When Every Recall Has a New Root Cause: A Mechanism-Novelty Estimator for Requirement-Level Recurrence in Automotive Safety Compliance

Jherrod Thomas

Independent Researcher — The Lion of Functional Safety

jherrodthomas.com · August 2026

Abstract*—Automotive safety practice treats each noncompliance action as a discrete defect with its own root cause, its own corrective action, and its own effectiveness check. That framing is adequate when a requirement fails once. It is actively misleading when a requirement fails repeatedly through causes that share nothing with one another. Between 2020 and July 2026, Ford filed five separate noncompliance actions against a single Federal requirement — FMVSS No. 111 rear visibility — attributed to printed-circuit-board conductivity, a remedy regression, module thermal shutdown, image inversion, and a menu overlay occluding a mandated test object. Four of the five are causally independent: no corrective action addressing any one of them would have detected or prevented any other. Three were filed while the manufacturer was operating under a three-year NHTSA consent order carrying a $165 million civil penalty whose subject was rear-visibility recall handling. We argue that such a record is not a sequence of defect signals but a single verification signal, and that it is quantitatively readable. We formalize a causal-independence partition over a requirement's recurrence record, then import two estimators from sampling theory: the Good–Turing missing-mass statistic, which estimates the probability that the next action against the requirement arrives through a previously unseen mechanism, and the bias-corrected Chao lower bound, which estimates residual mechanism richness. Four numbered relations convert the record into a novelty rate, a residual-mechanism bound, an effort-reallocation trigger comparing mechanism-level corrective action against requirement-level verification redesign, and an exposure-weighted residual population. Applied to the Ford series the estimators return a novelty rate of unity and a residual bound of at least six undiscovered mechanisms; applied as a contrast to BMW's three-action integrated-brake series they return a novelty rate of zero, diagnosing an entirely different failure — corrective-action ineffectiveness rather than verification incompleteness. Five derived requirements follow, anchored in FMVSS 111 S6.2, ISO 26262-9 freedom-from-interference, and IATF 16949 problem-solving clauses.

Index Terms*—FMVSS 111, ISO 26262, verification coverage, recurrence analysis, Good–Turing estimation, species richness, corrective action effectiveness, combinatorial testing, freedom from interference.*

I Introduction

A safety requirement that fails once tells the engineer about a defect. A safety requirement that fails five times through five unrelated causes tells the engineer about something else entirely, and the vocabulary of automotive quality practice has no word for it. Each of those five failures will be processed as its own event: its own containment, its own root-cause investigation, its own permanent corrective action, its own effectiveness check thirty or ninety days later. Each investigation will conclude correctly. Each corrective action will work. And the requirement will fail again, because nothing in the process ever asked the question that the *sequence* poses rather than any of its members: how large is the space of ways this requirement can be broken, and how much of that space has verification actually visited?

The public record supplies an unusually clean instance. Federal Motor Vehicle Safety Standard No. 111 requires that a rear visibility system display a specified field of view behind the vehicle when the transmission is in reverse, within stated response and linger times, and — the clause that matters most here — as the *default view*, without the driver taking any action to summon it [1]. NHTSA's compliance test procedure places seven cylindrical test objects behind the vehicle and checks whether the displayed image contains them [2]. Between 2020 and July 2026, one manufacturer filed five separate noncompliance actions against this requirement. The image was intermittent because a printed circuit board lacked electrical conductivity. The image lingered after the backing event because vehicles had been repaired incorrectly under a prior campaign. The image vanished for up to five minutes because the module hosting it reached 105°C and entered thermal shutdown, affecting 849,310 vehicles. The image appeared mirrored — picture, buttons, and guidelines all inverted — after an ignition cycle, affecting 889,950 vehicles. And finally, in NHTSA campaign 26V489 filed July 28, 2026, a camera *Views* menu left open before the shift into reverse failed to auto-dismiss and sat on top of the image, partially covering a mandated test object on 47,587 F-150 and F-250 trucks [3].

The last of these is the purest specimen. Nothing crashed. Nothing timed out. The camera worked, the bus worked, the decoder worked, the display worked, and the composite that reached the driver's eye was still noncompliant. Ford's own chronology records competent post-hoc engineering: a Critical Concern Review Group referral, a demonstration that the trigger is a deterministic input sequence rather than a random fault, and a read-across of every SYNC 4 build configuration for display size, orientation, and available camera views that isolated exactly two configurations in which the overlay reached far enough to

obscure a required object [3]. What that chronology also records, read against the four actions preceding it, is that the read-across was performed *after* the fifth failure rather than before the first.

The organizational context sharpens the point rather than softening it. On November 13, 2024, the manufacturer entered a three-year consent order with NHTSA carrying a $165 million civil penalty — the second largest in the agency's history — whose subject was the handling of a rear-visibility recall, and whose performance obligations funded a safety data analytics infrastructure, a multi-modal imaging laboratory for low-voltage electronics, and VIN-based traceability [4]. Three of the five noncompliance actions were filed while that order was in force. The remedial investment was substantial, well-directed at finding defects faster, and structurally incapable of addressing a menu drawn on the wrong compositing layer. Investment in *detection latency* does not reduce *mechanism richness*.

This paper's claim is that the distinction is measurable. We treat a requirement's recurrence record as a sample drawn from an unknown population of mechanisms capable of violating that requirement, and we import the estimators that ecology and computational linguistics developed for exactly this inferential situation: given a sample in which most observed classes appear exactly once, how much of the class space remains unseen? The Good–Turing statistic answers the first form of the question and the Chao lower bound answers the second [5]–[8]. Neither has, to our knowledge, been applied to regulatory noncompliance records, though both have well-established software-engineering analogues in capture–recapture defect estimation [9]–[13].

The contributions are as follows.

- A **causal-independence partition** over a requirement's recurrence record, with an operative test — would a corrective action addressing mechanism A have detected or prevented mechanism B? — that separates novel-mechanism events from remedy regressions and repeat events.
- Four **numbered relations** converting the partitioned record into a mechanism-novelty rate, a bias-corrected residual-mechanism bound, an effort-reallocation trigger comparing mechanism-level corrective action against requirement-level verification redesign, and an exposure-weighted residual population.
- A **two-regime diagnostic** distinguishing novelty-dominated recurrence, which indicts verification coverage, from repeat-dominated recurrence, which indicts corrective-action effectiveness — two conditions that current practice conflates under the single heading "repeat issue."
- A **worked example** on the five-action FMVSS 111 series with the three-action BMW integrated-brake series as a contrast case, yielding five derived requirements traceable to FMVSS 111 S6.2, ISO 26262-9, and IATF 16949.

Section II reviews background and related work. Section III presents the framework. Section IV applies it. Section V discusses limitations and the boundary of the standards lens. Section VI concludes.

II Background and Related Work

A Recall Root-Cause Analysis Stops at the Event

The empirical literature on automotive recalls is overwhelmingly cross-sectional: it classifies a corpus of campaigns by component and defect type and looks for association structure. Chi, Sigmund, and Astardi analyzed 345 NHTSA passenger-vehicle recalls, developed orthogonal classification schemes for defective components and for defect types spanning manufacturing defects, design flaws, and mislabeling, translated each case into functional block diagrams and FMEA statements, and applied Cramér's V and phi-coefficient tests to identify significant component-by-defect associations [14]. The stated purpose is prevention of recurrence, and the method is sound for its question. But the unit of analysis is the campaign, and the association tested is between component class and defect class across manufacturers. The longitudinal structure that this paper centers — one requirement, one manufacturer, N actions ordered in time, partitioned by causal independence — is invisible to that design. NHTSA's own methodological guidance on risk-based defect analysis is likewise organized around detecting and prioritizing individual emerging defects from field signals [15], which is the correct posture for an agency triaging a national complaint stream and the wrong posture for a manufacturer asking whether its verification of a specific clause is complete.

The software-recall literature notes the trend that makes the question urgent. Software and electronics have moved from roughly five percent of all campaigns since 1966 to the single most prevalent recall category in recent reporting years, and the modern instances increasingly involve no component failure at all — the parts perform to specification and the emergent behavior violates the requirement. Rear visibility is a canonical site for this because the requirement is stated over a *rendered composite* while the safety architecture is stated over *components*, a mismatch we return to in Section III-A.

B Estimating What You Have Not Yet Seen

The statistical problem of inferring unobserved class richness from a sample has two mature lineages. Good, working from Turing's wartime analysis, showed that the total probability mass of unseen classes is estimated by the proportion of the sample consisting of singletons — classes observed exactly once [5]. Gale and Sampson supplied the practical smoothed variant that made the estimator usable on sparse linguistic data [6], and Orlitsky, Santhanam, and Zhang later established asymptotic optimality properties for the family [7]. Independently, Chao derived a nonparametric lower bound on total class richness from singleton and doubleton counts, with a bias-corrected form that remains defined when no doubletons are observed — precisely the regime a short recurrence record occupies [8].

Software engineering imported the same machinery through capture–recapture. Vander Wiel and Votta applied capture–recapture to design inspections [9]; Wohlin, Runeson, and Brantestam evaluated it experimentally on code inspections [10]; Briand, El Emam, Freimut, and Laitenberger conducted the comprehensive model comparison that remains the reference

evaluation, characterizing estimator bias across model families and inspection-team sizes [11]; Petersson, Thelin, Runeson, and Wohlin surveyed a decade of theory, evaluation, and application [12]; and later work extended the estimators to naturally occurring defect populations rather than seeded ones [13]. Stringfellow and colleagues addressed the closely related question of estimating post-release defect content in components that showed no defects in testing [16]. The transfer we propose is direct in structure and different in unit: capture–recapture estimates *defects in an artifact* from multiple inspectors; we estimate *mechanisms violating a requirement* from a sequence of regulatory filings, where the "inspectors" are the field, the fleet, and the compliance-test laboratory, and where every observation is by construction a singleton unless a remedy has failed.

Adjacent recent work confirms the framing is live. The reliability-growth tradition models defect discovery as a non-homogeneous Poisson process indexed by testing effort [17], [18], and contemporary extensions add covariate structure and learned discovery curves [19]. Missing-mass reasoning has begun to appear explicitly in machine-learning deployment risk, where a Good–Turing framework has been proposed for quantifying the coverage gap between a training distribution and a deployment distribution [20], and in operational-design-domain coverage arguments for safety-critical AI, where the verification question is likewise "how much of the space did we visit" rather than "did this test pass" [21]. Large-scale empirical study of why bugs escape testing supplies the mechanism-level complement [22].

C Verification Coverage and Interaction Faults

If recurrence indicts verification coverage, the natural next question is which coverage notion. Structural coverage — statement, branch, MC/DC — measures how thoroughly tests exercise implemented code, and safety standards from DO-178C to ISO 26262 use it as a completeness check on requirements-based testing. It is silent about the failure at issue here, because the offending code executed correctly: the menu renderer drew the menu, the camera renderer drew the image, and the compositor composited. What was never tested was the *cross-product of configuration and interaction state* under the compliance condition.

Kuhn, Wallace, and Gallo established the empirical result that governs this space: across several fault corpora, failures were triggered by interactions among relatively few conditions, with pairwise and three-way combinations accounting for the large majority and no observed failure requiring more than six [23]. The Ford read-across is a direct field confirmation — the overlay obscured a required test object only in the intersection of an eight-inch landscape display with a 360-degree camera configuration, a two-factor interaction, and only under a specific three-step input sequence. Combinatorial interaction testing over the declared configuration and HMI-state space is therefore not an exotic recommendation but the indicated technique, and its absence is exactly what a novelty-dominated recurrence record predicts.

The integrity-allocation literature supplies the second half. ISO 26262-9 requires freedom from interference between software elements of differing integrity, and ISO 26262-6 Annex D catalogues the interference classes — timing and execution, memory, and exchange of information [24]. A rendered pixel is a shared resource in the exact sense the clause contemplates: a Quality Management-integrity infotainment menu and a compliance-bearing camera image contend for the same display surface, and if the composite is not itself the subject of a safety requirement, no artifact in the workflow requires anyone to argue that the QM element cannot corrupt it. Mixed-criticality partitioning research has developed the isolation mechanisms [25]; what the recurrence record suggests is missing is not the mechanism but the requirement that would have demanded one.

III Approach

A The Recurrence Record and Its Partition

Let R be a safety requirement with regulatory or safety-goal standing. Let the *recurrence record* of R over an observation window be the time-ordered sequence of noncompliance or defect actions filed against R, written $E = (e_1, \ldots, e_N)$, each e_i carrying an attributed causal mechanism m_i, a filing date, and an affected population p_i.

The record is useless until partitioned, because two actions may name different components while sharing a cause, or name the same component while sharing nothing. We define the partition by an operative counterfactual rather than by taxonomy.

Definition 1 (causal independence). *Mechanisms m_a and m_b violating requirement R are* causally independent *if and only if a complete and correctly executed corrective action addressing m_a, applied at the time m_a was identified, would neither have detected nor have prevented m_b.*

Definition 1 is deliberately counterfactual and deliberately generous to the manufacturer: it asks not whether the two mechanisms are *conceptually* similar but whether the actual remedial work on one would have reached the other. It resolves the cases that taxonomy-based classification handles badly. A remedy regression — an action arising because vehicles were repaired incorrectly under a prior campaign — is *dependent* on that prior campaign by construction, since correct execution of the prior corrective action is precisely what would have prevented it. Conversely, a printed-circuit-board conductivity defect and a compositing-layer defect are independent even though both are labelled "rearview camera image," because no plausible execution of a PCB corrective action inspects the render stack.

Applying Definition 1 partitions E into an *independent-mechanism* subsequence of size N_{ind} over S_{obs} distinct mechanisms, and a *dependent* subsequence of remedy regressions and repeats. Let f_k denote the number of distinct independent mechanisms observed exactly k times in the record.

The partition also localizes each mechanism on the requirement's *realization chain* — the ordered set of stages through which the requirement's satisfaction is physically produced. For a rear visibility requirement the chain runs: image capture, electrical link, decode, geometric transform, composite, display surface, and photons at the driver's eye. Fig. 1 places the observed mechanisms on this chain and makes the paper's central visual claim: the observed mechanisms cluster at stages the verification

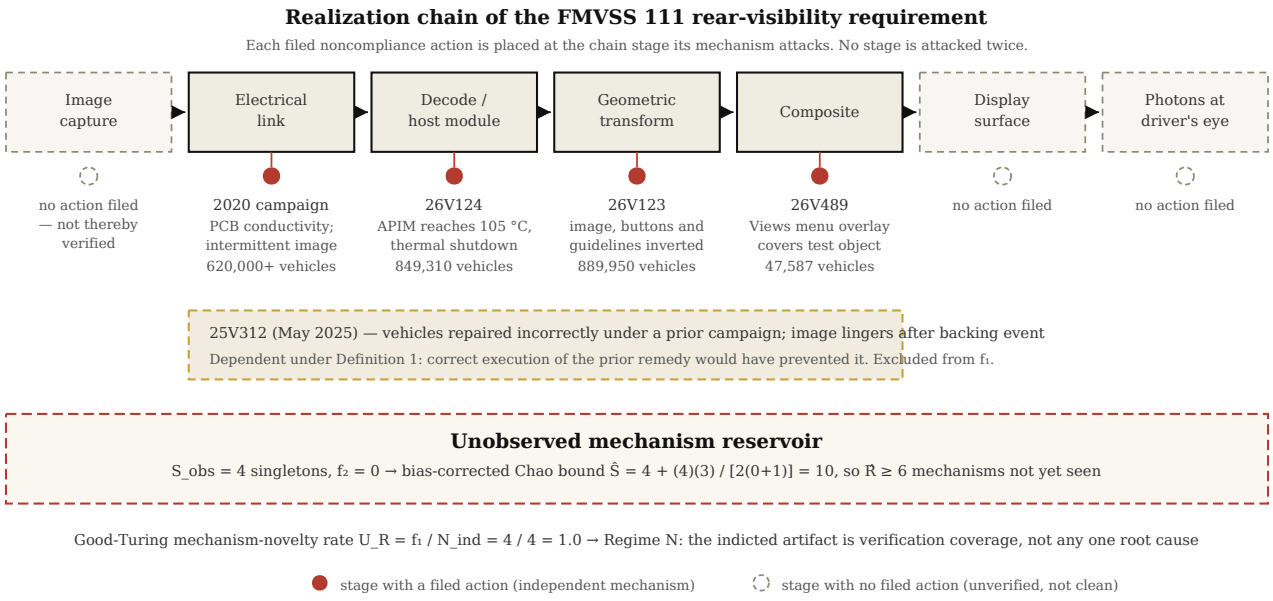

Fig. 1. The realization chain of the FMVSS 111 rear-visibility requirement, with the five filed noncompliance actions placed at the stage each mechanism attacks. Four are causally independent under Definition 1; one (25V312) is a remedy regression on a prior action and is drawn as a dependent event. Stages carrying no observed mechanism are not thereby verified — under a novelty-dominated record they are the reservoir the Chao bound of (2) estimates.

program happened to instrument, and the unvisited stages are where the residual richness lives.

B Mechanism Novelty and Residual Richness

The first quantity of interest is the probability that the *next* action against R arrives through a mechanism never previously seen. Good's estimator gives this as the singleton proportion of the sample [5], [6]:

$$U_R = \frac{f_1}{N_{\text{ind}}} \tag{1}$$

We call U_R the *mechanism-novelty rate* of requirement R. It ranges over the unit interval. A value near zero says the mechanism space is effectively enumerated and future failures will repeat known causes; a value near unity says every failure so far has been a first-of-kind and the next one should be expected to be as well. The estimator's optimality properties in the small-sample, high-novelty regime are what make it usable on records of the length regulatory practice actually produces [7].

The second quantity is the size of the mechanism space itself. The bias-corrected Chao lower bound remains defined when no mechanism has yet recurred, which is the normal condition for a novelty-dominated record [8]:

$$\hat{S}_R = S_{\text{obs}} + \frac{f_1\,(f_1-1)}{2\,(f_2+1)}, \qquad \hat{R} = \hat{S}_R - S_{\text{obs}} \tag{2}$$

where $\hat{R}$ is the estimated number of mechanisms capable of violating R that have not yet been observed. Two properties deserve emphasis. First, $\hat{S}_R$ is a *lower* bound, not a point estimate: the true richness is at least this large, so the quantity is conservative in the direction safety analysis requires. Second, the bound is driven entirely by singletons. A record of five actions with five distinct causes carries far more information about undiscovered richness than a record of five actions with one cause, which is the formal statement of the intuition this paper began with.

C The Reallocation Trigger

Equations (1) and (2) become an engineering decision only when they arbitrate between the two remedial strategies available. Let n be the number of future actions against R under the status quo. A *mechanism-level* strategy — the 8D chain of containment, root cause, permanent corrective action, and prevention of recurrence, executed per action — can prevent only those future actions that repeat an already-identified mechanism, so its expected yield is $(1-U_R)\,n$. A *requirement-level* strategy — redesigning the verification program to cover the realization chain and the configuration-by-state cross-product, and allocating an integrity claim to the requirement's output rather than to its components — prevents a fraction c of all future actions, where c is the coverage the redesigned verification achieves over the mechanism space, including its unobserved part. Requirement-level dominates when

$$c > 1 - U_R \tag{3}$$

Relation (3) is deliberately austere, and its austerity is the point. It contains no free safety factor and no tuning constant. As U_R approaches unity the right-hand side approaches zero and *any* verification redesign with nonzero coverage of the unobserved mechanism space dominates a perfectly executed mechanism-level program — not because the corrective actions are bad, but because they are addressing a population that is not generating the future events. As U_R approaches zero the inequality demands c near unity, which correctly makes requirement-level redesign hard to justify against a record whose failures all repeat: there,

the defect is in execution, not in coverage.

Finally, the residual richness bound converts to exposure. Let $\bar{p}$ be a central estimate of affected population per independent-mechanism action. The *exposure-weighted residual* is

$$\hat{P}_{\text{res}} = \hat{R} \cdot \bar{p} \tag{4}$$

Relation (4) is the crudest of the four and carries the heaviest caveat, developed in Section V: recall populations are heavy-tailed, so $\bar{p}$ is unstable and (4) should be read as an order-of-magnitude exposure statement rather than a forecast. Its function is to put the residual on the same axis — vehicles at risk — that recall decision-making already uses, so that the estimate can enter a prioritization conversation at all.

D The Two-Regime Diagnostic and the Six-Step Audit

Together (1)–(3) define two regimes that current practice conflates under the phrase "repeat issue."

Regime N (novelty-dominated), U_R above the trigger. The mechanism space is large relative to what verification has visited. Each root-cause investigation is correct and each corrective action is effective, and the requirement keeps failing. The indicted artifact is the verification program's coverage of the realization chain and the configuration cross-product, and the indicted allocation is usually an integrity claim placed on components rather than on the requirement's output.

Regime R (repeat-dominated), U_R at or near zero. The mechanism space is small and known. The requirement keeps failing through the *same* cause. The indicted artifacts are corrective-action effectiveness verification, production conformity, and the escape-point analysis that should have prevented remedy regressions — IATF 16949's problem-solving and corrective-action clauses and ISO 26262-7's production and field-monitoring provisions [24], [26], not the verification program.

The audit runs in six steps, consistent with the series this paper extends. **S1 — record assembly:** collect every action filed against R over the window from the regulatory database, with attributed mechanism, date, and population. **S2 — independence partition:** apply Definition 1 pairwise; record the counterfactual justification for each dependent classification, because that justification is the audit's most contestable content. **S3 — chain localization:** place each independent mechanism on the requirement's realization chain and mark stages with no observed mechanism. **S4 — estimation:** compute (1), (2), and (4). **S5 — regime assignment and trigger:** evaluate (3) against a defensible estimate of achievable verification coverage c, and assign Regime N or R. **S6 — requirement derivation:** derive requirements against the indicted artifact for the assigned regime — verification-program and integrity-allocation requirements in Regime N, effectiveness-verification and production-conformity requirements in Regime R.

IV Worked Example: Two Recurrence Records

A The FMVSS 111 Rear-Visibility Series

We instantiate the framework on the five-action rear-visibility record summarized in Table I. In S1, the requirement R is FMVSS No. 111 rear visibility, specifically the S6.2 family covering image content, response and linger time, and default view, as exercised by compliance test procedure TP-111V-01 [1], [2]. The window runs from the 2020 campaign through the July 28, 2026 filing of 26V489 [3].

In S2, Definition 1 yields $N_{\text{ind}} = 4$ independent actions over $S_{\text{obs}} = 4$ distinct mechanisms, so $f_1 = 4$ and $f_2 = 0$, with action 2 removed to the dependent subsequence as a remedy regression. We stress that this partition is conservative: classifying action 2 as independent would raise both estimators, and we decline that because the counterfactual test does not support it. In S3, the four independent mechanisms localize to four *different* stages of the realization chain — electrical link, host module thermal behavior, geometric transform, and composite — with no stage visited twice. This is the structural signature of Regime N, visible before any arithmetic.

In S4, relation (1) gives $U_R = 4/4 = 1.0$. Every independent action to date has introduced a first-of-kind mechanism, so the estimated probability that the next action does the same is unity — an estimate that is obviously saturated at the boundary and should be read as "at least one region of near-certainty," not as a claim that novelty is literally certain. Relation (2) gives $\hat{S}_R = 4 + (4)(3)/[2(0+1)] = 10$, hence $\hat{R} \geq 6$ undiscovered mechanisms. Relation (4), taking $\bar{p}$ as the mean of the four independent-action populations, approximately 6.0×10^5 vehicles, returns $\hat{P}_{\text{res}} \approx 3.6 \times 10^6$ vehicles of residual exposure; the median-based figure is approximately 4.4×10^6, and the gap between the two is itself the evidence for the heavy-tail caveat we develop in Section V.

In S5, relation (3) reduces to $c > 0$. Any verification redesign achieving nonzero coverage of the unobserved mechanism space dominates a mechanism-level program, however well the latter is executed. This is the framework's sharpest and most uncomfortable output, and it is worth stating plainly what it does and does not say. It does not say the five root-cause investigations were wrong; the public record indicates they were careful, and the read-across performed for 26V489 is a model of the technique [3]. It says that a remedial portfolio consisting entirely of such investigations has an expected yield against future actions of $(1 - 1.0)\,n = 0$, because the population those investigations address is not the population generating the events. It also explains, without recourse to any claim about the manufacturer's diligence, why $165 million of consent-order investment in analytics, imaging, and traceability [4] coincided with three further filings: every funded capability improves detection latency and defect attribution, and none of them enumerates the mechanism space.

B Contrast Case: The Integrated-Brake Series

The framework earns its two-regime structure only if it distinguishes. We apply the same estimators to a second 2026 record: a manufacturer's third recall in twenty-nine months on the same supplier-provided integrated braking unit, a campaign that explicitly re-opens vehicles already repaired under two prior campaigns. Here Definition 1 does the opposite work. The three actions are *not* independent — correct and complete execution of the first campaign's corrective action is precisely what would

TABLE I
RECURRENCE RECORD FOR FMVSS NO. 111 REAR VISIBILITY, ONE MANUFACTURER, 2020 THROUGH JULY 2026

#	Action	Attributed mechanism	Chain stage	Clause stressed	Population	Independent?
1	2020 camera campaign	Insufficient PCB electrical conductivity; intermittent or inoperative image	Electrical link	S6.2.1 image / availability	620,000+	Yes (baseline)
2	25V312 (May 2025)	Vehicles repaired incorrectly under a prior campaign; image persists after backing event ends	Remedy execution	S6.2.4 linger time	2021–22 Bronco	**No** — regression
3	26V124 (Mar. 6, 2026)	APIM reaches 105°C and enters thermal shutdown up to five minutes; no image in reverse	Decode / host module	S6.2.1, S6.2.3	849,310	Yes
4	26V123 (Mar. 2026)	Displayed image flipped or inverted after ignition cycle; image, buttons, guidelines mirrored	Geometric transform	S6.2.1 geometry	889,950	Yes
5	26V489 (Jul. 28, 2026)	Camera Views menu overlay does not auto-dismiss into reverse; covers a required test object	Composite	S6.2.1, S6.2.6 default view	47,587	Yes

The independence column applies Definition 1. Mechanism 2 is classified dependent because correct execution of the prior campaign's remedy is exactly what would have prevented it; the remaining four are mutually independent because no corrective action on any one inspects the artifact the others attack. Populations are as reported in the public record.

TABLE II
TWO-REGIME DIAGNOSTIC APPLIED TO THE TWO 2026 RECURRENCE RECORDS

Quantity	FMVSS 111 rear-visibility series	Integrated-brake series
Actions in window / N_{ind}	5 / 4	3 / 3
Distinct mechanisms S_{obs}	4	1
Singletons f_1 / doubletons f_2	4 / 0	0 / 0
Novelty rate U_R, eq. (1)	1.0	0.0
Residual richness $\hat{R}$, eq. (2)	≥ 6	0
Trigger, eq. (3)	$c > 0$ — satisfiable by any redesign	$c > 1$ — unsatisfiable
Regime	**N** — verification coverage	**R** — corrective-action effectiveness
Indicted artifact	Chain and cross-product coverage; integrity allocation on the requirement's output	Effectiveness verification; production conformity; escape-point analysis

Identical estimators, applied to records of comparable length, return opposite diagnoses and route to disjoint remedial artifacts. Current practice describes both records with the same phrase — "repeat issue" — and routes both to the same 8D chain.

have prevented the second and third — so the partition yields $S_{\text{obs}} = 1$ mechanism observed three times, $f_1 = 0$, $f_2 = 0$.

Relation (1) gives $U_R = 0$. Relation (2) gives $\hat{S}_R = 1 + 0 = 1$, hence $\hat{R} = 0$: the record supplies no evidence of undiscovered mechanism richness. Relation (3) demands $c > 1$, which is unsatisfiable, correctly rejecting verification redesign as the indicated remedy. The diagnosis is Regime R, and the indicted artifacts are entirely different: the corrective-action effectiveness verification that should have closed after the first campaign, the production-conformity provisions of ISO 26262-7, and the escape-point analysis that IATF 16949's problem-solving clause requires [24], [26]. Table II summarizes the contrast.

C Derived Requirements

Regime N assignment for the rear-visibility record routes to five derived requirements, listed in Table III. They are written to be verifiable and to attach to artifacts that already exist in an IATF 16949 and ISO 26262 workflow.

Requirement RV-105 is the one that generalizes beyond rear visibility, and it is deliberately written as a *blocking* condition rather than an analysis obligation. The failure mode the framework diagnoses is not that nobody looked at the recurrence record; it is that looking at it produced a series of correct per-event conclusions and no aggregate one. A gate that refuses to close on mechanism-level corrective action alone, when the novelty rate says mechanism-level corrective action cannot help, is the minimum organizational translation of relation (3).

V DISCUSSION

A Limitations and Threats to Validity

The most serious threat is that Definition 1 is a judgment, and the estimators are extremely sensitive to it. Reclassifying the remedy regression as independent would give $f_1 = 5$ and $\hat{S}_R = 15$, raising $\hat{R}$ from 6 to 10 — a 67% swing from one analyst decision. We have mitigated this by requiring that every dependent classification record its counterfactual justification (step S2) and by choosing the conservative direction at the one contestable case, but the framework cannot escape the fact that causal independence is adjudicated by the same organization whose verification program is under examination. Independent confirmation review of the partition, in the ISO 26262-8 sense, is the obvious safeguard and is not something we can supply from the public record.

TABLE III
DERIVED REQUIREMENTS FOR A REGIME N RECURRENCE RECORD

ID	Derived requirement	Verification method	Anchor
RV-101	A safety requirement shall be allocated to the rendered composite delivered to the display surface during a backing event, not solely to the components producing it. The composite shall be the subject of the compliance claim.	Safety-requirement review; traceability from FMVSS 111 S6.2 to a requirement whose object is the composite	FMVSS 111 S6.2.1, S6.2.6 [1]; ISO 26262-3 [24]
RV-102	Any element of lower integrity sharing the display surface with the compliance-bearing image shall be subject to a documented freedom-from-interference argument covering exchange of information and shared-resource contention.	FFI analysis per ISO 26262-6 Annex D; partitioning evidence	ISO 26262-9 §6; ISO 26262-6 Annex D [24], [25]
RV-103	Compliance verification shall execute over the declared cross-product of display size, orientation, available camera views, and entry HMI state, at minimum to pairwise strength, rather than from the default HMI state alone.	Combinatorial interaction test suite on HIL; documented factor model and strength justification	[23]; ISO 26262-8 §9 [24]
RV-104	A rendered-frame monitor shall verify at runtime that the mandated field of view is unobstructed in the delivered frame — for example by fiducial regions that must remain unoccluded — and shall annunciate on failure.	Fault injection against the compositor; monitor coverage measurement	FMVSS 111 S6.2.1 [1], [2]; ISO 26262-4
RV-105	For every requirement with two or more noncompliance actions in a rolling window, the mechanism-novelty rate and residual-richness bound shall be computed and reviewed, and a Regime N assignment shall block closure of mechanism-level corrective action as the sole remedy.	Management review record; recurrence register with independence justifications	IATF 16949 §10.2.3 [26]; ISO 26262-7 [24]

Second, the estimators assume the sample is drawn from a fixed mechanism population. A vehicle program under active software development does not have a fixed population: each release can *create* mechanisms, so the observed richness partly reflects mechanism generation rather than mechanism discovery. This biases $\hat{S}_R$ in an unknown direction — downward if the population is growing faster than sampling, upward if retired configurations have removed mechanisms that the record still counts. The reliability-growth literature handles the analogous non-stationarity with explicitly time-indexed intensity functions [17]–[19], and coupling the Chao bound to such a model is the natural formal repair.

Third, the sample sizes are very small. Four independent observations is a thin basis for any richness estimate, and the bias-corrected Chao form was designed for defensibility rather than precision in exactly this regime [8]. We therefore claim only what a lower bound licenses: at least six mechanisms remain, not approximately six. Relation (4) is weaker still, because recall populations span more than an order of magnitude within this single record — 47,587 against 889,950 — so the mean is dominated by its largest terms and the mean-versus-median gap reported in Section IV-A is the honest expression of the uncertainty. $\hat{P}_{\text{res}}$ should enter a prioritization discussion and should not enter a risk calculation.

Fourth, our reconstruction of mechanisms and independence relations is external, assembled from Part 573 filings and contemporaneous reporting [3], [4]. The internal engineering record may contain read-across analyses, coverage arguments, and integrity allocations we cannot see. The framework's claim is precisely that such artifacts should be *visible and gating* — that a recurrence register with computed novelty rate should exist and should block a closure decision — not that no engineer ever noticed the pattern.

Finally, the coverage parameter c in relation (3) is not estimated in this paper. We treat the trigger qualitatively because at $U_R = 1$ the inequality resolves without needing c; at intermediate novelty rates the trigger becomes genuinely sensitive to a quantity that would require measured combinatorial coverage against a declared factor model to estimate. That measurement is the most valuable piece of future empirical work the framework implies.

B *Where the Standards Lens Stops*

Neither FMVSS 111 nor ISO 26262 forbids any of this analysis, and neither requires it. FMVSS 111 states an outcome over a rendered image and is silent on architecture, which is proper for a performance standard; it is the manufacturer's business how the photons get there, and the standard's very silence is what allows a compliance-bearing output to be produced by a Quality Management-integrity stack without anyone violating a clause. ISO 26262 supplies the freedom-from-interference machinery that would govern the display surface [24], but only once someone has allocated a safety requirement to the composite — and if the hazard analysis never treated the rendered frame as a safety-related output, Part 9 is never invoked. IATF 16949 requires problem solving and corrective-action effectiveness [26], but its unit is the nonconformity, so a series of correctly closed 8Ds is fully conformant and the aggregate signal has no clause to be reported under.

The framework's own lens stops in three places. It says nothing about *why* a mechanism space is large — whether the driver is architectural coupling, supplier fragmentation, or configuration proliferation — and two programs with identical novelty rates may need very different redesigns. It has no treatment of security-induced recurrence, where an adversary rather than a defect process generates the mechanism sequence and the inde-

pendence assumption underlying the estimators fails outright; that case belongs to ISO/SAE 21434. And it cannot make the organizational decision it recommends. The estimators can show that mechanism-level corrective action has zero expected yield against future actions, and only a release process willing to treat that number as blocking can act on it.

VI Conclusion and Future Work

Five noncompliance actions against one Federal rear-visibility requirement, four of them causally independent and each localized to a different stage of the requirement's realization chain, three of them filed under a consent order whose subject was that same requirement: the record is not five defect signals but one verification signal, and it is readable with estimators that sampling theory has had for seventy years. We have defined a causal-independence partition over a requirement's recurrence record, applied the Good–Turing missing-mass statistic and the bias-corrected Chao richness bound to the partitioned record, and derived an effort-reallocation trigger that compares mechanism-level corrective action against requirement-level verification redesign on the single axis of expected future actions prevented. The rear-visibility record returns a novelty rate of unity, a residual bound of at least six undiscovered mechanisms, and a trigger satisfiable by any verification redesign with nonzero coverage. A contrast record — three campaigns on one braking unit, none of them independent — returns a novelty rate of zero and routes to entirely different artifacts, demonstrating that the diagnostic discriminates rather than merely alarms.

Three directions of future work follow. First, empirical: the estimators should be run at scale across the NHTSA campaign corpus, grouped by requirement and manufacturer, to establish base rates for the novelty rate and to test whether Regime N assignment predicts subsequent filings — a falsifiable claim this paper makes but cannot test from a single record. Second, formal: coupling the Chao bound to a time-indexed intensity model would address the non-stationarity threat of Section V-A and would let the residual bound carry a stated confidence rather than a bare inequality. Third, methodological: measuring the coverage parameter c against declared configuration-and-state factor models would make relation (3) quantitative at intermediate novelty rates, where the interesting decisions actually live, and would connect the framework to the combinatorial testing literature that already knows how to bound interaction faults [23].

The broader observation is a small one about vocabulary. Automotive quality practice has an excellent language for the defect and no language for the defect *series*. Every artifact in the workflow — the 8D, the effectiveness check, the Part 573 report — takes a single event as its subject, and every one of them can be executed perfectly while the requirement continues to fail. Giving the series a name, a statistic, and a gate is a modest change to the paperwork and a substantial change to what the paperwork can notice.

References

[1] *Electronic Code of Federal Regulations, Title 49 §571.111 — Standard No. 111; Rear visibility*. [Online]. Available: https://www.ecfr.gov/current/title-49/section-571.111

[2] National Highway Traffic Safety Administration, *Laboratory Test Procedure for FMVSS No. 111, Rear Visibility*, TP-111V, Office of Vehicle Safety Compliance, Washington, DC, USA. [Online]. Available: https://www.nhtsa.gov/laws-regulations/fmvss

[3] Ford Motor Company, "Part 573 Safety Recall Report 26V-489: Obstructed Rearview Camera Image / FMVSS 111 (Ford reference 26C37)," National Highway Traffic Safety Administration, Washington, DC, USA, Jul. 28, 2026. [Online]. Available: https://www.nhtsa.gov/recalls?nhtsaId=26V489

[4] National Highway Traffic Safety Administration, "NHTSA announces consent order with Ford, $165 million civil penalty," Press release, Washington, DC, USA, Nov. 13, 2024. [Online]. Available: https://www.nhtsa.gov/press-releases/ford-consent-order-165-million-civil-penalty

[5] I. J. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, vol. 40, no. 3–4, pp. 237–264, Dec. 1953, doi: 10.1093/biomet/40.3-4.237.

[6] W. A. Gale and G. Sampson, "Good–Turing frequency estimation without tears," *J. Quant. Linguistics*, vol. 2, no. 3, pp. 217–237, 1995, doi: 10.1080/09296179508590051.

[7] A. Orlitsky, N. P. Santhanam, and J. Zhang, "Always Good Turing: Asymptotically optimal probability estimation," *Science*, vol. 302, no. 5644, pp. 427–431, Oct. 2003, doi: 10.1126/science.1088284.

[8] A. Chao and C.-H. Chiu, "Species richness: Estimation and comparison," in *Wiley StatsRef: Statistics Reference Online*. Chichester, U.K.: Wiley, 2016, pp. 1–26, doi: 10.1002/9781118445112.stat03432.pub2.

[9] S. A. Vander Wiel and L. G. Votta, "Assessing software designs using capture-recapture methods," *IEEE Trans. Softw. Eng.*, vol. 19, no. 11, pp. 1045–1054, Nov. 1993, doi: 10.1109/32.256852.

[10] C. Wohlin, P. Runeson, and J. Brantestam, "An experimental evaluation of capture-recapture in software inspections," *Softw. Test. Verif. Rel.*, vol. 5, no. 4, pp. 213–232, Dec. 1995, doi: 10.1002/stvr.4370050403.

[11] L. C. Briand, K. El Emam, B. G. Freimut, and O. Laitenberger, "A comprehensive evaluation of capture-recapture models for estimating software defect content," *IEEE Trans. Softw. Eng.*, vol. 26, no. 6, pp. 518–540, Jun. 2000, doi: 10.1109/32.852741.

[12] H. Petersson, T. Thelin, P. Runeson, and C. Wohlin, "Capture-recapture in software inspections after 10 years research — Theory, evaluation and application," *J. Syst. Softw.*, vol. 72, no. 2, pp. 249–264, Jul. 2004, doi: 10.1016/S0164-1212(03)00090-6.

[13] C. Andersson, T. Thelin, P. Runeson, and N. Dzamashvili-Fogelström, "Evaluation of capture-recapture models for estimating the abundance of naturally-occurring defects," in *Proc. 2nd ACM-IEEE Int. Symp. Empirical Softw. Eng. Meas. (ESEM)*, Kaiserslautern, Germany, 2008, pp. 200–209, doi: 10.1145/1414004.1414031.

[14] C.-F. Chi, D. Sigmund, and M. O. Astardi, "Classification scheme for root cause and failure modes and effects analysis (FMEA) of passenger vehicle recalls," *Rel. Eng. Syst. Saf.*, vol. 200, art. 106929, Aug. 2020, doi: 10.1016/j.ress.2020.106929.

[15] National Highway Traffic Safety Administration, *Risk-Based Processes for Safety Defect Analysis and Management of Recalls*, Report No. DOT HS 812 984, Washington, DC, USA, Nov. 2020. [Online]. Available: https://www.nhtsa.gov/document/risk-based-processes-safety-defect-analysis-and-management-recall

[16] C. Stringfellow, A. Andrews, C. Wohlin, and H. Petersson, "Estimating the number of components with defects post-release that showed no defects in testing," *Softw. Test. Verif. Rel.*, vol. 12, no. 2, pp. 93–122, Jun. 2002, doi: 10.1002/stvr.235.

[17] A. L. Goel and K. Okumoto, "Time-dependent error-detection rate model for software reliability and other performance measures," *IEEE Trans. Rel.*, vol. R-28, no. 3, pp. 206–211, Aug. 1979, doi: 10.1109/TR.1979.5220566.

[18] K. Y. Song, I. H. Chang, and H. Pham, "A software reliability growth model assuming uncertain operating environments and dependent failure occurrences," *Ann. Oper. Res.*, 2026, doi: 10.1007/s10479-026-07079-z.

[19] M. Salboukh, L. Fiondella, and V. Nagaraju, "Predicting software defect discovery incorporating covariates with recurrent neural networks," *Qual. Rel. Eng. Int.*, 2026, doi: 10.1002/qre.70063.

[20] "Blind-spot mass: A Good–Turing framework for quantifying deployment coverage risk in machine learning systems," arXiv:2604.05057, Apr. 2026. [Online]. Available: `https://arxiv.org/abs/2604.05057`

[21] "From high-dimensional spaces to verifiable ODD coverage for safety-critical AI-based systems," arXiv:2604.02198, Apr. 2026. [Online]. Available: `https://arxiv.org/abs/2604.02198`

[22] "What makes software bugs escape testing? Evidence from a large-scale empirical study," arXiv:2604.26672, Apr. 2026. [Online]. Available: `https://arxiv.org/abs/2604.26672`

[23] D. R. Kuhn, D. R. Wallace, and A. M. Gallo, Jr., "Software fault interactions and implications for software testing," *IEEE Trans. Softw. Eng.*, vol. 30, no. 6, pp. 418–421, Jun. 2004, doi: 10.1109/TSE.2004.24.

[24] International Organization for Standardization, *ISO 26262: Road Vehicles — Functional Safety*, Parts 1–12, 2nd ed., Geneva, Switzerland, 2018.

[25] "Jiao: Bridging isolation and customization in mixed criticality robotics," arXiv:2605.03641, May 2026. [Online]. Available: `https://arxiv.org/abs/2605.03641`

[26] International Automotive Task Force, *IATF 16949:2016 — Quality Management System Requirements for Automotive Production and Relevant Service Parts Organizations*, 1st ed., Oct. 2016.