FLEETIDY · TECHNICAL METHODOLOGY

The FRED Carrier-Scoring Engine

A complete specification of the FRED scoring engine: how a motor carrier's forward crash frequency is estimated from public FMCSA data as three named factors, graded with posterior screens, attributed week to week, and gated before every write

---

Document FRED Scoring — Mathematical White Paper

Audience Underwriting, actuarial, and executive readers

Population 1,824,374 property carriers graded · 107,352 in the ranked 10+ power-unit book (2026-11-01 run: inspections through 2026-09-30, crashes through 2026-08-31)

Engine Class rate × violation relativities × Gamma–Poisson crash experience, every constant from one frozen model artifact

Date November 2026

§0 Executive Summary

The **FRED score** answers one question for every property motor carrier registered with FMCSA: *how many recordable crashes is this carrier likely to generate per power-unit-year over the coming twelve months, and how does that compare with a carrier of its class?* The answer is built entirely from public FMCSA records and refreshed weekly across the whole property universe — for-hire, private and exempt, interstate and intrastate — about 1.82 million carriers, of which the 107,352 with ten or more power units form the ranked underwriting book.

The engine is the product of the Safety Core Study, a stand-alone study of carrier crash-risk modelling on public data in which FMCSA's own measures appear only as benchmarks. Its prediction is the product of **three named factors**: a **class rate** μ_{class} per power-unit-year from a Poisson regression on the carrier's size, class, tenure, filing recency, how much of its filed fleet the roadside record sees, and road activity (§4); a **violation relativity** V from three credibility-shrunk ratios — driver behaviour, and out-of-service and violation ratios standardised for where the carrier was inspected (§5); and a **crash-experience** term $\bar{\theta}$, the Gamma–Poisson posterior mean of the carrier's own recordable crashes over 36 months, decayed on a linear ramp, against the crashes the first two factors predict (§6). Their product gives the **relativity to class** $\bar{R} = V\bar{\theta}$ and the forward expectation $\mu_{\text{class}}\bar{R}$ per power-unit-year (§7).

The grade is a band from fitted cuts on $\bar{R}$, then screens: the top tiers require proof (60% of the posterior below the cut, a driver-behaviour relativity no worse than class, no fatal crash, enough inspections, a plausible fleet), the bottom tiers require observed events and the same 60% posterior test in the other direction, so a wide posterior falls back toward the class rate whichever way it points (§8); a Satisfactory whose 80% range still runs from Excellent to Poor territory is shown as Not enough record, with no score (§8). Every weekly change is attributed to one of six causes and improvements are ramped one band per run (§9); nineteen acceptance checks run before anything is written (§10). Out of sample the engine explains 28.2% of forward crash deviance with an exposure-weighted normalized Gini of 0.653, 0.021 behind a gradient-boosted ceiling while remaining fully decomposable; out of time, in the latest run's gate year, its grades realize 0.48, 0.63, 0.86, 1.17, 1.66 and 3.18 times the class expectation, Excellent to Critical, in order (§8, §11).

> **Three numbers, one model.** For each carrier the engine emits (i) an expected **crash count** over the next twelve months with a 10th–90th percentile range, (ii) a **grade and 0–100 score** from its relativity to class $\bar{R}$, and (iii) the **three factors** with their values, the credibility behind them, the screens that fired, and the cause of the latest change. All of it rests on one posterior, so the pieces tell one story.

§ Contents

1 What the score estimates
2 Data, universe and timing
3 Exposure and size cohorts
4 Factor 1 — the class rate
5 Factor 2 — violation relativities
6 Factor 3 — crash experience
7 Relativity to class and forward expectation
8 From relativity to grade
9 Change attribution and the governor
10 Acceptance gates
11 Validation: the study and the live run
12 Severity display and what is not in the grade
13 Carriers, end to end
A Symbols and constants
B Reporting lag and the cutoffs

§1 What the Score Estimates

For carrier i at a cutoff date C, the target is the number of **federally recordable crashes** the carrier generates in the twelve months after C, per power-unit-year of exposure. Writing y_i for the realized count and E_i for forward exposure, the engine estimates the rate λ_i with $y_i \sim \text{Poisson}(\lambda_i E_i)$:

$$\lambda_i \;=\; \mu_{\text{class}}(x_i)\cdot V_i\cdot \bar\theta_i, \qquad \bar R_i \;=\; V_i\,\bar\theta_i \;=\; \frac{\lambda_i}{\mu_{\text{class}}(x_i)}.$$

μ_{class} is what a carrier with this one's attributes is expected to do; $\bar R$ is how far this carrier's own record moves it from that expectation. $\bar R = 1$ means “typical for a carrier of this class”; 0.5 means half the crashes, 2 means twice. The grade reads $\bar R$; the carrier page also shows the absolute expectation $\lambda_i \cdot \mathrm{PU}_i$ as expected crashes over the next twelve months.

The target is **frequency**, not a severity-weighted burden. The study's variable-importance work found severity weights in the frequency model to be noise, and a frequency target is what the Gamma–Poisson experience term is exact for. Severity is modelled separately for display and loss-cost work (§12).

Five properties were fixed in the study charter before any model was fit, and the design below is the attempt to satisfy them: (1) a thin or unreliable record sits near its class rate; (2) the top tier requires proof; (3) the bottom tier requires observed adverse events, never attributes alone; (4) events decay smoothly to zero weight; (5) no grade jumps upward without a cause that can be named.

---

§2 Data, Universe and Timing

All inputs are FMCSA files published on data.transportation.gov, downloaded weekly and archived point-in-time: the company census (power units, class, operation type, hazmat, domicile, registration date, MCS-150 filing date and mileage, safety rating), the MCMIS crash file (full history, with the date FMCSA added each record), the MCMIS inspection file (36 months) and its inspection-violation detail, SMS inspection units (vehicle VINs), the licensing and insurance history and out-of-service orders (exits), and the SMS violation lookup that maps each violation code to its BASIC. FMCSA BASIC percentiles are never read: the public SMS carries none for property carriers, and they would leak the outcome into the inputs.

The **universe** is every property carrier — for-hire, private and exempt; interstate and intrastate — with at least one power unit; operation type is a class-rate feature, not an exclusion. Passenger strata (buses, coaches, school buses, vans, limousines) are held out for their own class models. A carrier under an out-of-service order at the cutoff is excluded from grading, as is one whose operating authority was revoked with no roadside inspection or crash recorded since and no active authority on the current FMCSA file (activity or a reinstatement cancels the exit). A carrier registered on or after C has no exposure inside a window yet, a census record outside the population at the cutoff (inactive, no power units or no forward exposure) has nothing to grade, and a filing of more than 5,000 power units lies above the model's fitted size range (§3); each is shown with the reason instead of a grade. The 2026-11-01 run graded 1,824,374 property carriers. The reasons are written for the carriers of the site's searchable census, a list that differs from the graded universe (a graded carrier outside the census is not searchable), and 148,164 census carriers were shown with a reason instead of a grade: 100,630 operating authority revoked, no roadside activity since; 34,209 passenger operation — outside the property-carrier model; 8,017 registered after the scoring cutoff; 4,606 under a federal out-of-service order; 658 fleet above the model's fitted size range (5,000+ power units); 44 outside the scoring population at the cutoff.

Cutoffs and windows

Each weekly run at run date R uses two cutoffs. C is the first of the month containing the knowledge date $\min(R,\ \text{data-as-of} + 1\ \text{day})$, so the run reads whole months the FMCSA files cover. Inspections, violations, the census facts and the VIN count are read through C; crashes stop one month earlier, at $C_{\text{crash}} = C - 1\ \text{month}$, because crash reports reach FMCSA weeks late. The run reads the records on FMCSA's file at the data date and places each by the date it happened — a crash by its report date, an inspection by its inspection date — so a record that reaches the file late enters the window at its own date in the next weekly run; the validation replays of Appendix B read the add dates instead, to reproduce what each run could have known. The 2026-11-01 run used C = 2026-10-01 and C_{crash} = 2026-09-01 with data as of 2026-10-29. The windows are half-open and in calendar months:

$$\underbrace{[\,C_{\text{crash}} - 36\ \text{m},\ C_{\text{crash}}\,)}_{\text{crash window (36-month ramp)}} \qquad \underbrace{[\,C - 24\ \text{m},\ C\,)}_{\text{inspection, violation and VIN window}} \qquad \underbrace{[\,C,\ C + 12\ \text{m}\,)}_{\text{forward outcome (validation only)}}$$

Why this timing: nationally 85% of crashes are known to FMCSA within 60 days (67.7% within 30, 92.5% within 90) and completeness is strongly state-dependent — Colorado reached only 25% at 60 days in the study's measurement — while inspection records arrive within days to weeks. The reporting-hold study (Appendix B) replayed 13 historical monthly runs, each with only the records it would have had, and judged five timings on the crashes carriers had afterwards. Every timing that counted more recent events beat holding everything back two months on all 13. Counting crashes right up to the run date gained most in normal months but drifted generous when FMCSA's crash reporting stalled in mid-2025; holding crashes one month kept the longer hold's calibration and grade mix. A state-specific completeness factor fitted from the weekly archive is the planned refinement.

Table 2.1 — Core notation.

Symbol	Meaning
$C,\ C_{\text{crash}}$	Cutoffs: C = first of the month containing $\min(R,\ \text{data-as-of} + 1\ \text{d})$; inspections, violations, census and VINs through C. Crashes through $C_{\text{crash}} = C - 1$ month
PU_{floor}	Exposure power units: max(filed power units, power-unit VINs observed in 24 months) (§3)
c	Fleet-size cohort of the carrier: one of nine (§3)
$\mu_{\text{class}}(x)$	Class rate: expected recordable crashes per power-unit-year from the frozen Poisson GLM (§4)
$\tilde{r}_s,\ V$	Shrunk relativity of violation signal s; violation relativity $V = \prod_s \tilde{r}_s^{\gamma_s}$ (§5)
$N,\ \Lambda$	Decayed recordable crashes in 36 months; decayed class-and-violation-expected crashes (§6)
$a,\ \bar{\theta},\ Z$	Credibility prior $a(c, \Lambda)$ for the cohort and the record's expected crashes; experience relativity $(a + N)/(a + \Lambda)$; credibility $\Lambda/(a + \Lambda)$ (§6)
$\bar{R}$	Relativity to class $V\bar{\theta}$: the graded statistic, the posterior mean (§7)
$q_{10},\ q_{90}$	10th and 90th percentiles of the posterior for R (§7)

§3 Exposure and Size Cohorts

Crash risk is rated per **power-unit-year**, and the filed power-unit count alone is a poor measure of it. Counting distinct power-unit VINs seen at inspections and crashes in the trailing 24 months (trailers excluded — the second unit of an inspection is usually a trailer), the study measured realized ÷ predicted crashes under a cohort-rate model:

Table 3.1 — Forward crashes against a filed-PU prediction, by fleet-trust group (select split, 2025-08 origin).

Group	Carriers	Realized ÷ predicted
Observed fleet consistent with filing	54,720	1.08
More power units observed than filed	47,544	**1.71**
Fewer than a fifth observed (large fleets — inspections sample a big fleet)	894	0.97
Never seen in 24 months	358,638	**0.50**

Hence the **exposure floor** $\text{PU}_{\text{floor}} = \max(\text{filed PU},\ \text{observed power-unit VINs})$: a “one-truck” carrier seen with seven tractors is not a one-truck exposure. The never-seen population is not excluded — its dormancy is carried into the class rate through activity, filing recency and tenure (§4), and the VIN count is only ever a floor. A filing of more than 5,000 power units is above the model's fitted size range and is not graded (92% of such filers were never seen on the road); the carrier is shown with that reason. Exposure months are weighted by the same ramp as crashes, with tenure measured from the registration date, so a steady fleet contributes about 1.54 decayed power-unit-years per power unit over the 36-month window. FMCSA sometimes re-dates the registration of a carrier that has been operating for years; when a carrier has crashes or inspections on record in the two years before its registration date, tenure runs from its first crash or inspection on record instead, so the crashes it is charged with are always set against the time it was on the road. A carrier with no activity in those two years keeps its registration date.

Every carrier belongs to one of **nine fleet-size cohorts**. The cohort indexes the credibility prior (§6) and the reference rate for the driver-behaviour relativity (§5), and carries an indicator in the class rate (§4). The six cohorts of ten or more power units are the **ranked book**; the three below ten are graded but **advisory**, searchable and shown with their grade but outside the ranked book and the default charts.

Table 3.2 — The nine fleet-size cohorts. The prior a shown is the cohort's own value, used wherever the record-dependent table of §6 has no fitted value.

Cohort	Power units	Scope	Cohort prior a
`pu1`	1	Advisory; capped at Strong	0.79
`pu2_5`	2–5	Advisory	1.58
`pu6_9`	6–9	Advisory	1.88
`pu10_19`	10–19	Ranked	2.23
`pu20_49`	20–49	Ranked	1.88
`pu50_99`	50–99	Ranked	2.23
`pu100_199`	100–199	Ranked	1.88
`pu200_499`	200–499	Ranked	1.88
`pu500p`	500+	Ranked	1.88 (pooled)

In the 2026-11-01 run the ranked book held 107,352 carriers; the advisory cohorts together hold 1.72 million of the 1.82 million graded.

§4 Factor 1 — The Class Rate

The class rate is a Poisson regression with a log link and the decayed power-unit-years as offset, fitted once on the study's fit split (8.11 million carrier-origin rows, 540,090 forward crashes) and frozen. Its 60 design columns are:

Table 4.1 — Class-rate design (frozen artifact, 60 columns including the three violation terms of §5).

Block	Columns	Reference level
Size	$\log \mathrm{PU}$, $\log^2 \mathrm{PU}$, eight cohort indicators	cohort `pu10_19`
Class	for-hire flag; intrastate-only; hazmat; government; carrier class private / mixed	for-hire freight
Tenure	registered <1 y, 1–2 y, 2–5 y, 10 y+	5–10 years
MCS-150 age (as of C)	1–2 y, 2–3 y, 3–5 y, 5 y+, unknown (latest filing postdates C)	under 1 year
MCS-150 miles per power unit	$<$ 1k, 1–10k, 10–30k, 60–100k, 100–300k, $>$ 300k, unknown / implausible	30–60k
Roadside fleet evidence	share of the filed fleet seen roadside in 24 months: never, ≤ 0.1, 0.1–0.25, 0.25–0.5, >1; each band also interacted with a filing of 50+ power units and with $\log(1+\text{inspections per filed unit})$	0.5–1 of filed PU seen
Country	Canada, Mexico, other	United States
Inspection regime of the domicile state	low tertile, high tertile	middle tertile
Activity, 24 months	$\log(1+\text{traffic-enforcement inspections/PU})$, roadside, fixed-facility — each ratio capped at 4 per PU	—
Violation relativities (§5)	$\log \tilde{r}$ for driver behaviour, OOS ratio, violation ratio	—

The engine is chunked iteratively-reweighted least squares with a count-scale ridge of one expected crash (it pins rare dummies and leaves populated terms untouched) and carrier-cluster standard errors. The domicile **state** is deliberately absent: it would add about 0.75 points of deviance explained, but the study showed it is confounded with how completely each state reports crashes, so it waits for the completeness factor.

Why these inputs

The study's Poisson gradient-boosting model on the full feature library explained 31.4% of forward-crash deviance on the select split. Exposure and activity alone explained 20.4% and ranked carriers almost as well as the full model; inspections alone and violations alone each explained 14.0% — the same 14%, two measurements of one thing. Crash history alone explained 5.3% but almost all of it was unique: it is the one family that is not a proxy for activity, and it is the experience term of §6.

The inspection-selection analysis then settled *how* activity enters. Between inspection regimes the count means almost nothing (tertile effects of -5%), within regimes a great deal; type matters more than count — a traffic-enforcement stop is an observation of driving (relative risk about 2.4 per log unit of stops per truck), a weigh-station pull-in an observation of being on the road (about 1.15), and carriers inspected mostly at scales realized 0.69 of the activity model; among 308,800 rows with inspections and *no violation at all*, the slope of inspections on crashes is zero and realized $\div$ expected falls with the count. The count's predictive power is activity plus what the inspections find. Above about four inspections per power unit the realized curve flattens, so the activity terms are capped there — an audit of the Excellent list found a 61-truck carrier whose 1,219 inspections had extrapolated its class rate to 15× and hidden a crash rate above the book average.

Reading the roadside record

The roadside record is the only independent check on a filed fleet, and the class rate reads it in two dimensions: how much of the filed fleet the road actually sees, and how actively the carrier is inspected. A single “observed fleet” band was tested and rejected: it puts a 200-unit filing seen with 18 units and a 12-unit fleet seen with 5 in the same cell, and under it the clean, thinly seen large filings realized 1.5–2.4× their forecast and graded in

inverted order. With six seen-share bands, each interacted with a 50-plus filing and with inspections per filed unit, that slice realizes 0.99 of its forecast and 0.94 / 1.09 / 1.02 by tercile of $\bar{R}$. Replacing the filed fleet by a roadside capture–recapture estimate of the true fleet as the exposure was also built and rejected: it lost deviance in a model-free test and, refitted, changed no cohort's Gini.

Table 4.3 — Class-rate factor for the share of the filed fleet seen roadside in 24 months (against 0.5–1 seen, other inputs fixed).

Filing, activity	never seen	≤ 0.1	0.1–0.25	0.25–0.5	> 1
Under 50 PU, 0.25 inspections per unit / 24 m	0.34	0.43	0.61	0.82	1.64
Under 50 PU, 2 inspections per unit	0.91	1.56	0.90	0.89	1.16
50+ PU, 0.25 inspections per unit	0.06	0.36	0.62	0.81	2.21
50+ PU, 2 inspections per unit	0.16	1.31	0.92	0.88	1.57

Table 4.2 — Selected class-rate effects as rate ratios, e^{β}, from the frozen artifact.

Input	Rate ratio
Last MCS-150 filing 1–2 y / 2–3 y / 3–5 y / 5 y+ old (vs. under 1 y)	0.83 / 0.72 / 0.64 / 0.40
Latest filing postdates the cutoff (an active filer)	1.76
Registered under 1 y / 1–2 y / 2–5 y / 10 y+ (vs. 5–10 y)	1.58 / 1.17 / 1.10 / 0.93
Share of the filed fleet seen roadside	Table 4.3
Traffic-enforcement / roadside / fixed-facility inspections per PU (per log unit, before the seen-share interaction)	1.61 / 1.40 / 0.86
MCS-150 mileage unknown or implausible (vs. 30–60k miles per truck)	0.60
Intrastate-only operation	0.77
Domiciled in Canada / Mexico / elsewhere (vs. United States)	0.66 / 0.12 / 0.29
Low / high inspection-regime tertile (vs. middle)	0.99 / 0.97

Each ratio holds the other inputs fixed; the for-hire flag and the carrier-class indicators interact and are not read singly, and the cohort indicators are small (0.93–1.21) once the size terms are present. The model-free counterpart of the filing-recency effect in the study was starker: carriers whose last MCS-150 was more than five years old realized 0.18× a cohort-rate prediction — most are dormant registrations that still carry power units.

§5 Factor 2 — Violation Relativities

Three violation signals over the trailing 24 months enter the class-rate GLM as $\log \tilde{r}_s$, each a credibility-shrunk ratio of observed to expected. The shrinkage is Bühlmann's linear form, and the credibility constant K_s and the exponents γ_s were chosen on the select split by out-of-sample forward-crash deviance:

$$\tilde{r}_s \;=\; Z_s \cdot \frac{\text{observed}_s}{\text{expected}_s} \;+\; (1 - Z_s), \qquad V \;=\; \prod_s \tilde{r}_s^{\,\gamma_s}.$$

Table 5.1 — The three violation signals (frozen artifact).

Signal	Observed	Expected	Credibility Z_s	γ_s
Driver behaviour	Unsafe Driving + Hours-of-Service violations	driver-level inspections × the cohort's rate per inspection	$n/(n+10)$, n = driver-level inspections	0.112
Out-of-service ratio	inspections with an OOS result	Σ over the carrier's inspections of the OOS rate for that inspection's state × level cell	$e/(e+1)$, e = expected OOS inspections	0.083
Violation ratio	all violations	Σ over inspections of the violation rate for that state × level cell	$e/(e+3)$, e = expected violations	0.077

The two ratios are **indirectly standardised** for where and how the carrier was inspected: the expected count is built from the state × inspection-level mix of the carrier's own inspections (an inspection by a federal inspector is filed under a federal code rather than a state, so it is compared with federal inspections of the same level wherever they took place). In the study the standardised ratio beat the raw per-inspection rate every time, decisively for out-of-service results — once the standardised OOS ratio is in the model the raw OOS rate takes a negative sign. Model-free, a standardised violation ratio below 0.5 of expected realized 0.87–0.95 of the activity model and above 2× realized 1.10–1.18, monotone in every regime tertile; real, consistent, and modest beside the exposure and activity effects, which is why the γ are small.

Table 5.2 — Cohort reference rates for the driver-behaviour signal (Unsafe Driving + HOS violations per driver-level inspection, frozen artifact).

Cohort	pu1	pu2_5	pu6_9	pu10_19	pu20_49	pu50_99	pu100_199	pu200_499	pu500p
Rate	0.322	0.318	0.286	0.266	0.245	0.219	0.195	0.189	0.148

> **Which violations earn a place.** With activity and regime controlled, the forward-predictive BASICs are Unsafe Driving (relative risk 1.87 per log unit of violations per inspection), Hours-of-Service (1.21) and Vehicle Maintenance (1.16) — the same three UMTRI found in 2011 — and the maintenance signal is carried by the OOS ratio. Driver Fitness (0.66) and Hazardous Materials (0.64) are negative and get no term of their own — their violations count only inside the all-violations ratio: licence and medical-card violations mark administrative carriers, not dangerous ones. At the violation-group level the top of the table is driver behaviour, in order — speeding, seat belt, phone, texting, reckless driving, relative risks 1.15–1.24 — which is the content of the two BASICs in the driver-behaviour signal. Violation detail adds little to frequency once inspection and OOS counts are present; its role in the rating is face validity and the Excellent screen (§8).

§6 Factor 3 — Crash Experience

A carrier's own crash history is the one family of information that is not a proxy for activity, and it enters as classical Gamma–Poisson credibility. Let the carrier's true relativity to its class-and-violation expectation be $\theta \sim \text{Gamma}(a, a)$ — mean 1, variance $1/a$ — and let its decayed crash count be Poisson given θ and the decayed expectation Λ. Conjugacy gives the posterior and its mean in closed form:

$$\theta \mid \text{data} \;\sim\; \text{Gamma}(a+N,\; a+\Lambda), \qquad \boxed{\bar{\theta} \;=\; \frac{a+N}{a+\Lambda}}, \qquad Z \;=\; \frac{\Lambda}{a+\Lambda},$$

so $\bar{\theta} = Z \cdot (N/\Lambda) + (1-Z)\cdot 1$: the carrier's own ratio weighted by its credibility Z, the class expectation by the rest. Here

$$N \;=\; \sum_{\text{recordable crashes}} w(\text{age}), \qquad \Lambda \;=\; \mu_{\text{class}} \cdot V \cdot \sum_{\text{months}} \text{PU}_{\text{floor}} \cdot \tfrac{\text{days}}{365} \cdot w(\text{age}), \qquad w(\text{age}) \;=\; 1 - \frac{\text{age in months}}{36}.$$

Because Λ includes V, $\bar{\theta}$ is the residual after class *and* violations have spoken.

The decay

Events and exposure share the weight w: a crash loses one thirty-sixth of its weight each month and is gone at 36 months, with no step anywhere. The shape was chosen by the out-of-time negative-binomial predictive likelihood of the next twelve months' crashes over 13.3 million carrier-origin rows, jointly with the credibility prior. On a like-for-like comparison every 36-month shape beat every 24-month shape; among 36-month shapes the step-free linear ramp sat 1,257 log-likelihood units (0.19%) behind an unconstrained exponential that keeps a quarter of an event's weight at month 36 and would therefore step when it leaves the window, and 3,597 units ahead of equal weights. Fred's decision (2026-10-22): a crash signal on 36 months with a linear ramp to zero; the inspection and violation signals stay on 24 months, since the inspection file carries only 36 months of history.

The credibility prior

The prior strength a is fitted after the class model by the same forward negative-binomial likelihood on the fit split, per cohort **and** per band of the record's expected crashes Λ (< 0.1, 0.1–0.3, 0.3–1, 1–3, 3–10, 10+), and interpolated linearly in $\log a$ over $\log \Lambda$ between band midpoints so that a carrier crossing a band edge does not jump:

$$\bar{\theta} = \frac{a(c,\Lambda) + N}{a(c,\Lambda) + \Lambda}, \qquad Z = \frac{\Lambda}{a(c,\Lambda) + \Lambda}.$$

A single constant per cohort, the textbook choice, was tested and rejected: right on average, it was wrong in both directions by record. A crash on a ranked record expecting under 0.1 crashes was followed by 3.6× the class rate against 1.26× credited, and clean records expecting one to three crashes had 21% more crashes than credited (0.60 of class against 0.49). The fitted table gives a thin record a diffuse prior — one crash moves it a long way — and a record with a few expected crashes a narrow one.

Table 6.1 — Credibility prior a by cohort and expected crashes Λ (blank: too few rows to fit; the cohort's own a of Table 3.2 is used).

Cohort	Λ < 0.1	0.1–0.3	0.3–1	1–3	3–10
pu1	0.79	0.94			
pu2_5	0.67	1.88	2.65		
pu6_9	0.33	1.12	2.65		
pu10_19	0.24	0.94	2.23	3.75	
pu20_49	0.10	0.40	1.12	3.75	
pu50_99	0.14	0.40	0.94	1.58	7.48
pu100_199				1.12	5.30
pu200_499, pu500p					

$Z = \Lambda/(a + \Lambda)$ is shown on the carrier page as a **credibility tier** — High (≥ 0.60), Moderate (≥ 0.30), Low (≥ 0.10), Prior-only (below 0.10, marked provisional). The tier is descriptive; the grade uses the full posterior. The displayed Z is scaled by an **exposure corroboration** factor $c = \min\{1, \max(s/0.5,\ i/0.5)\}$, where s is the share of filed power units seen roadside in 24 months and i the inspections per filed unit-year: a filed fleet the roadside record never sees cannot carry High confidence, whatever Λ says. From 10 filed units, $c < 0.5$ raises the `EXPOSURE_UNVERIFIED` flag and, as an interim rule, holds an Excellent or Strong grade one band lower (Table 8.2). The relativity and forecast are not scaled by c; exposure trust enters them through the roadside-fleet terms of the class rate (§4) and the record-dependent prior above. The cap exists because those terms were fitted when the roadside VIN record covered only part of each 24-month window (it begins in late 2024), so for a filing the road rarely sees the class rate's discount is not yet tested on a full window; it is lifted once an origin with full roadside coverage has a mature outcome year that shows the favourable grades hold.

EXAMPLE 6.1 — WHAT A THIN CLEAN RECORD EARNS

A 10–19 power-unit fleet whose class rate and exposure give $\Lambda = 0.3$ expected crashes over the decayed 36-month window, with none observed and a clean violation record ($V = 1$): the prior there is $a = 1.42$, so $\bar{\theta} = 1.42/1.72 = 0.83$. A modest discount — Satisfactory, not a clean sheet (the Strong cut is 0.756). A single truck with one recent crash ($N \approx 0.9$ after decay) and $\Lambda = 0.04$ has $a = 0.79$ and $\bar{R} \approx 2.0$, in the Poor band; but only about half of its wide posterior lies above the Poor cut, so the downside screen grades it Marginal. A fleet with tens of expected crashes and none observed drives $\bar{\theta}$ toward zero and passes the Excellent screen on its own evidence.

§7 Relativity to Class and Forward Expectation

The graded statistic is the posterior **mean** relativity to class — never the median, which on thin Gamma posteriors sits far below the mean:

$$\boxed{\bar{R} \;=\; V \cdot \bar{\theta}}, \qquad \mathbb{E}[\text{crashes}_{12\,\text{mo}}] \;=\; \mu_{\text{class}} \cdot \bar{R} \cdot \text{PU}_{\text{floor}}, \qquad \mathbb{E}[\text{crashes}_{12\,\text{mo}} \mid \text{class only}] \;=\; \mu_{\text{class}} \cdot \text{PU}_{\text{floor}}.$$

The posterior for the carrier's relativity is the Gamma of §6 scaled by V; its 10th and 90th percentiles q_{10}, q_{90} are reported with $\bar{R}$, and its tail probabilities are the screens of §8:

$$P(R < c) \;=\; F_{\text{Gamma}(a+N,\, a+\Lambda)}\left(\frac{c}{V}\right).$$

On the study's select split the posterior was neither too narrow nor too wide: among rows with at least 0.5 expected crashes, 86% realized at or below the posterior 80th percentile (0.80–0.90 is the calibrated range once integer counts are allowed for). The 0–100 **FRED Score** interpolates $\bar{R}$ linearly within its band between the anchors 100 / 90 / 75 / 50 / 30 / 10 / 0 at the band edges, reaching 0 at twice the Critical cut, so score and grade never disagree.

§8 From Relativity to Grade

Cuts

Bands come from cuts on $\bar{R}$. The four lower cuts were **fitted** on the fit split: the isotonic (pool-adjacent-violators) fit of realized ÷ class-expected crashes on $\bar{R}$, weighted by class-expected counts rescaled so that **every cohort carries equal total weight** (an exposure-weighted fit would let the largest fleets set the small fleets' cuts), is crossed at the nominal boundaries 0.80 / 1.10 / 1.40 / 1.90. The fitted values are 0.756 / 1.057 / 1.570 / 2.109: the two lower ones within 6% of nominal, which is the check that $\bar{R}$ is calibrated through the middle; the two upper ones above nominal, because carriers with high posterior relativities realize somewhat less than $\bar{R}$ says and the fitted cuts absorb it. The Excellent cut is set at **0.60 by decision** — half the crashes of a comparable fleet — with the rate-based proof screens below carrying the burden of selectivity; count-based screens punish size.

Table 8.1 — Grade bands, score range and realized ÷ class-expected crashes in the out-of-time gate year of the latest run (2026-11-01: 1,666,169 carriers graded as they stood at the latest origin whose outcome year is at least 120 days mature, on the crashes realized in the following twelve months).

Grade	Relativity to class $\bar{R}$	Score	Realized ÷ class-expected
Excellent	$\bar{R} < 0.60$	90–100	0.48
Strong	$0.60 \le \bar{R} < 0.756$	75–90	0.63
Satisfactory	$0.756 \le \bar{R} < 1.057$	50–75	0.86
Marginal	$1.057 \le \bar{R} < 1.570$	30–50	1.17
Poor	$1.570 \le \bar{R} < 2.109$	10–30	1.66
Critical	$\bar{R} \ge 2.109$	0–10	3.18

Screens

The band says where the posterior mean falls; the screens ask whether the record supports it. They are applied in order — the downside posterior test first, then the upside tests, then the event requirements — and each is one-directional.

Table 8.2 — Screens, in the order applied (frozen artifact: $p_{\text{top}} = p_{\text{down}} = 0.6$, $p_{\text{bottom}} = 0.8$).

Rule	Condition	Otherwise
Downside posterior (Critical, Poor, Marginal in turn)	$P(R >$ band's lower cut$) \geq 0.6$	one band toward Satisfactory, then tested as that band — a wide posterior can fall more than one band
Excellent	band Excellent and $P(R < 0.60) \geq 0.6$ and at least 0.3 expected crashes over the window ($\Lambda \geq 0.3$) and driver-behaviour relativity $\tilde{r} \leq 1.0$ and no fatal crash in the window and ≥ 3 inspections and ≥ 0.1 inspections per *filed* power unit (24 months) and more than one power unit	Strong, then tested as Strong
Strong	band Strong or demoted from Excellent, and $P(R < 0.756) \geq 0.6$ and ≥ 0.1 inspections per filed power unit	Satisfactory
Critical	band Critical and (a fatal crash, or ≥ 2 recordable crashes in the window, or $P(p_{\text{oos}} > 2\mu_{\text{oos}}) \geq 0.8$)	Poor, then tested as Poor
Poor	band Poor or demoted from Critical, and (≥ 1 recordable crash in the window, or $P(p_{\text{oos}} > 1.5\mu_{\text{oos}}) \geq 0.8$)	Marginal
Exposure cap (interim)	from 10 filed power units, exposure corroboration $c < 0.5$ (§6): under a quarter of the filed units seen roadside in 24 months and under 0.25 inspections per filed unit-year	Excellent → Strong, Strong → Satisfactory
Adverse rating	a *current* FMCSA Unsatisfactory or Conditional safety rating	capped at Marginal, referral flag
Evidence for Satisfactory (display)	grade Satisfactory and not ($q_{10} < 0.60$ and $q_{90} \geq 1.57$)	shown as Not enough record: no grade, no score; forecasts unchanged

Not enough record. Every other grade passes a 60% posterior test against its cut; Satisfactory is what remains when none passes, which mixes a record that supports the class rate with a record that supports nothing. When the 80% range spans both the Excellent cut and the Poor cut, the record cannot separate a best-in-class fleet from a poor one, and the carrier is displayed as *Not enough record* and sorted after graded carriers; the engine's grade, gates, history and forecasts are untouched. On the select split (ten monthly origins, 2024-11 to 2025-08, realized 12-month crashes), ranked-book carriers under this rule realized 1.00 of their class expectation and 1.01 of their forecast, the same as Satisfactory carriers whose range supports the grade; the difference is the dispersion of the true rate around the forecast, estimated from $E[(y - \hat{\mu})^2 - y] = \phi\hat{\mu}^2$: $\phi = 0.46$ (±68%) against 0.07 (±26%) for a range inside 0.60–1.57, and the posterior's share at Poor-level rates (17%) matched the realized spread (18%).

The out-of-service posterior is a Beta: with k OOS inspections of n, μ_{oos} the carrier's own regime-expected OOS rate (its expected OOS count over its inspections; the cohort mean when it has none) and prior strength $K = 1/\mu_{\text{oos}}$ inspections (one expected OOS event),

$$p_{\text{oos}} \sim \text{Beta}\big(k + K\mu_{\text{oos}},\ n - k + K(1 - \mu_{\text{oos}})\big), \qquad \text{test } P(p_{\text{oos}} > m\,\mu_{\text{oos}}) \geq 0.8,\ m \in \{2,\ 1.5\}.$$

Why 0.6. The bands are already cut on the posterior mean, so the quantile screen is a safeguard, not the main event: in the study's sweep 0.5 and 0.6 were nearly identical, while 0.8 left 0.2% of the book in Strong and 299 rows in Excellent. The evidence floor $\Lambda \geq 0.3$ exists because a diffuse prior on a thin record (Table 6.1) can pass the posterior test on almost no evidence. The binding upside screens in that sweep were exposure plausibility (13,629 band-Excellent/Strong rows with fewer than 0.1 inspections per filed power unit — one filer of 4,150 trucks had three inspections), the driver-behaviour screen (7,979 rows) and the crash-event screen (1,759); the posterior quantile blocked 292. The sweep ran the driver-behaviour and crash screens in count-based forms (any

Unsafe Driving or Hours-of-Service violation; any injury or fatal crash), which were rejected because they punished size — a large fleet always has some violation and some injury crash — in favour of the rate-based relativity and fatal-only rules of Table 8.2.

> **The size skew — "bigger the better, all else equal."** Between two carriers at the same point estimate, the larger fleet's relativity is known and the smaller fleet's is a wide guess. The posterior encodes exactly that, so the principled implementation is to leave Satisfactory in *either* direction only when 60% of the posterior lies beyond the cut. On the select split the downside screen took the 10+ book from 21.2% Marginal and 4.4% Poor to 9.2% and 1.9%, concentrated in fleets with evidence (50+ PU kept 15% Marginal and 5% Poor; 10–19 PU fell to 6.7% and 0.7%); a single truck is Poor only 0.2% of the time — one crash on one truck is not enough on its own — and the tiers that remain are purer (Marginal realizes 1.39 of class instead of 1.26, Poor 1.82 instead of 1.70). Going to 0.7 on both sides would have cut Strong from 15% to 6% of the book for little added purity.

Reachability of Excellent

The charter requires a published table of the minimum clean exposure at which Excellent is reachable. For each cohort, $\Lambda_{\min}$ is the decayed expected count at which a clean record with $V = 1$ first satisfies both the band ($\bar{\theta} < 0.60$) and the posterior screen $P(\theta < 0.60 \mid N = 0) \geq 0.6$, from the evidence floor upward, and the steady fleet size that produces it (about 1.54 decayed power-unit-years per power unit) at two class rates, each measured on its own population: the cohort's median over every graded carrier-origin of the study's select split (§11), dormant and never-seen registrations included, and its median over the fleets the road sees in a live run of the engine (graded carriers with exposure corroboration $c \geq 0.5$, §6), which is higher by a factor of 2.2 to 2.8 in the cohorts from 6 to 499 power units, 3.4 at 500 or more and 5.2 at 2–5.

Table 8.3 — Reachability of Excellent on a clean record with $V = 1$, at the cut of 0.60: the class rate as the cohort median over every graded carrier-origin of the study's select split, and over fleets seen roadside (exposure corroboration $c \geq 0.5$) in a live run of the engine.

Cohort	$\Lambda_{\min}$ (expected crashes)	a there	Median class rate / PU-yr, all filings	Steady fleet for $\Lambda_{\min}$	Median class rate / PU-yr, fleets seen roadside	Steady fleet, seen roadside
pu2_5	1.24	1.84	0.0059	136	0.0304	26
pu6_9	1.39	2.01	0.0117	77	0.0328	27
pu10_19	2.38	3.25	0.0138	112	0.0346	45
pu20_49	2.31	3.16	0.0148	101	0.0348	43
pu50_99	0.71	1.06	0.0178	26	0.0413	11
pu100_199	0.97	1.45	0.0185	34	0.0414	15
pu200_499	1.27	1.88	0.0171	48	0.0398	21
pu500p	1.27	1.88	0.0105	78	0.0357	23

At the class rate of a fleet the road sees, a clean record alone earns Excellent from the mid-40s of power units upward (every 50+ cohort reaches $\Lambda_{\min}$ at any size in its range); at the median over every filing it takes about 50. Below that it takes a clean record *and* a violation record better than class ($V < 1$), which is how the Excellent carriers in the 10–49 cohorts get there; the exposure cap of Table 8.2 then holds the grade at Strong while the filed fleet is uncorroborated. A single-power-unit carrier is capped at Strong by rule: nothing is proven on one unit, and 682 heavily inspected single trucks reached Excellent through the rate-based screens in a study dry run without it. Excellent is deliberately rare even among large fleets: 2.5% of the ranked book in the 2026-11-01 run.

---

§9 Change Attribution and the Governor

Against the previous run, the change in a carrier's log relativity decomposes exactly. Writing primes for the previous run and $\Lambda_{\text{class}} = \Lambda / V$ for the expectation at the class rate alone,

$$\Delta \log \bar{R} \;=\; \underbrace{\big[\log(a+N)-\log(a'+N')\big]}_{\text{new event, decay roll-off or data correction}} \;+\; \underbrace{\Delta \log V + \big[\Delta(-\log(a+\Lambda))-\Delta(-\log(a+\Lambda_{\text{class}}))\big]}_{\text{violations}} \;+\; \underbrace{\Delta\big(-\log(a+\Lambda_{\text{class}})\big)}_{\text{exposure \& class}}$$

The numerator term is labelled **new event** when the decayed or raw crash count rose, **decay roll-off** when it fell by no more than one month of the ramp can explain ($N'/36$), and **data correction** when the raw window count fell and the decayed count fell by more than that — a crash removed from the FMCSA file. The **recalibration** term is non-zero only in a methodology run — one whose artifact checksum, cutoff rule, tenure rule or exposure-cap rule differs from the previous run's: there the decay, violation and exposure components are folded into it, so everything but a new event or a data correction is credited to the scoring update. A grade moved by a screen alone, with no component of $\bar{R}$ moving, is credited to exposure & class (the screens read fleet size and activity), or to the recalibration in a methodology run. The dominant component with the sign of the move names the cause on the carrier page; a carrier whose relativity did not move is *unchanged*, one absent from the previous run is *new*.

> **The upside governor.** An improvement whose dominant cause is not decay roll-off, a data correction or a recalibration is ramped **one band per run**: displayed = max(model grade, previous displayed − 1), and the model grade is kept as the **pending grade** until the display reaches it. The downside is never governed — a worse model grade shows at once. In the engine's monthly simulation on the selection split the largest upward moves were all exposure moves (a single-truck carrier observed with a second power unit, a burst of inspections raising its activity term, halving $\bar{R}$ in a month without a new crash); they are real information about exposure, and the governor is right to let them arrive gradually with the attribution saying "exposure & class." With every parameter frozen, 0.66–0.80% of carriers moved at least one band per 30 days, 0.02–0.03% two bands, 0.38–0.49% upward and 0.27–0.31% downward; the governor touched 0.15–0.24% per month and left the realized gradient of the displayed grades within a hundredth of the model grades'. The live run's churn is reported in §11. No carrier ever improved because of a new event — that is also a hard gate (§10).

---

§10 Acceptance Gates

Every run grades, on the same artifact, the population as it stood at the latest month-start origin whose twelve-month outcome year ended at least 120 days before the data date, and judges it on the crashes realized in that year — the wait is so that the year is nearly fully reported (the add-date completeness curve reaches 92.5% at 90 days, 94.7% at 120 and 98.5% at a year). Hard gates abort the write; soft gates are reported in the weekly run report, which lists every gate with its value and bound, the grade mix for the universe and the ranked book, churn with its causes, and the ten largest moves each way.

Table 10.1 — The gate suite, with the values of the latest run (2026-11-01).

Gate	Bound	Kind	Latest run (2026-11-01)
Out-of-time Gini, ranked cohorts with enough outcomes to judge (10–19 / 20–49 / 50–99 PU)	≥ 0.55 each	hard	0.596 / 0.649 / 0.724
Grade monotonicity (realized ÷ class-expected by grade)	No inversion > 2 standard errors	hard	Monotone
Aggregate calibration (realized ÷ predicted)	0.90 – 1.10	hard	0.945
Calibration by evidence bucket (expected crashes), worst bucket	Within ±10% soft, ±20% hard	soft / hard	1.03 / 0.90 / 0.90 / 0.90 / 0.97 / 0.96 / 0.96 → 0.102 — outside the soft bound
Posterior coverage (share realized at or below the 80th percentile)	0.78 – 0.92	soft	0.873
Moves of two or more bands	≤ 0.5% of carriers	hard	0.001%
Moves of one or more bands; governed share	≤ 3%; ≤ 0.5%	soft	0.04%; 0.01%
Upward moves caused by a new event	0	hard	0
Rows graded; ranked rows	≥ 900,000; ≥ 80,000	hard	1,824,374; 107,352
Single-power-unit Excellent	0	hard	0
Scores in range; all six grades present	0 – 100; yes	hard	0 – 97.5; yes
Pooled Gini; dead-signal share	Reported	info	0.642; 0.517

Verdict of the 2026-11-01 run: all nineteen checks ran; every hard gate passed, and one soft check, calibration by evidence bucket, fell outside its bound (the worst bucket, 0.90 of forecast at 0.2–0.5 expected crashes, sits 10.2% from calibration against a 10% soft bound) and is reported in the weekly run report. The evidence buckets realize 1.03 / 0.90 / 0.90 / 0.90 / 0.97 / 0.96 / 0.96 of forecast from under 0.05 to 5+ expected crashes; a shortfall in the buckets between 0.05 and one expected crash is the class rate's known soft spot, where the log-linear class rate slightly over-predicts quiet, lightly inspected carriers (§11).

§11 Validation: the Study and the Live Run

Design

Ten monthly origins, 2024-11 through 2025-08, each with features as of the origin (by add date) and a mature 12-month forward outcome; all property carriers with filed power units ≤ 5,000, not under an exit at the origin, forward exposure > 0 (VIN-floored power units × the fraction of the year before any exit). Carriers were split by DOT number — odd: **fit**; divisible by four: **select** (model choice, K, a); remainder two: **eval**, locked and read once after the candidate was frozen. Rows of one carrier stay in one split, so every out-of-sample figure is also out-of-carrier. The ten origins overlap eleven-twelfths in outcomes and are one independent outcome year; confidence intervals come from the carrier-cluster bootstrap. Metrics: Poisson deviance explained against a single-rate null, exposure-weighted normalized Gini (pooled and per cohort), Σ realized ÷ Σ predicted overall, by cohort and by decile, top-5% exposure lift, and the posterior 80th-percentile coverage.

The bake-off, select split

Table 11.1 — Out-of-sample on the select split (4.11M carrier-origin rows, 273,755 forward crashes).

Model	Deviance explained	Normalized Gini
PU-cohort rate (the size-only baseline)	1.1%	0.13
The FRED engine (class rate × violation relativities × crash experience)	**28.2%**	**0.653**
Gradient-boosting ceiling (same rows, same inputs, no leak)	30.5%	0.674

The engine sits 0.021 Gini and 2.3 points of deviance behind the gradient-boosted ceiling. That gap is the price of linearity among busy carriers: the tree ranks better pooled and, in the study's bake-off, inside every cohort — its advantage is nonlinear structure among active, inspected carriers — and it cannot decompose or be governed. Per cohort the engine's select-split Gini is 0.605 / 0.668 / 0.735 / 0.792 for pu10_19 / pu20_49 / pu50_99 / pu100_199. In the study the experience term was the single largest step (+3.4 points of deviance and +0.035 Gini on top of the class rate) and kept aggregate calibration at 1.00; neither an isotonic recalibration of the class rate nor activity band dummies paid (each gained 0.3–0.4 points of deviance and moved the Gini by at most 0.004). Normalized Gini throughout is the exposure-weighted Lorenz integral.

Calibration holds where it matters for underwriting. By decile of predicted rate, realized ÷ predicted runs from 0.90 at the lowest decile to 1.19 in the busy upper middle and 0.95 at the highest — a mild U-shape: the log-linear class rate slightly over-predicts quiet, rarely inspected carriers and under-predicts the busy upper middle. Thin ranked records ($\Lambda < 0.3$) realize 0.89 of their forecast when clean and 1.11 after a crash, and the clean, thinly seen large filings of §4 realize 0.99.

The locked evaluation split, read once

The locked split (4.03M rows, DOT mod 4 = 2, 268,454 forward crashes) was read once, at the bake-off, and has not been reopened. On it the three-factor structure explained 27.4% of deviance with Gini 0.647 and calibration 1.003, against the ceiling's 30.6%, 0.674 and 1.014 on the same rows — the same picture as select. Per cohort the Gini ran 0.59–0.83 with realized ÷ predicted 0.98–1.08 in every cohort. The engine's own figures (Table 11.1) are select-split figures.

Grading on history

Every run grades the population as it stood at a mature origin before its cutoff and checks the grades against the crashes that followed (§10). In the 2026-11-01 run's gate year the grades realize 0.48 / 0.63 / 0.86 / 1.17 / 1.66 / 3.18 of the class expectation from Excellent to Critical (Table 8.1), monotone by the 2-SE rule. The charter's tier criteria are met: Excellent realizes below 0.6, Critical above 1.9, no upside without the screens and no downside without events. In the engine's monthly simulation on the selection split, with every parameter frozen, 0.66–0.80% of carriers moved at least one band per 30 days, 0.02–0.03% two bands, and the governor touched 0.15–0.24%. In the live engine, between the 2026-10-27 and 2026-11-01 runs, 0.04% of the 1,823,517 carriers graded in both moved at least one band (0.23% per 30 days), 0.001% two bands (0.006% per 30 days), and the governor held 0.010%; 268 carriers moved up and 442 down, none up because of a new event.

The live run

On November 1, 2026 (inspections, violations and census facts through September 30, 2026, crashes through August 31, 2026, data as of October 29, 2026) the engine graded 1,824,374 property carriers, 107,352 of them in the ranked book. Every hard gate passed and the soft check on calibration by evidence bucket fell outside its bound (§10): per-cohort Gini 0.596 / 0.649 / 0.724, aggregate calibration 0.945, posterior coverage 0.873, 0 single-power-unit Excellent. The ranked-book grade mix in the live table (the governed grade of §9 after the adverse-rating cap of Table 8.2, before the Not-enough-record display rule of §8) was Excellent 2,731 (2.5%) · Strong 13,318 (12.4%) · Satisfactory 83,247 (77.5%) · Marginal 6,570 (6.1%) · Poor 1,081 (1.0%) · Critical 405 (0.4%) of 107,352; across the whole universe, in the live table, 2,767 · 21,659 · 1,766,522 · 29,818 · 2,778 · 830. The exposure cap of Table 8.2 held 5,412 Excellent or Strong model grades one band lower, and of the Satisfactory carriers 1,734,192 (95.1% of all graded; 61,922, 57.7%, in the ranked book) are shown as Not enough record (§8), leaving 32,330 (21,325 ranked) whose record supports the grade.

§12 Severity Display and What Is Not in the Grade

The graded quantity is recordable-crash frequency. For display and the later cargo and auto loss-cost layers, the carrier page also shows an expected severity mix — property-damage, injury and fatal crashes, and the probability of at least one fatal crash — computed as the frequency expectation times the **cohort's observed severity mix**. A separate severity model (a multinomial on class attributes and crash configuration) is a planned follow-up; severity never enters the frequency model.

Table 12.1 — Excluded from the grade, and why.

Not an input	Reason
Self-reported MCS-150 mileage	Blank, stale or implausible too often to be exposure; enters only as a coarse miles-per-truck band in the class rate
FMCSA BASIC percentiles, ISS	Outcome-dependent (would leak the target); the public SMS carries no percentiles for property carriers
Machine-learned models	The gradient-boosted model is a ceiling in the bake-off only; the engine must decompose and must not spike
Domicile state	Confounded with crash-reporting completeness (25% to over 90% known within 60 days by state); revisited with the completeness factor
Driver Fitness and Hazardous Materials violations as signals of their own; raw OOS and violation rates; driver-OOS rate	Negative or null once activity and the standardised ratios are present (the violations still count in the all-violations ratio)
A pending out-of-service order or revocation dated inside the outcome window	Forward-looking (a leak found in the study); exits are used only for censoring and the at-cutoff exclusion
Crash severity in the frequency model	Noise for frequency; modelled separately
Anything driver-level (CDLIS/PSP, Clearinghouse, ELD), IRP/IFTA, DataQs and enforcement-case detail	Not public

§13 Carriers, End to End

Three illustrative carriers, computed with the production constants.

EXAMPLE 13.1 — A MID-SIZE FLEET THAT EARNS EXCELLENT

Inputs. 88 power units (cohort `pu50_99`), 120 inspections in 24 months, two recordable crashes in 36 months (1.32 after decay).

Factors. The class rate and violation relativity together expect $\Lambda = 2.3$ decayed crashes for this fleet's exposure; its violation relativity is $V = 0.66$ — fewer driver-behaviour violations and out-of-service results than its inspection regime expects. At $\Lambda = 2.3$ the prior is $a = 2.32$, so $\bar{\theta} = (2.32 + 1.32)/(2.32 + 2.3) = 0.79$.

Grade. $\bar{R} = 0.66 \times 0.79 = 0.52$: band Excellent. The screens: 67% of the posterior lies below 0.60, Λ is above the 0.3 floor, 120 inspections on 88 filed units is about 1.4 per unit, driver behaviour is better than class and there is no fatal crash — **Excellent**.

EXAMPLE 13.2 — A SMALL CLEAN RECORD THAT HAS PROVEN NOTHING YET

Inputs. 5 power units (`pu2_5`), two inspections, no crash; $\Lambda = 0.04$ expected crashes over the window and $V = 0.90$.

Grade. The prior at $\Lambda = 0.04$ is $a = 0.67$, so $\bar{\theta} = 0.67/0.71 = 0.94$ and $\bar{R} = 0.85$: the record is too thin to move the carrier from its class rate — **Satisfactory**, advisory (under 10 power units). A clean sheet on two inspections is the absence of evidence, and the posterior says so; with Λ below 0.3 it could not be Excellent in any case.

EXAMPLE 13.3 — A BAD RECORD THAT STOPS SHORT OF CRITICAL

Inputs. 40 power units (`pu20_49`), 70 inspections, twelve recordable crashes in 36 months, most of them recent (7.5 after decay), against $\Lambda = 3.1$ expected; $V = 1.10$.

Grade. The prior is $a = 2.64$, so $\bar{\theta} = (2.64 + 7.5)/(2.64 + 3.1) = 1.77$ and $\bar{R} = 1.94$: band Poor, below the Critical cut of 2.109. The posterior is narrow enough that 71% of it sits above the Poor cut (the downside screen holds), and twelve window crashes satisfy the event requirement many times over — **Poor**, score about 16. The engine does not round a bad record up to Critical: that tier begins at 2.109, the fitted point where realized crashes cross $1.9\times$ the class expectation.

§A Symbols and Constants

Table A.1 — Production constants (frozen model artifact).

Constant	Value	Section
Cutoffs C, C_{crash}	first of the month after the last full month on file; crashes one month earlier	§2, §B
Crash window and decay	36 months, $w = 1 - \text{age}/36$	§6
Inspection, violation and VIN window	24 months	§2, §5
Exposure	$\text{PU}_{\text{floor}} = \max(\text{filed}, \text{observed power-unit VINs})$, decayed PU-years	§3
Fitted size range (not graded above it)	≤ 5,000 filed power units	§2, §3
Fleet-size cohorts	9; ranked book = 10+ PU (6 cohorts)	§3
Class-rate design	60 columns; six seen-share bands × (≥ 50 filed PU) × log(1 + inspections per PU); ridge 1.0 expected crash; activity cap 4 inspections per PU	§4
Driver-behaviour credibility K	10 driver-level inspections	§5
Standardised-ratio credibility K	1 expected OOS event; 3 expected violations	§5
Violation exponents γ (driver / OOS / violations)	0.112 / 0.083 / 0.077	§5
Credibility prior a	per cohort × Λ band, interpolated in $\log\Lambda$ (Table 6.1); cohort values 0.79 / 1.58 / 1.88 / 2.23 / 1.88 / 2.23 / 1.88 / 1.88 / pooled 1.88	§6
Credibility tiers on $Z = \Lambda/(a+\Lambda)$	High ≥ 0.60; Moderate ≥ 0.30; Low ≥ 0.10; Prior-only (provisional) below	§6
Exposure corroboration (display credibility)	$c = \min\{1, \max(s/0.5, i/0.5)\}$; displayed $Z \times c$; flag below 0.5 from 10 filed PU, and Excellent/Strong held one band lower (interim)	§6, §8
Grade cuts on $\bar{R}$	0.60 (decision) / 0.756 / 1.057 / 1.570 / 2.109 (fitted at nominal 0.80 / 1.10 / 1.40 / 1.90)	§8
Score anchors at band edges	100 / 90 / 75 / 50 / 30 / 10 / 0 (0 at 2 × the Critical cut)	§7
Posterior screens p_{top}, p_{down}, p_{bottom}	0.6, 0.6, 0.8	§8
Excellent screens	$\Lambda \geq 0.3$; driver-behaviour $\tilde{r} \leq 1.0$; no fatal crash; ≥ 3 inspections; ≥ 0.1 inspections per filed PU; > 1 PU	§8
Strong screens	posterior; ≥ 0.1 inspections per filed PU	§8
Critical / Poor event tests	fatal or ≥ 2 crashes or $P(p_{\text{oos}} > 2\mu) \geq 0.8$; ≥ 1 crash or $P(p_{\text{oos}} > 1.5\mu) \geq 0.8$	§8

Current FMCSA Unsatisfactory / Conditional	referral; capped at Marginal	§8
Governor	upward moves ramp 1 band per run unless decay, data correction or a model recalibration; downside ungoverned	§9
Gate bounds	Gini ≥ 0.55 per ranked cohort with enough outcomes to judge (≥ 5,000 rows and 100 realized crashes in the gate year: 10–19, 20–49, 50–99 PU); monotone (2 SE); aggregate 0.9–1.1; evidence ±10% / ±20%; coverage 0.78–0.92; ≥ 2-band ≤ 0.5%; ≥ 1-band ≤ 3% (recalibration moves excluded and reported); governed ≤ 0.5%; rows ≥ 900k; ranked ≥ 80k; OOT year ≥ 120 d mature	§10

Nothing is refit week to week. The artifact's sha256 and the cutoff, tenure and exposure-cap rules are recorded in every run's metrics; a recalibration is a new artifact with its own checksum or a change of one of those rules, and its moves are attributed to recalibration on every carrier page and shown at once.

§B Reporting Lag and the Cutoffs

A grade computed today uses only the records FMCSA has published today; a validation computed on history can accidentally use records that arrived later. The reporting-hold study therefore replayed historical first-of-month run dates R with only the records each run could have seen — crashes by FMCSA add date, inspections and violations by their MCMIS add date — and scored every timing with the frozen model on the **same outcome**: the recordable crashes each carrier had after R. Population, class columns, forward exposure and the inspection-window length were held fixed so only the newest events differed, and the replay reproduced the engine's own frame builder exactly before any timing was compared. Eight run dates (2025-01 to 2025-08) were judged on twelve months of outcome; five later ones (2025-10 to 2026-02), whose inspection windows are full production length, on six. Select split, about 400,000 carriers per run date. The decision rule was fixed in code before the run: a timing is better only if its gain in deviance explained over the two-month hold has a 90% bootstrap interval above zero on at least 75% of run dates, the ranked book's and each cohort's Gini do not fall by more than 0.005, grades stay monotone and the slow-reporting states' calibration stays within ±0.05.

Table B.1 — Timings against holding everything back two months, 13 replayed run dates.

Timing	Crashes through	Inspections through	Better on	Δ deviance explained	Δ Gini, 10+ PU book (median)	Calibration shift, 10+ PU book
Two-month hold (reference)	$R-2$ m	$R-2$ m	–	–	–	–
One-month hold	$R-1$ m	$R-1$ m	13 / 13	+0.002 to +0.004	+0.002 / +0.003	< 1%; 2.5% in the 2025 stall
No hold	R	R	13 / 13	+0.003 to +0.007	+0.003 / +0.005	1–2% kinder; 3–6% in the 2025 stall
Inspections now, crashes two months	$R-2$ m	R	13 / 13	+0.002 to +0.004	+0.001 / +0.002	< 1%
Inspections now, crashes one month (production)	$R-1$ m	R	13 / 13	+0.003 to +0.006	+0.002 / +0.004	< 1%; 2.4% in the 2025 stall

Gini medians are for the six-month / twelve-month runs. Fresh inspections help on every run date. Fresh crash counts help while reporting is normal — a run knew 66–70% of the crashes dated in the two held months — but not when FMCSA's crash reporting stalled in mid-2025: a run on 2025-08-01 knew 7% of the June and July crashes, and counting crashes up to the run date left the ranked book 5.6% too kind. Holding crashes one month keeps most of the gain without that drift. A carrier with a crash or out-of-service inspection in the newest months

is predicted markedly better (deviance explained 0.33–0.35 under the two-month hold, 0.35–0.37 under the production timing).

Table B.2 — Grades that are too kind at the time and drop when a late crash report lands (same cutoff, records known at the run vs. everything known today; 10 origins).

Wait before grading	Kinder at the time	Kinder by 2+ bands	Harsher
None	0.22–0.34% (0.58% in the 2025 stall)	0.02–0.05%	0.01–0.02%
30 days	0.09–0.17% (0.33% in the stall)	≤ 0.03%	≤ 0.01%
60 days	0.07–0.16% (0.33% in the stall)	≤ 0.03%	≤ 0.01%

On the carrier page, an inspection dated after C or a crash dated after C_{crash} is shown as *not yet scored*; it enters the grade as the window moves forward each month.

This document specifies the production FRED scoring engine. Every formula, constant and threshold is taken from the frozen model artifact and the engine code; study figures are from the Safety Core Study and live figures from the latest weekly run (November 1, 2026). The engine re-scores the full property universe weekly with every parameter frozen between scheduled recalibrations.