



American Journal of Innovation in Science and Engineering (AJISE)

ISSN: 2158-7205 (ONLINE)

VOLUME 5 ISSUE 3 (2026)



PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

Carrier Safety Scoring Under Measurement Instability: From CSA Percentile Volatility to Route-Level Risk Allocation

Orlova Natalya Gennadyevna^{1*}

Article Information

Received: December 07, 2025

Accepted: September 05, 2026

Published: October 28, 2026

Keywords

Carrier Safety Scoring, Inspection Exposure, Measurement Reliability, Motor Carrier Regulation, Route-Level Risk Allocation, Stochastic Vehicle Routing

ABSTRACT

Carrier safety percentile scores generated by the Federal Motor Carrier Safety Administration's Safety Measurement System inform enforcement prioritization, insurance underwriting, and, increasingly, automated freight-assignment decisions. Several independent strands of research document that these percentiles behave inconsistently for carriers with limited inspection histories. This article connects that documented inconsistency to a specific decision a route-level or lane-level risk-allocation system must make: whether to treat every published percentile with equal statistical confidence. Drawing together evidence on carrier size, inspection frequency, longitudinal score movement, and carrier-level aggregation, the article argues that a below-threshold CSA percentile carries a different statistical status than an above-threshold one, and proposes this distinction as a design requirement for automated routing systems that already handle other forms of scenario uncertainty. The article specifies the data, unit of observation, and modeling approach a future empirical study would need to test this claim, framing the present contribution as the conceptual groundwork for that test.

INTRODUCTION

A carrier's Compliance, Safety, Accountability percentile compares its record against a peer group of carriers with similar numbers of safety events, a design choice the Safety Measurement System applies because raw violation counts are not comparable across carriers of different operating scale (Knipling *et al.*, 2011). The percentile sits atop a more elaborate scoring architecture than a simple count: violations are sorted into seven Behavior Analysis and Safety Improvement Categories, weighted by severity and by how recently they occurred, and then compared within a peer group before the final 0-100 rank is assigned. This article concentrates on one narrower, well-documented property of that architecture: a percentile built from few observed events carries more sampling variance than one built from many, and carriers operating the majority of U.S. commercial truck-miles face systematically lower inspection frequency than large fleets (Knipling *et al.*, 2011). The threshold at which that lower frequency produces enough variance to invert a carrier's rank between consecutive monthly updates is the empirical question the Conclusion specifies for future testing. The reviewed literature establishes that inspection count and score movement track each other consistently across independent studies, evidence strong enough to treat the relationship as a design problem for any system that consumes the score.

The seven BASICS apply different thresholds: six trigger an intervention at the 65th percentile, and hazardous materials compliance triggers at the 80th percentile, reflecting the lower base rate of hazmat-specific violations across the carrier population. Each BASIC also applies a time-decay weight that discounts

violations by age: a violation recorded within the past six months carries more weight than one recorded eighteen months earlier, and violations older than twenty-four months drop out of the calculation entirely. This decay function lets a carrier's percentile move even when no new inspection occurs, purely because older violations age out of the window. A carrier with few inspections has fewer violations available to decay out gradually, so a single recent violation can dominate its score for a longer stretch of the following two years than the same violation would for a carrier whose peer-group comparison draws on a denser inspection history. That mechanical property compounds the sampling-variance concern documented below: instability in a low-exposure carrier's percentile comes from how few events are observed and from how heavily each observed event weighs once it is recorded. Peer groups themselves are defined by the number of relevant inspections a carrier has accumulated in a rolling twenty-four-month window, so two carriers with the same number of trucks can sit in different peer groups if one has been inspected more often. This indirect construction means fleet size predicts peer-group membership imperfectly, through inspection frequency as an intermediate variable, a distinction the literature reviewed below treats inconsistently: some studies measure firm size directly, others measure inspection count, and the two correlate without being identical.

Where that problem becomes consequential depends on how the percentile gets used. Enforcement targeting tolerates a noisy signal because the cost of a false positive is a compliance review, an administrative burden the agency absorbs across a large portfolio of interventions. A route-level or lane-level risk-allocation system assigns

¹ Deputy Director for Research, Institute of Digital Transport Solutions and Automation, Nizhny Tagil, Russian Federation

* Corresponding author's e-mail: natalya.orlova.gen.1@gmail.com

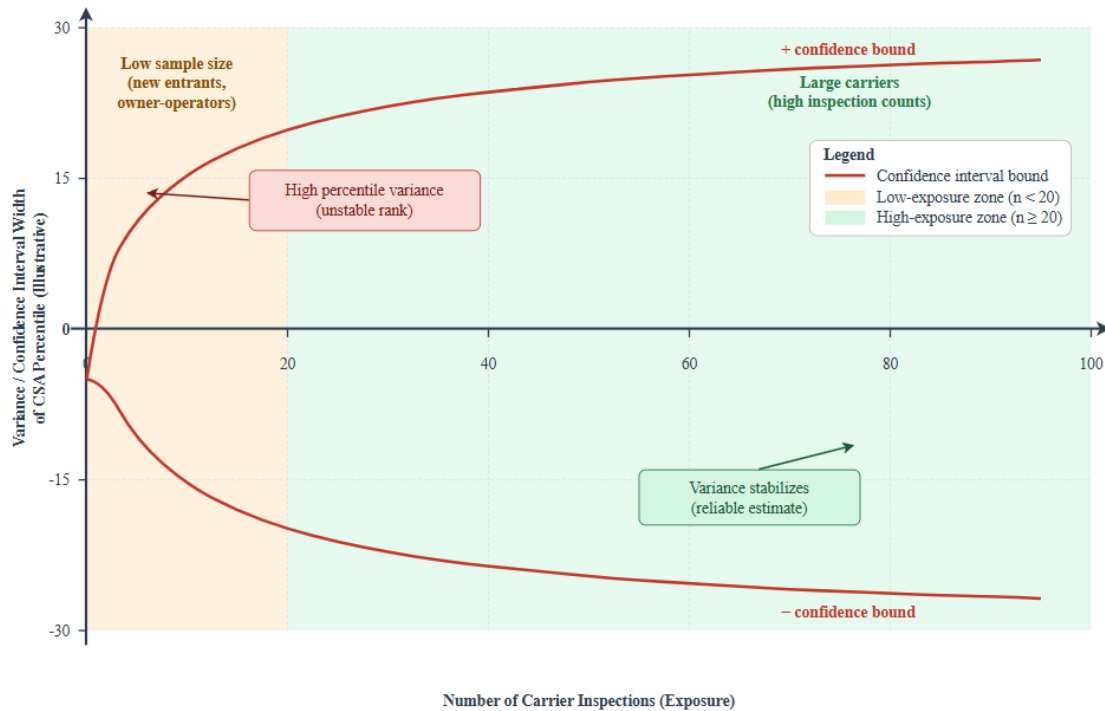


Figure 1: Conceptual relationship between CSA percentile reliability and inspection count

specific loads to specific carriers, so a single unstable rank at the point of assignment carries direct consequences that an enforcement queue spanning thousands of carriers can absorb through averaging. Dorskach (2026b) describes a stochastic capacitated vehicle routing architecture whose dispatch objective already treats one kind of uncertain input, demand forecasts, as a distribution instead of a fixed number, using sample-average approximation across fifty scenarios. The same logic extends directly to a differently structured kind of uncertainty rooted in regulatory measurement, and this article develops that extension. This narrative review draws on a body of motor-carrier safety and operations-research literature identified through the citation threads described in the Materials and Methods section, and its claims stay bounded by that literature.

LITERATURE REVIEW

The idea that inspection exposure shapes score reliability predates the CSA program. Moses and Savage (1994) found, using pre-CSA data, that firm size predicts truck accident rates independent of route characteristics. That finding set an expectation any later scoring system would need to confront: carriers of different sizes generate structurally different volumes of observable safety data, so a system comparing them directly risks confusing data volume with underlying risk. Knipling *et al.* (2011) document the CSA-era version of this problem directly, reporting that carriers operating most U.S. truck-miles receive disproportionately fewer inspections than their share of exposure would imply, which places them in smaller peer groups where each individual violation shifts the percentile by more.

Two studies extend this size-exposure link in different directions. Cantor *et al.* (2016) find that firm size and prior safety performance jointly predict technology adoption, which raises a secondary concern beyond exposure alone: carriers with the fewest inspections are often also the carriers least likely to have adopted the monitoring systems that could supply an alternative, higher-frequency data stream. Scott and Nyaga (2019), working from a different angle, find that firm size and asset ownership predict the probability of a regulatory violation net of market price effects, a result that complicates a purely statistical reading of the size-instability relationship. If size predicts the underlying violation rate and not merely the number of opportunities to observe it, then some of what looks like measurement noise at small scale may be a genuine difference in operating risk. This article treats that possibility as a live alternative explanation: separating a sampling-variance effect from a true risk-level effect calls for violation rates measured against a common, size-independent exposure denominator, the kind of benchmark the empirical study outlined in the Conclusion would supply.

An earlier study by the same research group anticipated this problem before the CSA program took its current form. Cantor *et al.* (2008) examined what predicts a motor carrier's adoption of on-board safety technology and found that adoption decisions cluster by firm size well before regulatory scoring created a direct incentive to close the exposure gap. Carriers slow to adopt monitoring technology in the pre-CSA period that study identifies are, on the size distribution the industry has retained since, disproportionately the same population Knipling *et al.* (2011) later found to be under-inspected.

The persistence of that adoption pattern across roughly two decades of regulatory change suggests the exposure gap is a structural feature tied to firm size that predates CSA and has outlasted several rounds of methodology revision. A route-level system built today inherits that history. Electronic logging devices, now mandatory across most of the interstate fleet, generate a continuous stream of hours-of-service data that could supplement sparse inspection records for exactly the carriers Cantor *et al.* (2008) identified as slow adopters of earlier monitoring technology. Whether that mandate has closed the adoption gap the 2008 study documents is a separate empirical question, since ELD compliance speaks directly to one BASIC category and leaves the other six, including vehicle maintenance and driver fitness, dependent on inspection-based data alone.

A second line of evidence tracks how individual carriers' status changes over time. Miller (2017) applied discrete-time hazard modeling to carriers flagged for poor hours-of-service compliance at the CSA program's 2010 launch and found substantial movement in both directions over the following 3.5 years: many flagged carriers exited that status, and a comparable number of previously unflagged carriers entered it, with carrier size shaping the rate of both transitions. This pattern is consistent with the exposure-driven instability described above. Miller's hazard model estimates transition rates, which leaves open whether a carrier's true risk genuinely changed or its percentile moved through sampling variance alone. Miller *et al.* (2017) respond to this open question by arguing that a carrier's safety trajectory, tracked over multiple periods, is a more defensible unit of analysis than any single percentile snapshot, an argument that treats the point-in-time score as inherently provisional. Miller and Saldanha (2016) add a further candidate mechanism: financial performance and safety outcomes move together longitudinally within carriers, so a firm's score volatility may reflect its financial volatility as much as its safety practice, a mechanism that remains entangled with the exposure effect in the hazard-rate literature.

Aggregation compounds the difficulty of separating these explanations, because the CSA percentile is computed at the carrier level while the underlying events happen at the level of individual vehicles, drivers, and trips. Lack *et al.* (2021) show that FMCSA inspection violations aggregated to the carrier level correlate with crash risk in the aggregate, and that this correlation coexists with substantial internal heterogeneity: a carrier with one severely deficient vehicle and a carrier with mild noncompliance spread evenly across a fleet can produce the same percentile. This finding cuts two ways for the present argument. It supports the claim that the carrier-level score obscures information the routing problem might need. It also complicates any simple story in which few inspections automatically means an unreliable score, since a carrier with many inspections but concentrated risk is unreliable for a different reason entirely, a pattern that exposure count alone leaves invisible.

Inspection intensity also varies by geography. Roadside inspections concentrate at weigh stations and fixed enforcement sites, so a carrier whose routes run through states with a denser network of inspection facilities accumulates exposure faster than a carrier of identical size and safety practice operating primarily through states with sparser enforcement infrastructure. Lack *et al.* (2021) aggregate violations to the carrier level across states and corridors, so their crash-risk correlation could partly reflect enforcement geography as much as underlying safety differences between carriers. A carrier operating long-haul lanes through a corridor where inspection stations are numerous and inspection rates run high would accumulate a denser exposure history than a carrier of comparable size operating primarily through less densely enforced territory, independent of any difference in vehicle condition or driver behavior. This geographic confound sits alongside the firm-size confound documented above and proves harder to correct for, since a carrier's operating geography shifts load by load, while its fleet size stays comparatively fixed. The elevated intervention threshold for hazardous materials compliance concentrates the exposure problem further, since the population of carriers authorized to transport hazardous materials is smaller than the general carrier population and receives a correspondingly narrower base of hazmat-specific inspections. Toumazis and Kwon (2016) treat the sparseness of historical hazmat accident data as a defining feature of the routing problem they model, choosing a worst-case framework precisely because a stable probability distribution cannot be estimated from so few observed events. The same sparseness constrains the hazmat BASIC score for any individual carrier operating in that population, so a hazmat-authorized carrier with a below-threshold percentile in that specific category carries a compounded exposure problem, few hazmat inspections layered on top of whatever general inspection frequency its fleet size and operating geography would otherwise predict. The CSA-focused studies reviewed above treat percentile instability as a property of the overall score, a framing that leaves the hazmat 's own exposure dynamics unexamined, even though its 80th-percentile threshold and smaller carrier population suggest it may be the most exposure-sensitive of the seven. That estimation step would let a routing objective built on Daskach's (2026b) sample-average-approximation logic be tested against a risk-neutral baseline to determine whether exposure-adjusted percentiles change assignment outcomes for the below-threshold population specifically, and whether any such adjustment invites the strategic response Scott *et al.* (2021) document for the electronic logging device mandate.

Low-exposure carriers are not distributed evenly across the industry. Cantor *et al.* (2017) find that new entrants subject to the New Entrant Safety Assurance Program show weaker safety performance than pre-program entrants, and that completing the program's audit

successfully predicts better subsequent performance. Miller and Saldanha (2018) track new entrants longitudinally and find the performance gap relative to established carriers narrows within their observation window. A developmental account, in which entrants simply learn safer practices over time, fits this pattern, and so does a purely mechanical account, in which inspection exposure for a new entrant stays low in its early years regardless of how quickly its practices improve. The two accounts look observationally similar across Cantor *et al.* (2017) and Miller and Saldanha (2018) alike.

Owner-operators present a related but distinct case. Kim *et al.* (2024) find that owner-operator experience moderates violation outcomes, which indicates that treating owner-operators as a single risk category obscures a tenure effect running alongside an exposure effect, since more experienced owner-operators have accumulated more inspections as well as more driving history. Miller *et al.* (2018) review competing theoretical accounts of why owner-operators show different safety profiles than company drivers and trace part of the disagreement across that literature to samples drawn from carrier populations with different exposure characteristics, a claim an exposure-adjusted outcome measure could test directly.

An alternative to carrier-level scoring exists in the literature: driver-level scoring, which shifts the unit of analysis from the fleet to the individual driver. Cantor *et al.* (2010) build a crash-prediction model on driver-level covariates, violation history and tenure among them. A driver's own violation history remains limited by how often that driver has been inspected, so this approach carries its own exposure sensitivity, and it demonstrates, at minimum, that a finer-grained modeling approach is technically available. Scott *et al.* (2021) supply a caution against assuming any new signal, driver-level or otherwise, would prove immune to strategic response: their finding that electronic logging device adoption reduced hours-of-service violations while unsafe-driving violations rose shows that a narrowly defined compliance signal invites behavior that satisfies the letter of the signal while leaving the underlying risk unaddressed. A driver-level or exposure-adjusted score would need evaluation against this same possibility before it could count as a reliable solution.

Miller (2020) finds that larger carriers invest more heavily in dedicated safety staff, standardized maintenance schedules, and driver training infrastructure, and that this investment accounts for a substantial share of the compliance gap between large and small carriers independent of fleet size itself. This finding complicates the exposure-driven account developed above: part of the size-compliance relationship stems from genuine organizational investment as well as from statistical artifact, so a discount function that shrinks small-carrier percentiles toward the peer-group mean would correct for only part of the observed gap and leave a real behavioral difference in place. Miller's (2020) finding

refines the exposure account built from Knippling *et al.* (2011): it narrows the share of the size-compliance relationship that exposure alone explains and identifies organizational investment as a second mechanism operating alongside sampling variance. Small carriers and owner-operators typically lack the capital to fund a dedicated safety department, which means a policy response built solely around statistical correction leaves the underlying resource gap Miller (2020) identifies in place. A route-level system that discounts a small carrier's percentile for low exposure models only one of the two mechanisms this literature identifies, leaving the other for a complementary policy lever.

Separately from the CSA literature, operations research has developed formal methods for incorporating measured uncertainty into routing decisions. Toumazis and Kwon (2016) build a worst-case conditional value-at-risk approach for hazardous materials routing specifically because historical accident data in that domain is too sparse to support a stable probability distribution, treating data scarcity as a parameter to model within the routing problem itself. Lianes *et al.* (2018) extend a risk-averse orientation to general network routing, modeling travel-time variance alongside expected travel time and showing that a risk-averse routing policy diverges from an expected-cost-minimizing one as input variance rises. Both treat variance as an exogenous input supplied to the model in advance, which leaves open a further step: estimating that variance from a peer-group structure such as the one CSA scoring already provides.

Both operations-research studies model variance as a property of the route or the hazard. A CSA percentile's variance is partly a property of the underlying safety event and partly a property of how many times that carrier has been inspected, a distinction that separates it from travel-time variance or hazardous-materials accident variance, since neither carries an analogous inspection-count denominator. Extending either framework to a carrier-safety application means adding a second source of uncertainty, measurement uncertainty layered on top of outcome uncertainty, to models that currently capture only the outcome side. A trucking-specific implementation of this variance-aware logic already exists: Dorskach's (2026b) sample-average-approximation dispatch architecture applies exactly this kind of scenario-weighted treatment to demand forecasts, and extending it to a regulatory percentile with the exposure structure documented above is the natural next step this article develops.

Dorskach's (2026b) SCVRPTW architecture provides the clearest working model in this review of a system built with the right kind of scaffolding, because its objective function already down-weights less-supported scenarios in a sample-average approximation, weighting scenarios according to how well-supported they are. The parallel to CSA-percentile uncertainty is structural: both cases involve a point estimate that should, in principle, be treated as a draw from a distribution whose width depends on

the amount of underlying data. The architecture's SAA mechanism, built for a smoothly forecastable variable like demand, extends directly to a regulatory score whose variance stems from inspection sampling, peer-group composition, and BASIC-specific weighting rules, and specifying that extension for carrier-level risk is the central task this synthesis takes up.

MATERIALS AND METHODS

This article follows a narrative synthesis design. Sources were identified through citation-thread tracing, beginning from two anchor works, Knipling *et al.* (2011) for the CSA-percentile literature and Toumazis and Kwon (2016) for the risk-aware vehicle-routing literature, and expanding outward through each anchor's own citations and the citing literature indexed in Transportation Research Part E, Journal of Business Logistics, Transportation Journal, and Journal of Operations Management. A source qualified for inclusion if it reported an empirical or theoretical relationship among carrier or firm size, inspection frequency or exposure, and CSA percentile movement or motor-carrier violation outcomes, or if it applied a risk-aware or variance-sensitive formulation to a vehicle-routing or transportation-planning objective. Sources focused on a single BASIC category needed to address exposure or aggregation directly to qualify, with one exception: the hazmat-specific literature, where exposure was the only lens available for that category. The synthesis proceeds by identifying convergence and divergence across the retained sources, since the sources use incompatible outcome measures. Three explanatory mechanisms recur across the literature, sampling variance tied to inspection count, firm-level differences in true risk, and geographic variation in enforcement density, and each retained source was read against which of these mechanisms, or combination of mechanisms, its design can distinguish from the others. That coding is reported qualitatively in the Results and Discussion section below.

RESULTS AND DISCUSSION

Read against the criteria above, the retained sources support four results that the discussion below interprets.

1. Inspection count predicts percentile instability independent of the underlying safety event. Knipling *et al.* (2011) and Miller (2017) document this relationship using different data structures and different time periods, which is stronger evidence than either study alone would provide.

2. Firm size operates through at least two distinct channels. It predicts inspection frequency, per Knipling *et al.* (2011), and it independently predicts the probability of a true violation, per Scott and Nyaga (2019), so a single size-based correction conflates two distinct mechanisms.

3. Organizational investment explains part of the size-compliance gap that a purely statistical correction would otherwise treat as noise. Miller (2020) attributes a substantial share of that gap to dedicated safety staffing and maintenance infrastructure that larger carriers can

fund and smaller carriers typically cannot.

4. The population with the least stable scores, new entrants and owner-operators, per Cantor *et al.* (2017) and Kim *et al.* (2024), is also the population with the least margin buffer against a misclassified score, per Dorskach (2026a), so measurement error and economic exposure to that error concentrate in the same carriers.

The literature reviewed above converges on a specific asymmetry: the CSA percentile is least statistically stable for the population, new entrants and owner-operators, that a route-level risk-allocation system would most need to differentiate reliably. Cantor *et al.* (2017) and Kim *et al.* (2024) document that this population carries measurably different and more heterogeneous risk than established, larger carriers. Knipling *et al.* (2011) document that this same population generates the fewest inspections. A routing system indifferent to CSA percentile forgoes differentiation for exactly the carriers most likely to need it, and a routing system that treats every published percentile with equal confidence risks penalizing small and new carriers whose extreme score may reflect small-sample variance more than elevated risk. Scott and Nyaga's (2019) finding that size also predicts true violation probability means this risk cuts in both directions: some below-threshold carriers with poor scores may in fact be genuinely riskier.

The economic stakes of this differentiation fall unevenly across the carrier population. Dorskach (2026a) documents that operating margins in the U.S. trucking industry run thin enough that a carrier's access to consistent freight volume determines survival more directly than most other operating decisions, and that small and new carriers hold the least margin buffer against a lost lane award. A route-level system that treats a below-threshold percentile as disqualifying, without adjusting for the exposure behind it, compounds the misclassification: it removes precisely the freight volume a thinly capitalized carrier needs to invest in the safety infrastructure Miller (2020) identifies as the more durable driver of compliance. The two dynamics reinforce each other in a feedback loop the reviewed studies document only in separate pieces: low exposure produces an unreliable score, an unreliable score reduces freight access, and reduced freight access limits the investment that would improve the carrier's underlying safety practice and, eventually, its inspection history. Dorskach (2026c) frames the CSA percentile as the primary channel through which parties without direct access to a carrier's inspection history, including insurers and freight brokers, assess that carrier's safety standing, which is precisely why the exposure behind the score matters to audiences well beyond the enforcement agency that publishes it. Insurance underwriters draw on the same percentile for a related but distinct purpose, setting premiums that further constrain a thinly capitalized carrier's margin. A carrier whose percentile is elevated by sampling variance and not genuine risk faces a compounding cost, a higher premium layered on top of reduced freight access, both flowing from the same

statistical instability documented in the studies above. Underwriting practice and route-level allocation both currently treat the percentile as a self-contained number, so both channels transmit the same measurement problem into carrier-level economic outcomes.

This ambiguity is central to how much weight the paper's design principle can carry, and deserves a plain statement. The correlation between low inspection exposure and score instability observed across Knipling *et al.* (2011) and Miller (2017) is consistent with several distinct mechanisms operating simultaneously, and separating them calls for a dataset broader than the sources this review draws on. Sampling variance is one candidate mechanism. Genuine risk differences correlated with firm size, as Scott and Nyaga (2019) suggest, are another. Differences in operating geography, cargo type, route composition, and fleet utilization intensity could independently produce both low inspection counts and unstable scores if inspectors target certain corridors, cargo classes, or vehicle types more heavily than others, a possibility the reviewed sources leave open because none measures these factors directly. A carrier's prior regulatory history could also shape both its current inspection frequency, through targeted follow-up inspections, and its current score, through the time-weighting the Safety Measurement System applies to recent violations. Disentangling these mechanisms requires a dataset that jointly observes inspection counts, percentile trajectories, and at minimum operating geography and cargo mix, the same carrier-BASIC-month panel the Conclusion specifies for future work.

Given that limitation, this article's central contribution is a design principle: a route-level system should treat a CSA percentile as dependent on the inspection count behind it, because Miller's (2017) hazard-rate evidence shows that low-exposure scores move in ways a stable underlying-risk signal would rarely produce. Converting that principle into an actual discount function, one that shrinks a carrier's percentile toward its peer-group mean in proportion to inspection count, is the natural next step built directly on Dorskach's (2026b) SAA-based demand-uncertainty treatment. Specifying the discount function requires, at minimum, an estimate of how percentile variance depends on inspection count within each BASIC category and peer group, the estimate the Conclusion identifies as the starting point for that empirical work.

Any such mechanism would also need to anticipate the strategic response Scott *et al.* (2021) document for the electronic logging device mandate. A discount function legible to carriers gives a below-threshold carrier an incentive to manage its inspection exposure directly, for example by minimizing exposure to stay inside a protective confidence band, which can compete directly with the incentive to invest in genuinely safer operations. This risk argues for pairing any exposure-based discount with a signal less easily manipulated at the point of measurement, such as the driver-level covariates Cantor *et al.* (2010) demonstrate are usable. That pairing introduces its own aggregation and exposure issues at the driver level, issues

that warrant the same kind of exposure-sensitive analysis developed here. Dorskach (2026c) treats percentile scoring as the primary mechanism through which the regulatory system differentiates carrier safety. The evidence gathered here is consistent with that account of the system's design intent and adds a qualification: the instrument performs that differentiation with uneven precision across the carrier population.

A shrinkage approach of this kind has a specific statistical form worth stating precisely. Empirical Bayes shrinkage would treat each carrier's observed percentile as a noisy estimate of an underlying true risk level, pooling information across the peer group to estimate how much of the observed spread in percentiles within a peer group reflects genuine risk variation and how much reflects sampling noise given the group's typical inspection count. A carrier's shrunk percentile would then sit somewhere between its raw observed score and the peer-group mean, with the weight on the raw score increasing as its own inspection count rises. This is the same logic behind the sample-average approximation Dorskach's (2026b) architecture applies to demand scenarios, restated for a peer-group variance structure instead of a scenario-weighted forecast. The estimation itself requires the carrier-BASIC-month panel described in the Conclusion below, which would let the shrinkage weights be calibrated to the actual variance-exposure relationship documented qualitatively above; the panel converts this relationship from an assumed one into an estimated one.

The limitations of a narrative synthesis matter for how much weight this design principle should carry before empirical testing. A systematic review with a registered search protocol and formal quality appraisal would likely surface additional studies, particularly work published in operations-research venues outside the transportation-economics journals that dominate the citation threads followed here. Because this synthesis follows citation threads, selection of which studies to include reflects judgment calls about relevance, and a different reviewer working from the same broad question might have drawn a different but overlapping set of sources. This limitation leaves the specific claim that inspection exposure and percentile reliability are linked intact: that claim rests on multiple independent studies using different methods and different time periods, a stronger form of evidence than any single study could provide on its own. It does mean the design principle proposed here is best read as a starting hypothesis for the empirical work outlined in the Conclusion.

CONCLUSION

This article converts a set of independently documented relationships, inspection exposure, firm size, organizational investment, and geographic enforcement density, into a single design proposition: a route-level or underwriting system that treats every published CSA percentile with equal statistical confidence will systematically misjudge exactly the carriers, new entrants and owner-operators, it most needs to differentiate correctly. That proposition is

conceptual, and testing it against carrier-level data is the next step, specified below.

Testing it requires a specific dataset and a specific sequence of steps the field has not yet assembled. FMCSA Safety Measurement System records at the carrier-BASIC-month level, spanning multiple years and linked to fleet size, primary cargo class, and operating geography, would let a future study estimate how percentile variance depends on inspection count within each peer group before any shrinkage weight is proposed. That estimate would need to hold across BASIC categories separately, since the hazmat category's elevated 80th-percentile threshold and smaller carrier population make its exposure sensitivity a distinct empirical question from the other six.

Even before that empirical work is complete, a route-level or underwriting system can take one practical step consistent with this synthesis: report the inspection count behind a below-threshold percentile alongside the score itself, so a downstream decision, automated or human, gains the information needed to distinguish an unreliable score from a genuinely elevated one.

REFERENCES

- Cantor, D. E., Corsi, T. M., & Grimm, C. M. (2008). Determinants of motor carrier safety technology adoption. *Transportation Research Part E: Logistics and Transportation Review*, 44(5), 932–947. <https://doi.org/10.1016/j.tre.2007.03.001>
- Cantor, D. E., Corsi, T. M., & Grimm, C. M. (2017). The impact of new entrants and the new entrant program on motor carrier safety performance. *Transportation Research Part E: Logistics and Transportation Review*, 97, 217–227. <https://doi.org/10.1016/j.tre.2016.11.005>
- Cantor, D. E., Corsi, T. M., Grimm, C. M., & Özpölat, K. (2010). A driver focused truck crash prediction model. *Transportation Research Part E: Logistics and Transportation Review*, 46(5), 683–692.
- Cantor, D. E., Corsi, T. M., Grimm, C. M., & Singh, P. (2016). Technology, firm size, and safety: Theory and empirical evidence from the U.S. motor-carrier industry. *Transportation Journal*, 55(2), 149–167.
- Doskatch, H. (2026a). *Building a profitable trucking company in the U.S.* LAP LAMBERT Academic Publishing.
- Doskatch, H. (2026b). Mitigating the U.S. truck driver shortage through AI-driven route optimization. *Scientia. Technology, Science and Society*, 3(5), 61–68. [https://doi.org/10.59324/stss.2026.3\(5\).04](https://doi.org/10.59324/stss.2026.3(5).04)
- Doskatch, H. (2026c). *Safety and compliance in commercial transportation.* LAP LAMBERT Academic Publishing.
- Kim, S., Park, A., & Wagner, S. M. (2024). Owner-operator experience and violations in trucking: The mixed role of perceived justice. *Transportation Research Part E: Logistics and Transportation Review*, 188, 103642. <https://doi.org/10.1016/j.tre.2024.103642>
- Knipling, R. R., Hyten, C., & Singer, J. (2011). Safety management in small motor carriers (CTBSSP Synthesis 22). *Transportation Research Board, National Academies Press.*
- Lack, C. D., Berkow, K. S., & Gao, Y. (2021). Insights into motor carrier crashes: A preliminary investigation of FMCSA inspection violations. *Accident Analysis & Prevention*, 155, 106105. <https://doi.org/10.1016/j.aap.2021.106105>
- Liancas, T., Nikolova, E., & Stier-Moses, N. E. (2018). Risk-averse selfish routing. *Mathematics of Operations Research*, 44(1), 38–57. <https://doi.org/10.1287/moor.2017.0913>
- Miller, J. W. (2017). Discrete time hazard modeling of large motor carriers' longitudinal CSA performance. *Transportation Journal*, 56(2), 107–139. <https://doi.org/10.5325/transportationj.56.2.0107>
- Miller, J. W. (2020). Why are larger motor carriers more compliant with safety regulations? *Transportation Journal*, 59(1), 28–72. <https://doi.org/10.5325/transportationj.59.1.0028>
- Miller, J. W., Golcic, S., & Fugate, B. (2017). Developing and testing a dynamic theory of motor carrier safety. *Journal of Business Logistics*, 38(2), 96–114. <https://doi.org/10.1111/jbl.12149>
- Miller, J. W., Golcic, S. L., & Fugate, B. S. (2018). Reconciling alternative theories for the safety of owner-operators. *Journal of Business Logistics*, 39(2), 101–122.
- Miller, J. W., & Saldanha, J. P. (2016). A new look at the longitudinal relationship between motor carrier financial performance and safety. *Journal of Business Logistics*, 37(3), 284–306.
- Miller, J. W., & Saldanha, J. P. (2018). An exploratory investigation of new entrant motor carriers' longitudinal safety performance. *Transportation Journal*, 57(2), 163–192. <https://doi.org/10.5325/transportationj.57.2.0163>
- Moses, L. N., & Savage, I. (1994). The effect of firm characteristics on truck accidents. *Accident Analysis and Prevention*, 26(2), 173–179.
- Scott, A., Balthrop, A., & Miller, J. W. (2021). Unintended responses to IT-enabled monitoring: The case of the electronic logging device mandate. *Journal of Operations Management*, 67(2), 152–181. <https://doi.org/10.1002/joom.1110>
- Scott, A., & Nyaga, G. N. (2019). The effect of firm size, asset ownership, and market prices on regulatory violations. *Journal of Operations Management*, 65(7), 685–709. <https://doi.org/10.1002/joom.1059>
- Toumazis, I., & Kwon, C. (2016). Worst-case conditional value-at-risk minimization for hazardous materials transportation. *Transportation Science*, 50(4), 1174–1187. <https://doi.org/10.1287/trsc.2015.0639>