International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 14 **Issue:** VIII **Month of publication:** August 2026

**DOI:** https://doi.org/10.22214/ijraset.2026.84685

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VIII Aug 2026- Available at www.ijraset.com

# Automation, Standardisation and Product Quality at Toyota: A Comparative Analysis of United States Vehicle Recall Records, 1984-2026

Perseus Bhavnagri
*Student, Master of Science (Information Technology), Wilson College (Autonomous), University of Mumbai, Mumbai, India*

***Abstract: The Toyota Production System is widely credited with setting the modern benchmark for manufacturing quality, yet the claim is usually supported by case description rather than by measurement against competitors. This study tests it directly. Using the complete safety recall register maintained by the United States National Highway Traffic Safety Administration, 4,916 vehicle recall campaigns filed between 1984 and 2026 were assembled at campaign level and grouped by parent company, so that Toyota, Lexus and Scion are compared with Ford, General Motors, Honda, Volkswagen Group and Nissan on equal terms. Seven hypotheses covering recall frequency, recall magnitude, defect detection speed, self-reporting behaviour, crisis response and failure profile were tested at a 5 per cent significance level using non-parametric methods, because both principal variables are severely right-skewed (skewness 7.96 for units affected). Three findings emerge. First, Toyota's exposure-adjusted recall frequency is 21.5 campaigns per 100 model-years offered against 40.0 for the comparison group, a rate 1.86 times lower ($p < 0.001$); on this measure Toyota is the strongest of the six manufacturers examined. Second, and in the opposite direction, a typical Toyota recall reaches 19,050 vehicles against 6,534 for the comparison group ($p < 0.001$), and this difference survives the removal of the 2009-2011 crisis years, indicating a structural rather than episodic pattern. Third, the widely assumed advantage in detection speed does not exist: the interval between the end of production and the filing of a recall does not differ significantly between Toyota and its competitors ($p = 0.58$, negligible effect size). Toyota self-reports a significantly higher share of its recalls than the comparison group (86.5 per cent against 81.5 per cent, $p = 0.010$), though the effect is small. The study concludes that Toyota's quality advantage is real but narrower than its reputation implies: the firm produces fewer defects rather than finding them faster, and the component standardisation that suppresses defect frequency simultaneously widens the population exposed when a defect does escape. Quality strategy in automotive manufacturing therefore involves a measurable trade-off between defect rate and defect reach.***

***Keywords: Toyota Production System, Product recall, Automotive quality, Lean manufacturing, Component standardisation, NHTSA, Manufacturing automation, Defect detection***

## I. INTRODUCTION

### *A. Background*

Building a motor vehicle is among the most demanding coordination problems in manufacturing. A modern car is assembled from roughly thirty thousand individual parts, drawn from hundreds of suppliers spread across several continents, and it must then survive a decade or more of use by people who did not build it and cannot inspect it. A fault that would be trivial in most products can, in a vehicle, injure or kill. Manufacturers therefore invest heavily in preventing defects, and the systems they have built to do so have shaped industrial practice far beyond the motor industry itself. No such system has been more influential than the one developed at Toyota Motor Corporation. The Toyota Production System grew out of the company's post-war effort to build vehicles economically at low volume, and it rests on a small number of ideas that have since been copied worldwide: produce only what is needed when it is needed, standardise work so that variation becomes visible, stop the line the moment something is wrong rather than repairing it later, and treat improvement as a continuous obligation of every employee rather than an occasional project. Automation occupies a particular place in this scheme. Machines are not introduced simply to replace labour; they are given the capacity to detect abnormality and halt themselves, so that a defect is contained at the moment it appears instead of being passed downstream. These principles have been studied extensively and adopted widely. Jadhav et al. [1] trace the sequence in which component manufacturers in India implement lean practices, and Habidin et al. [2] propose a structural model linking Lean Six Sigma practice to organisational performance among Malaysian automotive firms.

Kumar et al. [3] show how a Six Sigma programme within a single automotive company was directed at customer loyalty rather than at cost alone. The literature on adoption is therefore substantial.

### *B. Recalls as an Observable Measure of Quality*

What the literature offers far less of is external verification. Studies of lean adoption typically measure whether practices were introduced and how managers rate the outcome. Both are self-reported, both are collected from firms with an interest in the answer, and neither is comparable across companies that use different internal metrics. If the claim is that one manufacturing philosophy produces better vehicles than another, the evidence for it should come from outside the firms making the claim.

Safety recalls provide exactly such evidence. In the United States a manufacturer that discovers a defect posing an unreasonable risk to safety is legally obliged to notify the National Highway Traffic Safety Administration and to remedy the fault without charge. Every notification is recorded in a public register that names the manufacturer, the models and model years involved, the component at fault, the number of vehicles potentially affected, the dates between which the affected vehicles were built, and whether the manufacturer acted on its own initiative or after the regulator intervened. The register has been maintained continuously since 1967.

A recall is an admission that a defect escaped the factory and reached customers. It is a coarse measure, in that it captures only faults serious enough to be safety-related, and a delayed one, in that it records the moment of discovery rather than the moment of error. But it is recorded identically for every manufacturer, compiled by a regulator rather than by the firms themselves, and available over four decades. For a comparative question it is the most credible measure obtainable without privileged access to company records.

### *C. Problem Statement and Need for the Study*

Toyota's reputation for quality is treated in much of the management literature as settled fact rather than as a proposition requiring evidence. It is invoked to justify lean adoption in industries from healthcare to software, and it survived a recall crisis in 2009 and 2010 that was severe enough to prompt congressional hearings in the United States. Yet remarkably little published work compares Toyota's actual defect record against that of its competitors using a common, externally maintained measure.

Existing recall research has grown substantially but has developed along different lines. Ahsan [4] provides a trend analysis of car recalls in the United States market and demonstrates that recalls are common and unevenly distributed among carmakers, but the analysis treats brands individually, does not adjust for the very different numbers of models each firm sells, and closes in 2010. Ahsan and Gunawan [5] examine who initiates recalls across food and liquor, automobiles, toys and electrical goods, and find that manufacturers are the usual initiators, but they do not compare individual automakers with one another. A separate strand, represented by Chen and Nguyen [6], Mafael et al. [7], Raithel et al. [8], Gelbrich et al. [9] and Astira et al. [10], studies what happens after a recall is announced: how markets price it, how customers react, and how remedies should be managed. None of this work asks the prior question of whether one manufacturing philosophy yields fewer defects in the first place.

This study addresses that gap. It asks whether the manufacturing system for which Toyota is celebrated is visible in the firm's recall record when that record is compared, on equal terms and over four decades, with five other major manufacturers.

### *D. Significance of the Study*

For manufacturers, the study offers an externally verifiable benchmark rather than a self-assessment, and it identifies a trade-off in quality strategy that is not usually made explicit. For regulators, it provides evidence on how far manufacturers differ in reporting defects voluntarily. For researchers, it demonstrates that a public safety register can be used to test claims about manufacturing philosophy that are ordinarily supported only by case narrative. For students of operations management, it offers a worked example in which a widely repeated belief is partly confirmed and partly refuted by measurement.

## II. LITERATURE REVIEW

### *A. Lean Production and Quality Performance*

Research on lean production has largely concerned itself with implementation. Jadhav et al. [1] set out a sequence for adopting lean practices in Indian automotive component manufacturing and identify the obstacles that commonly interrupt it. Habidin et al. [2] examine Lean Six Sigma in the Malaysian automotive sector and propose a structural model relating lean and six sigma practice to organisational performance, deriving hypotheses for later testing rather than reporting measured outcomes.

Both studies advance understanding of how lean spreads. Neither observes product outcomes independently of the firms surveyed, and this limitation is characteristic of the field: adoption and perceived benefit are measured, while the defect record that quality management exists to improve is not.

Work on inspection and reliability approaches the problem from the engineering side. Tuominen [11] model the cost consequences of alternative inspection strategies in automotive quality control, showing that the placement of inspection within a process materially changes the economics of catching a fault. Summit [12] uses warranty claim data to model component reliability, and Deutschmann et al. [13] examine how electromagnetic interference threatens the functional safety of vehicle electronics. These studies illuminate mechanisms by which defects arise and are caught, but they operate at component or process level rather than comparing firms.

### *B. Product Recalls as a Field of Study*

Recall research has expanded considerably since the Toyota crisis. Ahsan [4] analyses recall trends in the United States car market and establishes that recalls are routine events concentrated among a limited number of manufacturers. Ahsan and Gunawan [5] extend the approach across product categories, classifying causes and identifying initiators, and conclude that manufacturers rather than regulators initiate most recalls. Beamish and Bapuji [14] examine the wave of toy recalls attributed to Chinese production and argue that design faults originating with the contracting firm accounted for a larger share than the public narrative allowed, a useful reminder that the location of blame in a recall is often assumed rather than established.

A second strand studies consequences. Chen and Nguyen [6] find that recall announcements depress share prices and shift analyst earnings forecasts. Mafael et al. [7] show that the remedy offered and the strength of the brand jointly determine whether customer satisfaction recovers. Raithel et al. [8] examine why owners often fail to bring vehicles in for correction, a finding with direct safety implications. Gelbrich et al. [9] study how dealers should manage the remedy, and Astira et al. [10] examine how perceived fairness after a recall influences continued commitment to the brand. Majid and Bapuji [15] analyse how firms respond when liability falls on the consumer.

A third strand concerns process and supply chain. Wynn et al. [16] specify the data and process requirements that effective recall coordination demands. Zhang et al. [17] argue that recall capability is built through supply chain quality management rather than assembled after a defect emerges. Bhattacharya et al. [18] describe the structural complexity of the Indian automotive supply chain, and Matsuo [19] documents how Toyota's coordination mechanism responded to the disruption caused by the Tohoku earthquake. Aalund et al. [20] analyse non-collision battery safety risks in electric vehicles, a category of defect that barely existed when the earlier recall literature was written.

### *C. Gap Analysis*

Three gaps follow from this body of work. First, recall studies analyse brands as they appear in the register, so that Toyota and Lexus, or Ford and Lincoln, are counted as unrelated firms even though they share engineering, components and quality systems. Second, recall counts are reported without adjustment for exposure: a manufacturer selling forty models will accumulate more recalls than one selling ten, irrespective of quality, and comparisons that ignore this are uninterpretable. Third, no published study appears to compare individual automakers on defect detection lag or on the proportion of recalls they initiate themselves, although both are directly implied by the quality philosophies these firms advocate. The gaps are summarised in Table I.

TABLE I
GAP ANALYSIS OF PRIOR LITERATURE

| Study | Focus | Contribution | Gap addressed by this study |
|---|---|---|---|
| Ahsan [4] | US car recall trends, 2000-2010, archival recall notices | Establishes that recalls are frequent and concentrated among a few carmakers | Analyses brands separately, does not group by parent company, does not adjust for how many models a firm sells, and ends in 2010 |
| Ahsan & Gunawan [5] | Content analysis of recall notices across product categories | Finds most recalls across all product types are initiated by the manufacturer | Compares product categories, not individual automakers; no automotive-only comparison of self-reporting |
| Majid and Bapuji [15] | Consumer liability and firm responsiveness in automobile recalls | Firms respond differently when liability rests with the consumer | Focuses on responsiveness after a recall is announced, not on how often recalls occur |
| Chen and Nguyen [6] | Stock price and analyst forecasts around recall announcements | Recalls carry a measurable financial penalty | Treats the recall as an event; does not examine the manufacturing behaviour that produced it |

| | | | |
|---|---|---|---|
| Wynn et al. [16] | Data and process requirements for recall coordination | Specifies the information a recall process must carry | Prescriptive and process-oriented; no empirical comparison between firms |
| Matsuo [19] | Toyota's coordination mechanism after the Tohoku earthquake | Documents how Toyota's supply chain absorbed a major shock | Single-firm and single-event; no quality benchmarking against rivals |
| Gelbrich et al. [9] | Remedy management for automotive recalls | Dealer handling of the remedy shapes customer response | Concerns the repair stage, not defect incidence |
| Zhang et al. [17] | Recall capability through supply chain quality management | Supply chain quality management builds recall capability | Survey-based and perceptual; no recall register data |
| Jadhav et al. [1]; Habidin et al. [2] | Lean and Lean Six Sigma implementation in automotive firms | Lean adoption is linked to improved operational performance | Measures adoption, perception or proposed models, never the recall record that quality should ultimately move |
| Tuominen [11]; Summit [12] | Inspection cost modelling and warranty-based reliability | Provides methods for reasoning about inspection and component reliability | Methodological; not applied comparatively across major manufacturers |

## III. OBJECTIVES, VARIABLES AND HYPOTHESES

### A. Objectives of the Study

The study pursues five objectives:

- To determine whether Toyota's recall frequency differs from that of comparable manufacturers once the number of models each firm sells is taken into account.
- To determine whether the number of vehicles affected by a single Toyota recall differs from that of comparable manufacturers.
- To determine whether Toyota identifies defects sooner after production than comparable manufacturers.
- To determine whether Toyota reports defects on its own initiative more often than comparable manufacturers.
- To establish whether the recall crisis of 2009 to 2011 represented a genuine departure from Toyota's own long-run record, and whether any differences identified persist once those years are excluded.

### B. Variables

Table II lists the variables used, their role in the analysis and their measurement.

TABLE II
VARIABLES USED IN THE STUDY

| Variable | Role | Measurement | Definition |
|---|---|---|---|
| Manufacturer group | Independent | Categorical (6 groups) | Parent company to which the recalled brand belongs (Toyota, Ford, GM, Honda, VW Group, Nissan) |
| Recall campaign count | Dependent | Count | Number of distinct recall campaigns filed by a group |
| Units affected | Dependent | Continuous (count of vehicles) | Potential number of vehicles covered by a single recall campaign (POTAFF) |
| Detection lag | Dependent | Continuous (days) | Days between the end of manufacture of the affected vehicles and the date the recall report was received |
| Recall initiator | Dependent | Binary | Whether the campaign was filed on the manufacturer's own initiative or followed regulator action |
| Component | Dependent | Categorical | Vehicle system named as defective in the campaign |
| Model-years offered | Exposure / control | Count | Distinct model-year combinations the group sold in the United States, from the EPA fuel economy register |
| Filing year | Control | Ordinal | Calendar year in which the recall report was received |

### C. Hypotheses

Seven hypotheses were formulated. H1 to H4 follow directly from the claims made for the Toyota Production System. H2 is stated in the opposite direction to the others: component standardisation, which is expected to reduce the frequency of defects, should by the same mechanism increase the number of vehicles exposed when a defect does occur, because a standardised part is fitted across many models. H5 tests the crisis period against Toyota's own history rather than against competitors, and H7 re-tests H2 with the crisis years removed in order to establish whether the pattern is structural or an artefact of a single episode. Each hypothesis is stated in null and alternative form in Table III.

TABLE III
HYPOTHESES TESTED

| # | Null hypothesis (H0) | Alternative hypothesis (H1) |
|---|---|---|
| H1 | Toyota's exposure-adjusted recall frequency does not differ from that of the comparison group. | Toyota's exposure-adjusted recall frequency is lower than that of the comparison group. |
| H2 | The number of vehicles affected by a Toyota recall campaign does not differ from the comparison group. | Toyota recall campaigns affect more vehicles per campaign than those of the comparison group. |
| H3 | Toyota's defect detection lag does not differ from that of the comparison group. | Toyota's defect detection lag is shorter than that of the comparison group. |
| H4 | The proportion of manufacturer-initiated recalls does not differ between Toyota and the comparison group. | Toyota initiates a higher proportion of its own recalls than the comparison group. |
| H5 | Toyota's recall frequency during 2009-2011 does not differ from its own 2004-2008 baseline. | Toyota's recall frequency during 2009-2011 exceeds its own 2004-2008 baseline. |
| H6 | The distribution of defective components is independent of manufacturer group. | The distribution of defective components depends on manufacturer group. |
| H7 | With the 2009-2011 crisis years removed, Toyota's units affected per campaign does not differ from the comparison group. | The difference identified in H2 persists after the crisis years are removed. |

## IV. METHODOLOGY

### A. Research Design

The study is quantitative, comparative and non-experimental, and is based entirely on secondary data. It was conducted between August and September 2026. The design is retrospective and observational: recall events that have already occurred are analysed as they were recorded, without intervention. Because the register is a complete administrative record rather than a sample drawn from a population, the analysis is a census of the recall activity of the six manufacturer groups over the study window.

Causal inference is not attempted. A recall register records that a defect was found and disclosed; it cannot establish why the defect arose. The study therefore tests whether the differences predicted by the Toyota Production System are observable in the record, and treats the manufacturing explanation as an interpretation of those differences rather than as a demonstrated cause.

### B. Data Collection

Two public datasets were used, both obtained directly from United States federal agencies and both free of licensing restriction.

The primary source is the recall flat file published by the Office of Defects Investigation of the National Highway Traffic Safety Administration, downloaded on 21 August 2026 from static.nhtsa.gov in two parts covering campaigns filed before and after 2010. The combined file contains 326,209 records across twenty-nine tab-delimited fields and covers every safety defect and compliance campaign filed since 1967.

The secondary source is the vehicle fuel economy register published by the United States Environmental Protection Agency, downloaded on the same date from fueleconomy.gov. It contains 50,242 records across eighty-four fields and lists every vehicle model certified for sale in the United States between 1984 and 2027. This file is used only to establish how many distinct models each manufacturer group offered in each year, which serves as the exposure denominator described below.

### *C. Data Preparation*

Four preparation steps were applied, each of which affects the interpretation of the results and is therefore stated explicitly.

First, records were restricted to those with a recall type code of V, denoting vehicle recalls. Campaigns concerning tyres, equipment, child restraints and other categories were excluded, since these are frequently filed by component suppliers rather than by vehicle manufacturers and would not represent the manufacturing behaviour under study. This reduced the file from 326,209 to 285,109 records.

Second, the register stores one row for each combination of campaign, make, model and model year, so that a single recall spanning eight models generates eight rows. Analysing rows would count large recalls repeatedly. Records were therefore collapsed to campaign level using the NHTSA campaign number. The number of vehicles potentially affected was confirmed to be constant within each campaign, so no aggregation rule was required for that field.

Third, brands were mapped to parent companies, so that Toyota, Lexus and Scion form one group; Ford, Lincoln and Mercury a second; Chevrolet, GMC, Buick, Cadillac, Pontiac, Saturn and Oldsmobile a third; Honda and Acura a fourth; Volkswagen, Audi and Porsche a fifth; and Nissan and Infiniti a sixth. Grouping is necessary because engineering, components and quality systems are shared within these families, and treating a luxury division as an independent manufacturer would misattribute its record.

Fourth, the analysis window was set to 1984 to 2026, matching the period covered by the EPA register and thereby allowing every recall in the analysis to be paired with an exposure measure. The resulting analytical dataset contains 4,916 recall campaigns. Four records reported a detection lag exceeding twenty years, the longest implying an interval of more than eighteen centuries; these are evidently data entry errors and were removed from the detection lag analysis only. Their removal does not alter any conclusion.

### *D. Sample Size*

Although the study analyses a complete register rather than a drawn sample, the adequacy of the resulting number of observations was verified against Cochran's formula for an infinite or very large population:

- $n = (Z^2 \times p \times (1 - p)) / e^2$
- $Z = 1.96$, corresponding to a 95 per cent confidence level
- $p = 0.5$, the value producing maximum variability and therefore the most conservative requirement
- $e = 0.05$, a 5 per cent margin of error

Substituting these values gives $n = (3.8416 \times 0.25) / 0.0025$, or approximately 385 observations. The analytical dataset contains 4,916 recall campaigns, of which 458 belong to Toyota. Both the full dataset and the Toyota subset exceed the minimum, so the analysis is adequately powered for the comparisons undertaken.

### *E. Analytical Techniques*

Both principal continuous variables depart severely from normality. Units affected has a skewness of 7.96 and detection lag a skewness of 58.6; a Shapiro-Wilk test on a random subsample of five hundred campaigns rejects normality decisively ($p < 0.001$). Distributions of this shape are expected, since most recalls are small while a few reach millions of vehicles. Parametric tests based on means would be dominated by the extreme upper tail, so non-parametric methods were used throughout.

- Mann-Whitney U test, for differences in units affected (H2, H7) and detection lag (H3). Cliff's delta is reported as the effect size, with magnitudes interpreted as negligible below 0.147, small below 0.33, medium below 0.474 and large above it.
- Binomial proportion test, for the exposure-adjusted recall frequency in H1.
- Chi-square test of independence, for the initiator comparison in H4 and the component profile in H6. Cohen's h and Cramér's V are reported as the corresponding effect sizes.
- Poisson test, for the comparison of Toyota's crisis-period recall rate against its own baseline in H5.

All tests were conducted at a 5 per cent significance level using Python 3.14 with SciPy 1.18 and pandas 3.0. Effect sizes are reported alongside every p-value, because a dataset of nearly five thousand campaigns will render trivially small differences statistically significant, and significance alone would therefore be misleading.

### *F. The Exposure Problem*

Comparing raw recall counts between manufacturers is invalid. A firm selling a wide range of models over four decades will accumulate more recalls than a firm selling a narrow range, whatever the quality of either. Some measure of exposure is required.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VIII Aug 2026- Available at www.ijraset.com

The ideal denominator would be vehicles sold in the United States by parent company and year, but no such series is available without charge for the full period; the International Organization of Motor Vehicle Manufacturers discontinued its manufacturer-level production series after 2017.

The number of distinct model-year combinations a group offered, taken from the EPA certification register, is used instead. It is an exact count of product breadth and is measured identically for every manufacturer. Its limitation is that it measures how many different models a firm sold rather than how many vehicles, and Toyota sells very large volumes of a comparatively small number of models. The implications of this are examined in Section IX.

## V. DATA ANALYSIS AND FINDINGS

### A. *Description of the Dataset*

The analytical dataset comprises 4,916 vehicle recall campaigns filed between 1984 and 2026 by six manufacturer groups, covering 590,864,514 vehicles in total. Ford accounts for the largest number of campaigns at 1,530 and General Motors for 1,401, together representing three-fifths of the dataset. Toyota accounts for 458 campaigns, Honda 445, Volkswagen Group 670 and Nissan 412.

The distribution of campaign size is extremely uneven. The median campaign affects 7,310 vehicles while the mean is 120,192, a gap that reflects a small number of very large campaigns; the largest single campaign in the dataset covers 7,900,000 vehicles. A quarter of all campaigns affect fewer than 487 vehicles. Table IV reports the position of each manufacturer group.

TABLE IV

RECALL PROFILE BY MANUFACTURER GROUP, 1984-2026

| Manufacturer group | Campaigns | Total vehicles affected | Median per campaign | Mean per campaign | Median lag (days) | Self-reported (%) |
|---|---|---|---|---|---|---|
| Toyota | 458 | 80,279,350 | 19,050 | 175,282 | 112 | 86.5 |
| Ford | 1,530 | 187,692,946 | 5,014 | 122,675 | 122 | 77.1 |
| General Motors | 1,401 | 174,757,858 | 6,004 | 124,738 | 162 | 81.9 |
| Honda | 445 | 79,764,997 | 11,511 | 179,247 | 216 | 83.1 |
| Volkswagen Group | 670 | 26,239,536 | 4,230 | 39,163 | 173 | 88.5 |
| Nissan | 412 | 42,129,827 | 14,230 | 102,257 | 113 | 83.0 |

### B. *Exposure-Adjusted Recall Frequency (H1)*

Toyota offered 2,129 distinct model-years in the United States between 1984 and 2026 and filed 458 recall campaigns, a rate of 21.5 campaigns per 100 model-years offered. The five comparison groups together offered 11,155 model-years and filed 4,458 campaigns, a rate of 40.0. A binomial proportion test rejects the null hypothesis decisively ($p < 0.001$), with a rate ratio of 0.538: Toyota recalls 1.86 times less often per model offered than the comparison group.

Table V and Fig. 1 show that Toyota records the lowest rate of the six groups. Volkswagen Group follows at 25.7 and Nissan at 30.3, while Ford records 67.6, more than three times Toyota's rate. Honda, whose reputation for reliability is often mentioned alongside Toyota's, records 50.1, more than double. H1 is rejected in favour of the alternative: Toyota's exposure-adjusted recall frequency is lower than that of the comparison group.

TABLE V

EXPOSURE-ADJUSTED RECALL FREQUENCY

| Manufacturer group | Recall campaigns | Model-years offered | Recalls per 100 model-years |
|---|---|---|---|
| Toyota | 458 | 2,129 | 21.5 |
| Volkswagen Group | 670 | 2,612 | 25.7 |
| Nissan | 412 | 1,360 | 30.3 |
| General Motors | 1,401 | 4,033 | 34.7 |
| Honda | 445 | 888 | 50.1 |
| Ford | 1,530 | 2,262 | 67.6 |

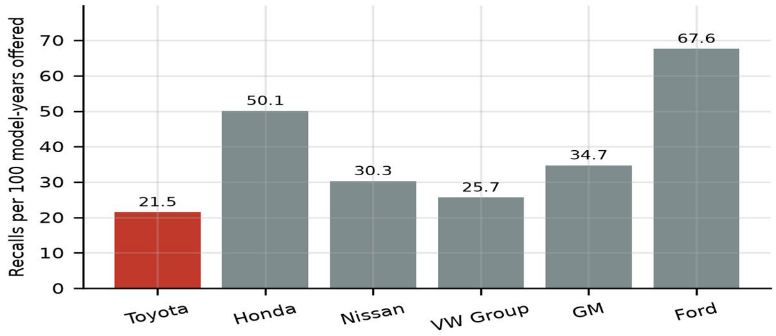



Fig. 1 Exposure-adjusted recall frequency by manufacturer group

*C. Vehicles Affected per Campaign (H2)*

The pattern reverses when the size of each recall is examined. A median Toyota campaign affects 19,050 vehicles against 6,534 for the comparison group, a ratio of approximately 2.9 to 1. A Mann-Whitney U test rejects the null hypothesis ($U = 1{,}206{,}942$, $p < 0.001$), with Cliff's delta of +0.182, a small but clearly positive effect.

Table IV and Figs. 2 and 3 show that Toyota records the largest median campaign of the six groups, followed by Nissan at 14,230 and Honda at 11,511. Ford and General Motors, which recall most often, record the smallest typical campaigns at 5,014 and 6,004 respectively. The ordering is close to the inverse of that observed for recall frequency.

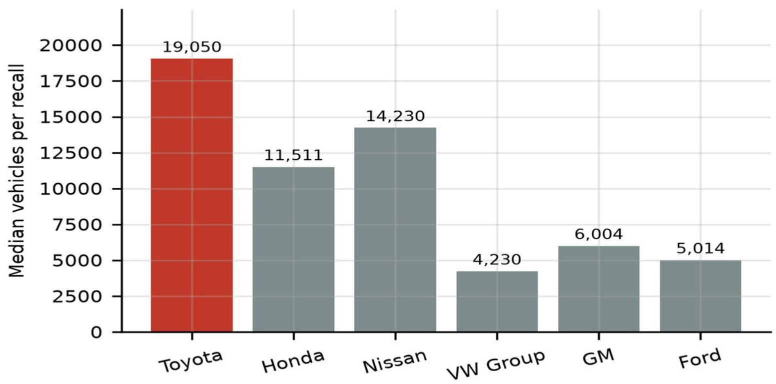



Fig. 2 Median vehicles affected per recall campaign

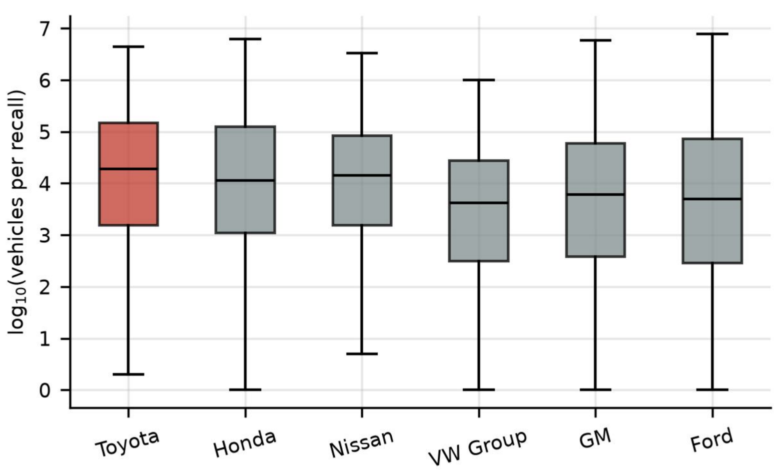



Fig. 3 Distribution of recall size (logarithmic scale)

### *D. Defect Detection Lag (H3)*

The interval between the end of manufacture of the affected vehicles and the filing of the recall report was calculated for 3,614 campaigns in which both dates were recorded and the resulting interval was plausible. Toyota's median interval is 112 days against 148 days for the comparison group, which appears to favour Toyota. The difference is not statistically significant: a Mann-Whitney U test returns $p = 0.580$, and Cliff's delta of -0.018 indicates a negligible effect. Retaining the four implausible records described in Section IV-C changes nothing ($p = 0.555$, Cliff's delta -0.020).

The null hypothesis cannot be rejected. Fig. 4 shows why the descriptive comparison is misleading: the comparison group is heterogeneous, with Nissan at 113 days and Ford at 122 days performing much as Toyota does, while Honda at 216 days and Volkswagen Group at 173 days pull the pooled median upward. Toyota is quick relative to some competitors and comparable to others, but it does not detect defects faster than its competitors as a class. This result is reported in full because it contradicts an expectation the study set out with.

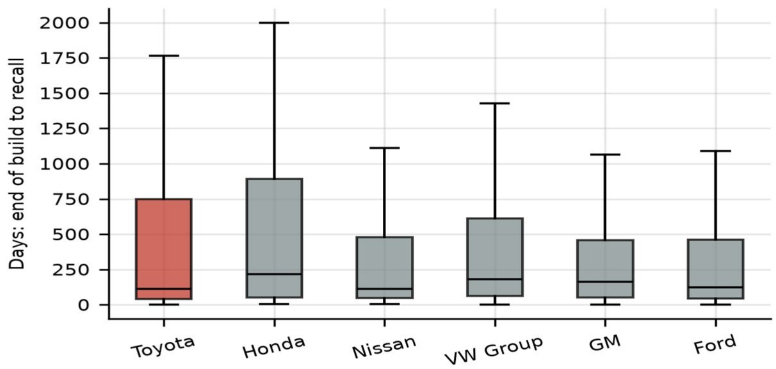



Fig. 4 Interval from end of production to recall filing

### *E. Who Initiates the Recall (H4)*

Of Toyota's 458 campaigns, 396 were filed on the manufacturer's own initiative and 62 followed regulator action, a self-reporting rate of 86.5 per cent. The comparison group self-reported 3,631 of the 4,457 campaigns for which an initiator was recorded, or 81.5 per cent. A chi-square test of independence rejects the null hypothesis (chi-square = 6.67, df = 1, $p = 0.010$), with Cohen's h of +0.137 indicating a small effect.

Fig. 5 shows that Volkswagen Group records the highest rate at 88.5 per cent, with Toyota second at 86.5 per cent and Ford lowest at 77.1 per cent. Toyota is therefore the highest of the mass-market manufacturers examined but not the highest overall. The finding is consistent with the direction predicted by the Toyota Production System, and consistent also with the observation of Ahsan and Gunawan [5] that manufacturers initiate the majority of recalls across product categories, though the effect size indicates that the difference between firms is modest.

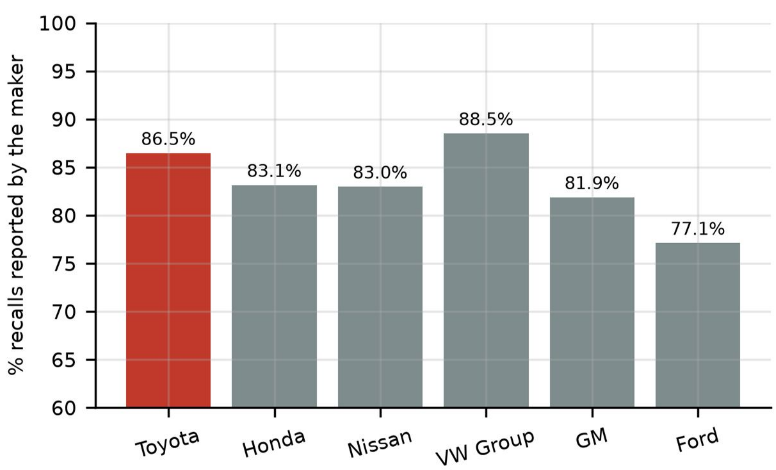



Fig. 5 Proportion of recalls initiated by the manufacturer

### F. The 2009 to 2011 Crisis (H5)

Between 2004 and 2008 Toyota filed 43 recall campaigns, an average of 8.6 per year. Between 2009 and 2011 it filed 47, an average of 15.7 per year. A Poisson test against the earlier baseline gives $p < 0.001$; the crisis period rate is 1.82 times the preceding rate. The episode is therefore a genuine and statistically significant departure from Toyota's own record, not an artefact of reporting or of general industry trends. Fig. 6 places it in context: Toyota's recall count rises sharply in 2010 and remains elevated for several years, but the line stays well below those of Ford and General Motors throughout, including at its peak.

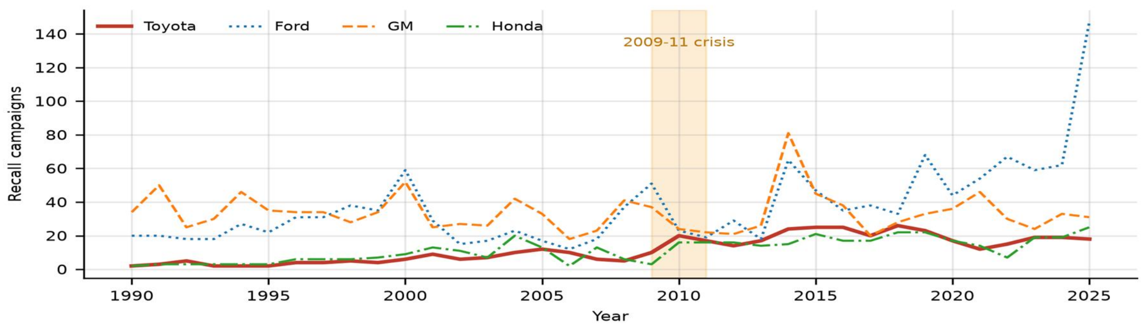



Fig. 6 Recall campaigns per year, 1990-2025

### G. Failure Profile by Manufacturer (H6)

The ten most frequently recalled components across the dataset were cross-tabulated against manufacturer group. A chi-square test of independence rejects the null hypothesis (chi-square = 212.0, df = 45, $p < 0.001$), but Cramer's V of 0.114 indicates a weak association. Manufacturers differ in what fails, but not dramatically.

Air bags are the most frequently recalled component overall, appearing in 465 campaigns, followed by gasoline fuel systems at 444 and electrical systems at 417. Air bags are also Toyota's most common recall category, at 64 campaigns, a pattern shaped substantially by the industry-wide Takata inflator recalls, which affected many manufacturers simultaneously and originated with a shared supplier rather than with any individual carmaker.

### H. Robustness (H7)

Because the 2009 to 2011 crisis involved several unusually large campaigns, the possibility arises that the H2 result is driven by that episode. H2 was therefore re-tested with all campaigns filed between 2009 and 2011 removed, leaving 411 Toyota campaigns and 4,193 comparison campaigns. The difference persists. Toyota's median campaign affects 16,582 vehicles against 6,807 for the comparison group, with $p < 0.001$ and Cliff's delta of +0.156, still a small positive effect. The tendency for Toyota recalls to reach more vehicles is therefore a structural feature of the firm's product architecture rather than a residue of one crisis.

## VI. RESULTS

Six of the seven hypotheses were rejected in favour of their alternatives; H3 could not be rejected. Table VI summarises the outcome of each test.

TABLE VI
SUMMARY OF HYPOTHESIS TESTING RESULTS

| Hypothesis | Test used | Statistic / effect size | p-value | Decision at alpha = 0.05 |
|---|---|---|---|---|
| H1 | Binomial proportion test | rate ratio = 0.538 | < 0.001 | Reject H0 |
| H2 | Mann-Whitney U | U = 1,206,942; d = +0.182 | < 0.001 | Reject H0 |
| H3 | Mann-Whitney U | U = 536,238; d = -0.018 | 0.580 | Fail to reject H0 |
| H4 | Chi-square test of independence | chi2 = 6.67; h = +0.137 | 0.010 | Reject H0 |
| H5 | Poisson test | rate ratio = 1.82 | < 0.001 | Reject H0 |
| H6 | Chi-square test of independence | chi2 = 212.0; V = 0.114 | < 0.001 | Reject H0 |
| H7 | Mann-Whitney U | U = 1,001,431; d = +0.156 | < 0.001 | Reject H0 |

Taken together the results describe a manufacturer whose quality advantage is real but narrower and differently shaped than its reputation suggests. Toyota recalls substantially less often than its competitors once product breadth is taken into account, and this is the largest and most robust effect in the study. It also reports a significantly higher share of its own defects, though by a small margin. It does not, however, detect defects faster, and when it does recall, the campaign reaches roughly three times as many vehicles as a typical competitor campaign.

## VII. DISCUSSION

### *A. Fewer Defects Rather Than Faster Detection*

The combination of H1 and H3 is the study's most informative result. Toyota's recall advantage is one of prevention, not of detection. The firm does not find faults sooner than its competitors; it has fewer faults to find. This distinction matters because the two mechanisms imply different explanations. Faster detection would point to superior monitoring, field surveillance and internal reporting. A lower defect rate points instead to what happens before a vehicle is built: standardised work, disciplined process control, mature supplier relationships and the containment of abnormality at the point where it occurs. The finding also qualifies a common reading of jidoka, the principle that a machine or operator should stop production the moment something abnormal appears. Whatever jidoka achieves inside the factory, it does not translate into a measurably faster response to defects that reach customers. Once a fault escapes the plant, Toyota appears to identify it on much the same timescale as everyone else.

### *B. The Standardisation Trade-off*

H1 and H2 point in opposite directions, and the tension between them is the study's most interesting contribution. Toyota recalls less often but each recall reaches more vehicles, and the pattern holds when the crisis years are removed. A plausible explanation is that both results follow from the same cause. Component standardisation across model lines reduces the number of distinct parts a manufacturer must design, validate and control, which should lower the probability that any given part is defective. The same standardisation means that a part which does prove defective is installed across a wide range of vehicles. Fewer faults are created, but each fault propagates further. On this reading the two findings are not in conflict but are two consequences of a single design philosophy, and the apparent paradox dissolves into a trade-off. An alternative and more favourable explanation deserves equal weight, and the present data cannot distinguish between them. A larger campaign may reflect a broader definition of the affected population rather than a wider defect. A manufacturer that recalls every vehicle that might contain a suspect part, rather than only those known to contain it, will record larger campaigns while behaving more cautiously. Distinguishing conservative scoping from genuine propagation would require engineering records that lie outside the public register, and this is identified in Section X as the most valuable extension of the present work.

### *C. Interpreting the Crisis*

The 2009 to 2011 crisis was real by the measure used here, with recall frequency reaching 1.82 times Toyota's own baseline. Yet Fig. 6 shows that even at its peak Toyota's absolute recall count remained below that of Ford and General Motors in most years. The crisis was a substantial deterioration relative to Toyota's own standard rather than a fall below the industry's. That both statements are true simultaneously illustrates why comparison against a single firm's history and comparison against its competitors answer different questions, and why reporting only one of them would misrepresent the record.

### *D. Self-Reporting and Regulatory Behaviour*

Toyota self-reports a higher proportion of its recalls than the comparison group, and the difference is statistically significant, but Cohen's h of 0.137 places the effect in the small range. The finding should therefore be stated carefully. Manufacturer-initiated reporting is the norm across the industry, ranging from 77.1 per cent at Ford to 88.5 per cent at Volkswagen Group, which is consistent with the pattern Ahsan and Gunawan [5] observed across other product categories. Toyota sits towards the upper end of a range in which every manufacturer reports most of its own defects, rather than standing apart from the industry.

## VIII. RECOMMENDATIONS

Four recommendations follow from the findings.

1) Manufacturers should evaluate component standardisation on two dimensions rather than one. The efficiency and quality gains of shared parts are well understood; the concentration of recall exposure that accompanies them is less often quantified. A part used across many models warrants validation effort proportionate to the population it will reach, not merely to its unit cost.

2) Firms should not assume that internal quality practice converts into faster field detection. The absence of any detection advantage at Toyota, despite a manufacturing system explicitly built around immediate containment, suggests that in-plant detection and field defect discovery are largely separate capabilities requiring separate investment.
3) Regulators should consider publishing recall rates adjusted for exposure alongside raw counts. Raw counts systematically flatter manufacturers with narrow product ranges and penalise those with broad ones, and the ordering of firms changes materially once exposure is taken into account.
4) Researchers using recall registers should collapse records to campaign level and group brands by parent company. Neither step is difficult, both are frequently omitted, and omitting either produces figures that are not comparable across firms.

## IX. LIMITATIONS OF THE STUDY

Six limitations qualify these findings and should be read alongside them.

1) The exposure denominator counts models offered, not vehicles sold. Toyota sells very high volumes of a comparatively small number of models, so its recall rate per vehicle sold may differ from its rate per model offered. This is the most serious limitation of the study, and it could be resolved only with a manufacturer-level sales series that is not publicly available for the full period.
2) The data covers the United States market alone. Vehicles built for other markets differ in specification, and regulatory regimes differ in what they require to be reported, so the findings should not be generalised to global operations without further evidence.
3) A recall records a defect that was found and disclosed. Defects that were never identified, or that were handled through service campaigns falling short of a safety recall, are invisible to this analysis. The register measures disclosed defects, which is not identical to defects.
4) The number of vehicles potentially affected is reported by the manufacturer and represents the population covered by the campaign rather than the number of vehicles actually defective. Differences in scoping practice between manufacturers are therefore embedded in the H2 result and cannot be separated from genuine differences in defect propagation.
5) Detection lag is measured from the end of the manufacturing period to the filing of the report. It is an approximation of discovery speed, since the manufacturer may have identified the fault at any point before filing, and the true internal discovery date is not recorded.
6) The study is observational and establishes association rather than causation. Manufacturers differ in market position, product mix, plant location and supplier base as well as in production philosophy, and the register cannot isolate the contribution of any one of these.

## X. FUTURE SCOPE

Five extensions would strengthen or usefully challenge these findings.

1) Repeating the analysis with vehicles sold as the denominator, using a licensed sales series, would test directly whether Toyota's frequency advantage survives the change of exposure measure. This is the single most valuable extension available.
2) Distinguishing conservative recall scoping from genuine defect propagation would resolve the competing explanations offered in Section VII-B. Comparing the number of vehicles covered by a campaign with the number actually presented for remedy, a figure manufacturers report to the regulator in quarterly completion filings, would offer a route into this question.
3) Extending the comparison to other regulators, in particular the European and Japanese recall registers, would establish whether the pattern observed in the United States reflects manufacturing behaviour or the reporting environment.
4) Applying text analysis to the defect and consequence narratives recorded in the register would permit classification of root causes as design, manufacturing or supplier faults, following the approach Ahsan and Gunawan [5] applied across product categories, and would allow the standardisation hypothesis to be tested directly rather than inferred.
5) Examining whether recall patterns are changing as vehicles become software-defined and electrified would be timely. Aalund et al. [20] show that battery safety introduces failure modes with no mechanical precedent, and software faults are increasingly remedied by over-the-air update, a remedy category the register has only recently begun to record.

## XI. CONCLUSION

This study set out to test, against an external and independently maintained record, a claim that is usually asserted rather than measured: that the Toyota Production System delivers superior product quality. Using 4,916 recall campaigns filed with the United States National Highway Traffic Safety Administration between 1984 and 2026, and comparing Toyota with five other manufacturer groups on equal terms, the claim is substantially supported but requires qualification in two respects.

The support is clear. Adjusted for the number of models each firm sells, Toyota recalls 1.86 times less often than its competitors and records the lowest rate of the six groups examined, less than a third of Ford's. It also discloses a significantly higher proportion of its own defects than the comparison group. On the central question of whether Toyota produces fewer defective vehicles than its competitors, the evidence says that it does.

The qualifications are equally clear. Toyota does not identify defects faster than its competitors; the apparent advantage in the descriptive figures disappears under test, and the firm's advantage lies in prevention rather than detection. And Toyota's recalls reach roughly three times as many vehicles as those of a typical competitor, a difference that persists when the 2009 to 2011 crisis years are excluded and that is therefore structural. The most plausible reading is that component standardisation both suppresses defect frequency and widens defect reach, so that the strength and the weakness share a single origin.

The wider implication is that quality in automotive manufacturing is not a single quantity along which firms can be ranked. A manufacturer may be excellent at not creating faults while being ordinary at finding them, and the design choices that reduce how often faults occur may increase how far they spread. Reputation tends to compress these distinctions into a single verdict. Measurement separates them again, and the separation is where the useful knowledge lies.

## REFERENCES


[1] J. R. Jadhav, S. S. Mantha, and S. B. Rane, "Roadmap for Lean implementation in Indian automotive component manufacturing industry: comparative study of UNIDO Model and ISM Model," Journal of industrial engineering international, 2014. doi: 10.1007/s40092-014-0074-6.

[2] N. F. Habidin, S. M. Yusof, C. M. Z. C. Omar, et al., "Lean Six Sigma Initiative: Business Engineering Practices and Performance in Malaysian Automotive Industry," IOSR Journal of Engineering, 2012. doi: 10.9790/3021-02711318.

[3] M. Kumar, J. Antony, F. J. Antony, et al., "Winning customer loyalty in an automotive company through Six Sigma: a case study," Quality and Reliability Engineering International, 2006. doi: 10.1002/qre.840.

[4] K. Ahsan, "Trend Analysis of Car Recalls: Evidence from the US Market," International Journal of Managing Value and Supply Chains, 2013. doi: 10.5121/ijmvsc.2013.4401.

[5] K. Ahsan and I. Gunawan, "Analysis of Product Recalls: Identification of Recall Initiators and Causes of Recall," Operations and Supply Chain Management An International Journal, 2014. doi: 10.31387/oscm0180115.

[6] Y. Chen and N. H. Nguyen, "Stock Price and Analyst Earnings Forecast around Product Recall Announcements," International Journal of Economics and Finance, 2013. doi: 10.5539/ijef.v5n6p1.

[7] A. Mafael, S. Raithel, and S. J. Hock, "Managing customer satisfaction after a product recall: the joint role of remedy, brand equity, and severity," Journal of the Academy of Marketing Science, 2021. doi: 10.1007/s11747-021-00802-1.

[8] S. Raithel, S. J. Hock, and A. Mafael, "Product recall effectiveness and consumers' participation in corrective actions," Journal of the Academy of Marketing Science, 2023. doi: 10.1007/s11747-023-00967-x.

[9] K. Gelbrich, S. Voigt, and A. Nazifi, "Remedy management for product recalls in the automotive industry: How car dealers should time the repair and communicate its outcome," Journal of Business Research, 2023. doi: 10.1016/j.jbusres.2023.114173.

[10] A. A. Astira, E. Yulianto, and A. Rahimah, "From Recall to Commitment: The Influence of Distributive Justice and Severity Recall on Affective Trust and Corporate Reputation in the Automotive Sector," Jurnal Manajemen Teori dan Terapan | Journal of Theory and Applied Management, 2025. doi: 10.20473/jmtt.v18i1.66154.

[11] V. Tuominen, "Cost Modeling of Inspection Strategies in Automotive Quality Control," Engineering Management Research, 2012. doi: 10.5539/emr.v1n2p33.

[12] R. A. Summit, "Modelling component reliability using warranty data," ANZIAM Journal, 2012. doi: 10.21914/anziamj.v53i0.5115.

[13] B. Deutschmann, G. Winkler, and P. Kastner, "Impact of electromagnetic interference on the functional safety of smart power devices for automotive applications," e+i Elektrotechnik und Informationstechnik, 2018. doi: 10.1007/s00502-018-0633-4.

[14] P. W. Beamish and H. Bapuji, "Toy Recalls and China: Emotion vs. Evidence," Management and Organization Review, 2008. doi: 10.1111/j.1740-8784.2008.00105.x.

[15] K. Majid and H. Bapuji, "Consumer liability and firm responsiveness: evidence from automobile recalls," Journal of International Business Policy, 2024. doi: 10.1057/s42214-024-00189-0.

[16] M. T. Wynn, C. Ouyang, A. H. M. ter Hofstede, et al., "Data and process requirements for product recall coordination," Computers in Industry, 2011. doi: 10.1016/j.compind.2011.05.003.

[17] M. Zhang, H. Hu, and X. Zhao, "Developing product recall capability through supply chain quality management," International Journal of Production Economics, 2020. doi: 10.1016/j.ijpe.2020.107795.

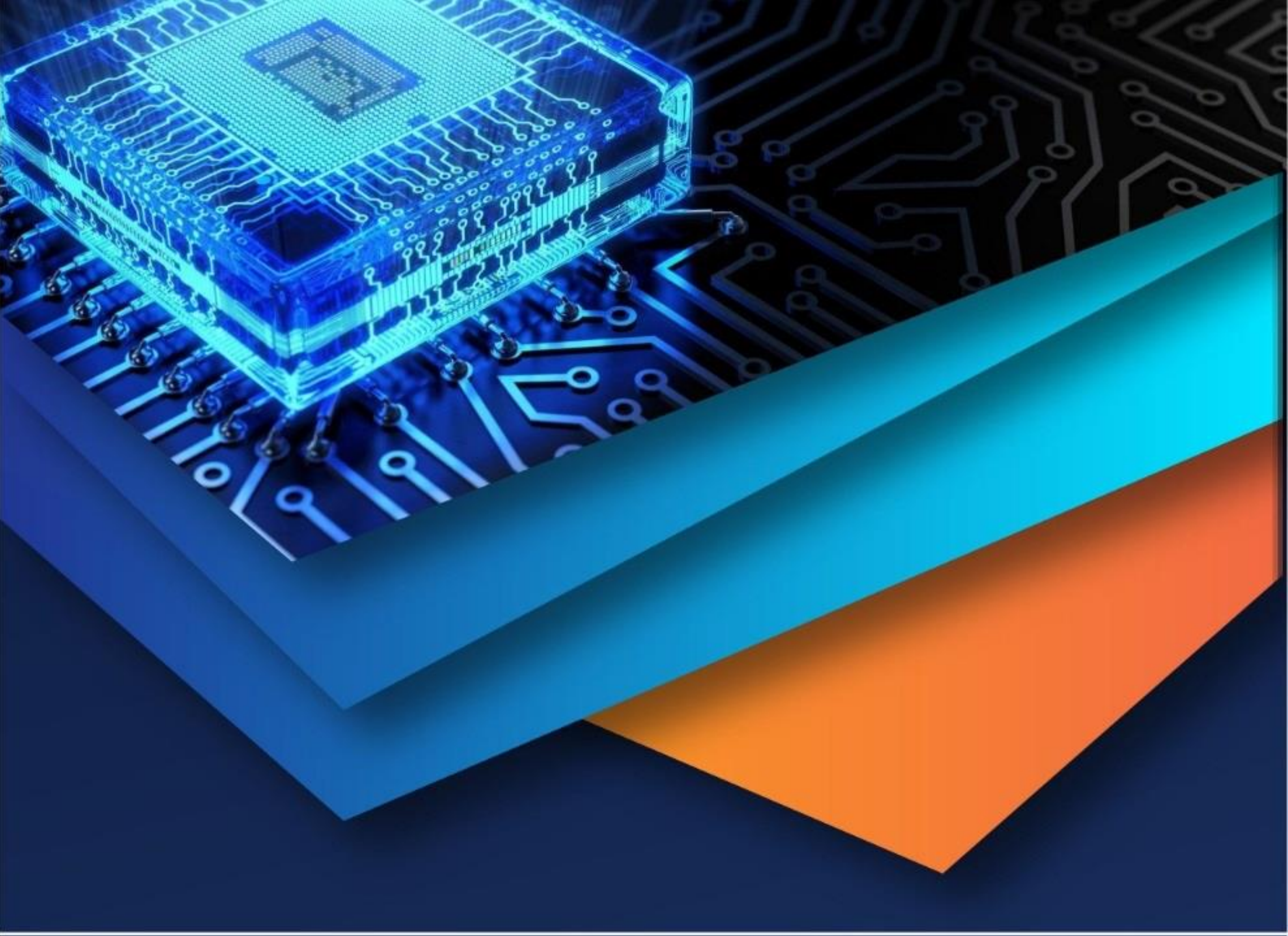
[18] S. Bhattacharya, D. Mukhopadhyay, and S. Giri, "Supply Chain Management in Indian Automotive Industry: Complexities, Challenges and Way Ahead," International Journal of Managing Value and Supply Chains, 2014. doi: 10.5121/ijmvsc.2014.5206.

[19] H. Matsuo, "Implications of the Tohoku earthquake for Toyota's coordination mechanism: Supply chain disruption of automotive semiconductors," International Journal of Production Economics, 2014. doi: 10.1016/j.ijpe.2014.07.010.

[20] R. Aalund, W. Diao, L. Kong, et al., "Understanding the Non-Collision Related Battery Safety Risks in Electric Vehicles a Case Study in Electric Vehicle Recalls and the LG Chem Battery," IEEE Access, 2021. doi: 10.1109/access.2021.3090304.

[21] National Highway Traffic Safety Administration, "Office of Defects Investigation recall flat files," U.S. Department of Transportation, 2026. [Online]. Available: https://static.nhtsa.gov/odi/ffdd/rcl/

[22] U.S. Environmental Protection Agency, "Fuel economy vehicle data," 2026. [Online]. Available: https://www.fueleconomy.gov/feg/epadata/vehicles.csv.zip




10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY