. Fan, G. A. Wang, and Z. Zhang, “What’s buzzing in the blizzard of buzz? Automotive component isolation in social media postings,” *Decision Support Systems*, 2013, doi: 10.1016/j.dss.2012.12.023.

[9] A. S. Abrahams, W. Fan, G. A. Wang, Z. Zhang, and J. Jiao, “An integrated text analytic framework for product defect discovery,” *Production and Operations Management*, vol. 24, no. 6, 2015, doi: 10.1111/poms.12303.

[10] X. Zhang, S. Niu, D. Zhang, G. A. Wang, and W. Fan, “Predicting vehicle recalls with user-generated contents,” 2015.

[11] S. Das, A. Mudgal, A. Dutta, and S. R. Geedipally, “Vehicle consumer complaint reports involving severe incidents,” *Transportation Research Record*, 2018, doi: 10.1177/0361198118788464.

[12] V. Astvansh and K. Eshghi, “The effects of regulatory investigation, supplier defect, and product age on stock investors’ reaction to an automobile recall,” *Journal of Business Research*, 2023, doi: 10.1016/j.jbusres.2023.114052.

[13] C.-F. Chi, D. Sigmund, and M. O. Astardi, “Classification scheme for root cause and failure modes and effects analysis of passenger vehicle recalls,” *Reliability Engineering & System Safety*, vol. 200, p. 106929, 2020, doi: 10.1016/j.ress.2020.106929.

[14] J. Ni and X. Huang, “Discovery-to-recall in the automotive industry,” *Journal of Supply Chain Management*, 2017, doi: 10.1111/jscm.12160.

[15] V. Astvansh, G. P. Ball, and M. Josefy, “The recall decision exposed: Automobile recall timing and process data set,” *Manufacturing & Service Operations Management*, vol. 24, no. 3, 2022.

[16] W. Shi, X. Xie, and X. Chen, “Unveiling the impact of defect complaints on vehicle recall delays,” *International Journal of Production Research*, 2025, doi: 10.1080/00207543.2025.2490978.

[17] M. Chakraborty and R. Gupta, “RECIPE: A vehicle RECall incident PrEdiction method,” in *Proceedings of the IEEE international conference on intelligent transportation systems*, 2023. doi: 10.1109/ITSC57777.2023.10421883.

[18] P. Maniyur, P. Nayar, and M. Bhat, “An advanced analytics approach for recall in the automotive industry,” 2017.

[19] M. Ghazizadeh, A. D. McDonald, and J. D. Lee, “Text mining to decipher free-response consumer complaints,” *Human Factors*, vol. 56, no. 6, 2014, doi: 10.1177/0018720813519473.

[20] K. M. Kwayu, V. Kwigizile, K. Lee, and J.-S. Oh, “Discovering latent themes in traffic fatal crash narratives using text mining analytics,” *Accident Analysis and Prevention*, 2021, doi: 10.1016/j.aap.2020.105899.

[21] A. Goswami, A. Verma, and M. Verma, “Modeling recalls for automobiles,” in *2025 international conference on computational, communication and information technology (ICCCIT)*, IEEE, 2025, pp. 164–169.

[22] J. Poumay and A. Ittoo, “Using meaning instead of words to track topics,” 2023.

[23] L. Zheng, Z. He, and S. He, “ITMDID: An improved topic model for defect information derivation,” *Expert Systems with Applications*, 2023, doi: 10.1016/j.eswa.2023.119947.

[24] P. Li, S. Chen, L. Yue, Y. Xu, and D. A. Noyce, “Analyzing relationships between latent topics in autonomous vehicle crash narratives using explainable XGBoost,” *Accident Analysis & Prevention*, 2024, doi: 10.1016/j.aap.2024.107605.

[25] Y.-N. Li, M. Jiang, L. Wang, and J. Wei, “XGBoost-based risk prediction model for massive vehicle recalls,” *Risk Analysis*, 2025, doi: 10.1111/risa.70052.

[26] J.-J. Tian, N. Sun, T. Bai, T. Liu, and F. Fei, “Research on prediction method of potential defect risk of motor vehicle based on BERT,” in *CICTP*, 2022.

[27] Meta, “Llama 3.1.” https://ai.meta.com/llama, 2025.

[28] Anthropic, “Claude 4.5 sonnet.” https://www.anthropic.com/claude, 2025.

[29] OpenAI, “ChatGPT.” https://openai.com/chatgpt, 2025.