



THE RELIABILITY OF CSA DATA AND SCORES

December 2013

Executive Summary

Compliance, Safety, Accountability (CSA) is FMCSA's safety monitoring and measurement system used to identify unsafe carriers and prioritize them for future interventions (e.g., audits). The agency also encourages third parties to use CSA Safety Measurement System (SMS) scores as a tool for making safety-based business decisions.¹ FMCSA hopes to leverage the power of the marketplace to make judgments about carriers and, as a result, compel them to improve their safety performance. SMS scores also have the potential to be used by plaintiffs' attorneys and prosecutors in the context of post-crash litigation.

The use of SMS scores by third party stakeholders and its evaluation by judges raise obvious questions about the accuracy and reliability of the data. For stakeholders such as shippers and brokers the question is whether or not the scores can be routinely relied upon to make sound, beneficial judgments about the safety posture of individual carriers. Similarly, courts must be concerned with whether or not SMS data meet Federal and jurisdictional rules of evidence which require that the data be "trustworthy"² and rest "on a reliable foundation."³

Researchers have arrived at mixed conclusions with respect to the reliability of SMS scores in identifying unsafe (crash prone) motor carriers. Some found virtually no correlation between scores and crash rates in any of the measurement categories.⁴ However the American Transportation Research Institute (ATRI), using a better prediction model, found a positive relationship between scores and crash risk in three of the publicly available measurement categories (BASICS) but also found that scores in two others bear an inverse relationship to crash risk.⁵ Of the non-publicly available categories, scores in one (the *Crash Indicator* BASIC) likely correspond well to future crash involvement,⁶ but scores in the other (the *HM Compliance* BASIC) do not. ATRI also pointed out that the number of alerts that a carrier has been assigned is a strong indicator of crash risk.⁷ However, the strength of the relationship varies depending on the BASICS in which the carrier has alerts, since scores in some BASICS more strongly correlate with crash risk than those in others.

The relationship between scores and crash risk is impacted by a number of data and methodology problems that plague the system. These include: a substantial lack of data, particularly on small carriers who comprise the bulk of the industry; regional enforcement disparities; the questionable assignment of severity weights to individual violations; the underreporting of crashes by states; the inclusion of crashes that were not caused by motor carriers; and the increased exposure to crashes experienced by carriers operating in urban environments.

¹ *Carrier Safety Measurement System Methodology, Version 3.0, Revised December 2012*, FMCSA, page 1-2.

² Federal Rules of Evidence, Rule 803 (8) (B).

³ *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993) United States Supreme Court.

⁴ Gallo, A. P. & Busche, M., *CSA: Another Look with Similar Conclusions*, Wells Fargo Securities Equity Research, July 12, 2012; Gimpel, J. *Statistical Issues in the Safety Measurement and Inspection of Motor Carriers*. Alliance for Safe, Efficient and Competitive Truck Transportation, July 10, 2012.

⁵ American Transportation Research Institute (ATRI), *Compliance Safety Accountability: Analyzing the Relationship of Scores to Crash Risk*, October 2012, page vii.

⁶ Note that crash involvement does not imply cause.

⁷ ATRI, page 30.

Though there are statistical correlations between SMS scores in certain categories and crash risk, as well as between the total number of alerts assigned and crash risk, individual carriers' scores can be unreliable indicators of their safety performance. The identified correlations between scores and crash risk represent industry-wide trends that often don't hold true for individual carriers. In most BASICs there are thousands of carriers ("exceptions") whose scores contradict the trends (i.e. carriers with high scores but low crash rates and vice-versa). The sheer number of "exceptions" and the presence of numerous data and methodology problems lead to the conclusion that SMS scores alone as measures of individual carrier safety performance are, at a minimum, unreliable.

For more information contact Rob Abbott, Vice President of Safety Policy, at rabbott@trucking.org or P. Sean Garney, Manager of Safety Policy at sgarney@trucking.org.

Introduction

Compliance, Safety, Accountability (CSA) is FMCSA's safety monitoring and measurement system used to identify unsafe carriers and prioritize them for future interventions. According to the CSA Safety Measurement System (SMS) methodology, "*The goal of CSA is to implement more effective and efficient ways for FMCSA, its State Partners, and the trucking industry to prevent commercial motor vehicle (CMV) crashes, fatalities, and injuries.*"⁸ Moreover, FMCSA uses the SMS to assign scores to motor carriers based on comparative safety performance in order to intervene with the least safe operators. The agency then strives to compel them to change their behavior and, failing that, takes steps to remove them from the industry.

The SMS methodology also sets out a second purpose of the system: use by third parties to make safety-based judgments about motor carriers. Specifically, the methodology says, "*In turn, this information will empower motor carriers and other stakeholders involved with the motor carrier industry to make safety-based business decisions.*"⁹ These "other stakeholders," such as shippers, brokers, financial institutions and insurers, are presumably encouraged to use SMS data for carrier selection, pricing, and the like. By doing so, it appears FMCSA hopes to leverage the power of the marketplace to compel motor carriers to improve their safety performance. However, in apparent contradiction, the FMCSA website that displays carriers' scores includes a disclaimer which says: "*Readers should not draw conclusions about a carrier's overall safety condition simply based on the data displayed in this system.*"¹⁰

SMS scores also have the potential to be used by plaintiffs' attorneys and prosecutors in the context of post-crash litigation. For instance, a plaintiff's attorney could contend that a motor carrier was to blame for a crash because it lacked effective, functioning safety management controls, as evidenced by poor SMS data and/or scores. In this scenario, SMS data would be presented by an expert witness who would contend that the carrier's measurements speak to its safety culture.

The potential use of SMS data by third party stakeholders and its evaluation by judges raise obvious questions about the accuracy and reliability of SMS data. For stakeholders such as shipper and brokers, the question is whether or not the scores can be routinely relied upon to make sound, beneficial judgments about the safety posture of individual carriers. Similarly, courts must be concerned with whether or not SMS data meet rules of evidence in their respective venues. These rules generally require that evidence be "trustworthy"¹¹ and rest "on a reliable foundation."¹²

Relationships Between Scores and Crash Risk

The principal use of SMS data is to develop scores of each motor carrier's performance in seven measurement categories called Behavioral Analysis Safety Improvement Categories or "BASICS" (See ATA's "CSA: How It Works" document for additional details).¹³ Scores represent percentile rankings of performance compared to carriers of similar size and/or exposure. For instance, a score of 87 suggests that the carrier has worse performance than 87% of carriers of similar size/exposure.¹⁴

⁸ *Carrier Safety Measurement System Methodology, Version 3.0, Revised December 2012*, FMCSA, page 1-1.

⁹ *Ibid*, page 1-2.

¹⁰ See <http://ai.fmcsa.dot.gov/sms/>

¹¹ Federal Rules of Evidence, Rule 803 (8) (B).

¹² *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993) United States Supreme Court.

¹³ Available at <http://www.trucking.org/>

¹⁴ *CSA: Introduction to the Safety Measurement System Version 3.0*, Slide 29, Federal Motor Carrier Safety Administration, Washington, D.C., March 2013, available at <http://csa.fmcsa.dot.gov/>.

Early Research

At issue is whether or not scores represent an accurate measure of crash risk and, if so, are they reliable. Early research on these questions suggested there was little, if any, relationship between scores and crash risk. For instance, a July 2012 analysis conducted by Wells Fargo Securities evaluated the scores and crash rates of the 4,600 largest trucking companies in North America. To summarize, Wells Fargo said: “*We did not find any meaningful statistical correlation between BASIC scores and actual accident incidence on the basis of miles driven or number of power units in our 4,600 carrier dataset.*” The authors went on to say: “*Based on our research, we do not believe stakeholders should rely on CSA BASIC scores as an indicator of carrier safety performance or future crash risk.*”¹⁵

A broader analysis of the system and data, performed by Professor James Gimpel of the University of Maryland, arrived at similar conclusions. After evaluating the SMS data and methodology, Gimpel said: “*There are serious problems with the design of these instruments themselves that render them unreliable. For many carriers in the MCMIS data, the association between crash risk and the BASIC scores is so low as to be irrelevant, which is peculiar given what is commonly understood about the notions of unsafe driving, and the other constructs that BASIC scores are supposed to indicate.*” The author added: “*Consequently, statistical relationships detected in the MCMIS data are not only a cloudy reflection of the true population, but may well be flat wrong.*”¹⁶

ATRI’s Findings

However, a subsequent analysis conducted by the American Transportation Research Institute (ATRI) applied a more rigorous statistical test than Wells Fargo and Gimpel. ATRI contended that these researchers used an inappropriate statistical test by looking only at simple linear correlations between scores and crash rates. Instead, ATRI relied on what it deemed a more appropriate tool, negative binomial modeling. Using this statistical analysis, ATRI found a positive relationship between BASIC scores and crashes in three of the publicly available measurement categories: the *Unsafe Driving*, *Hours of Service Compliance*¹⁷ and *Vehicle Maintenance* BASICS, with the strongest relationship being in the *Unsafe Driving* BASIC. On the other hand, ATRI found a negative relationship between scores in the other two publicly available BASICS and crash involvement. That is, in the *Driver Fitness* and *Controlled Substances and Alcohol* BASICS, higher (i.e. worse) scores were found to be associated with lower crash risks.¹⁸

Though ATRI was only able to evaluate the BASICS in which scores are publicly available, much is known about the remaining measurement categories (*Hazardous Materials Compliance* BASIC and the *Crash Indicator* BASIC). FMCSA has consistently demonstrated that *Crash Indicator* BASIC scores are a strong predictor of future crash involvement (See *Crash Accountability* section of this document). Conversely, however, scores in the *Hazardous Materials (HM) Compliance* BASIC reflect the likelihood of future hazardous materials *violations*, but not the propensity to be involved in *crashes*. In fact, FMCSA has acknowledged that “*The goal of the HM Compliance BASIC is not to predict future crash risk.*”¹⁹

¹⁵ Gallo, A. P. & Busche, M., *CSA: Another Look with Similar Conclusions*, Wells Fargo Securities Equity Research, July 12, 2012.

¹⁶ Gimpel, J. *Statistical Issues in the Safety Measurement and Inspection of Motor Carriers*. Alliance for Safe, Efficient and Competitive Truck Transportation, July 10, 2012. Available Online: <http://asectt.blogspot.com/2012/07/news-brief-university-of-maryland-study.html>

¹⁷ Previously called the *Fatigued Driving* BASIC.

¹⁸ ATRI, page vii.

¹⁹ See Federal Motor Carrier Safety Administration *Safety Measurement System Change*, June 2012 at http://csa.fmcsa.dot.gov/Documents/SMS_FoundationalDoc_final.pdf, page 7.

ATRI also presented a somewhat novel finding with respect to the significance of motor carriers' SMS scores. Through its analysis, ATRI identified a strong correlation between the number alerts assigned to carriers (based on high scores or serious violations found during compliance reviews) and crash risk. ATRI found that:

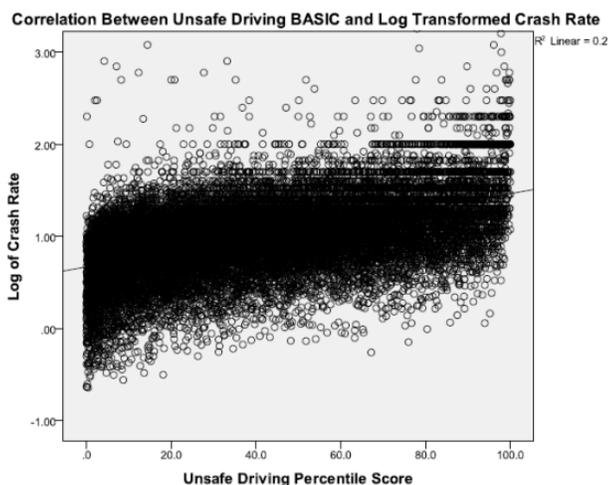
“Compared to a carrier with at least one BASIC score and no ‘Alerts,’ a carrier with a single ‘Alert’ is expected to have a crash rate 1.24 times higher; a carrier with two ‘Alerts’ is expected to have a crash rate 1.61 times higher; a carrier with three ‘Alerts’ is expected to have a crash rate 2.81 times higher; a carrier with four ‘Alerts’ is expected to have a crash rate 3.20 times higher; and a carrier with an ‘Alert’ in all five public BASICs has a crash rate nearly four times higher.”²⁰

With respect to carriers with only one or two alerts, however, this relationship can be misleading. For example, a carrier with an alert in the *Unsafe Driving* BASIC is likely far more crash prone than a carrier with an alert in the *Driver Fitness* BASIC, as indicated by ATRI's findings with respect to individual BASIC scores and crash correlations.

Assessing Individual Carriers

Though ATRI did find a positive statistical correlation between BASIC scores and crashes in three of the publicly available measurement categories, their report offered a number of important caveats. First, ATRI pointed out that its findings regarding the relationship between scores and crash risk may not hold true for every motor carrier. Though the statistical analysis indicates a trend based on data from hundreds of thousands of carriers, there are tens of thousands of carriers whose scores are contrary to the trend. In other words they have high BASIC scores but low crash rates or vice-versa.

The following chart taken from the ATRI study demonstrates the presence of these “exceptions.” Each point represents a carrier's performance in the *Unsafe Driving* BASIC with the BASIC scores depicted on the X axis and crash rates on the Y axis. Though the trend line suggests that higher scores are correlated with increased crash frequency, there is a great deal of variability. In other words, many fleets have high scores but low crash rates or vice-versa.



²⁰ ATRI, page 30.

It is important to point out that a fleet's crash rate may be as much a reflection of happenstance as their safety practices. Crashes are relatively rare events so it's possible that a low crash rate is more an artifact of the carrier's limited exposure (i.e. low mileage or operates in rural area) than the fact that it has a robust safety program. Conversely, a high crash rate may reflect that a carrier with low mileage was unfortunate enough to be involved in a crash (that it did not necessarily cause), resulting in a spike in its rate. This is more likely the case for small carriers that have limited exposure.

The second major caveat ATRI noted was that perceived safety risk is heavily dependent on the amount of available data on each motor carrier. This statement resulted from their finding that carriers with *any* amount of data appeared to have higher crash rates than carriers with no data in the system. As a result, ATRI surmised: *"For instance, it would be specious to conclude that carriers with insufficient roadside inspection data truly have the safest operations of all motor carriers simply because they are absent from both the SMS and MCMIS crash databases."*²¹ This is significant since only 19 percent of active carriers in the ATRI dataset had adequate data to be scored in one or more of the publicly available BASICS that ATRI reviewed and only 23% had some minimal amount of data in the system, but lacked sufficient violations to warrant a score.²²

Data Sufficiency

As ATRI identified, one of the more significant problems impacting the SMS is the lack of data available to assess the performance of the majority of regulated carriers. FMCSA contends that it has sufficient data to "assess" the performance of roughly 200,000 of the estimated 525,000 active motor carriers, or slightly less than 40% of the industry, in at least one BASIC.²³ To be "assessed" in any BASIC a carrier must first meet data sufficiency tests. For instance, in the *Unsafe Driving* and *Hours of Service Compliance* BASICS the carrier must have had at least three relevant inspections and in the *Driver Fitness, Vehicle Maintenance and HM Compliance* BASICS at least five relevant inspections.²⁴

To be assigned a score (i.e. not merely be assessed), however, a carrier must meet the data sufficiency tests *and* have negative data (e.g., violations or crashes). As of late 2013, FMCSA had sufficient data to assign scores to almost 18 percent of active motor carriers in at least one BASIC.²⁵ As of early 2013, FMCSA only had sufficient data to assign scores to 3 percent of active carriers in all BASICS.²⁶ The percentage of active carriers scored in a majority of categories (e.g., at least four BASICS) is not known.

The lack of data is significant since scores are based on comparative performance. In other words they suggest that, relative to others, a carrier performed well or poorly. However, relative measures are dependent on the composition of carriers on whom the system has data upon which to compare. For instance, if SMS has more data on generally safe carriers, a moderately safe company may score poorly compared to those on whom the agency has data – but not if it were to be compared against all other similarly situated carriers.

²¹ ATRI, Page 26.

²² Ibid, page 11.

²³ FMCSA Presentation at The American Trucking Association's Management Conference and Exhibition, October 21, 2013.

²⁴ *Carrier Safety Measurement System Methodology, Version 3.0, Revised December 2012*, Federal Motor Carrier Safety Administration.

²⁵ FMCSA Presentation at The American Trucking Association's Management Conference and Exhibition, October 21, 2013.

²⁶ House Committee on Transportation and Infrastructure, Subcommittee on Highways and Transit hearing *"Evaluating The Effectiveness of DOT's Truck And Bus Safety Programs"* September 13, 2012, written follow-up responses by FMCSA to questions posed by committee members, provided January 2013.

Small Carrier Impact

This limitation has a particularly acute impact on small carriers for two reasons. First, FMCSA lacks data on the vast majority of small carriers. According to data from FMCSA's Motor Carrier Management Information System, 90% of carriers have six or fewer trucks.²⁷ However, according to a Government Accountability Office (GAO) review of the CSA Operational Model Test, a beta test on CSA conducted before nationwide implementation of the program, FMCSA only had adequate data to score 5.7 percent of small carriers (those with five or fewer trucks) in any BASIC.²⁸ Though the Operational Model Test was limited to nine states, it is reasonable to assume that the composition of the data from all other states is similar.

Second, due to the limited amount of data they generate, small carriers' scores are very volatile. In other words, a single violation or two can cause a small carrier's score to swing widely compared to others of similar exposure. Conversely, for large carriers with lots of data in the system a single violation will have little impact on their measures (see explanation of measures in *CSA: How it Works*).²⁹

Underreporting of Crashes

FMCSA contends that the lack of data in the SMS is not a significant problem since the carriers on which it has data are involved in over 90 percent of the crashes reported to the agency – the program's target population.³⁰ However, this contention is misleading on two fronts. First, this claim fails to take into account that many crashes simply don't get reported to FMCSA. A series of assessments conducted by the University of Michigan Transportation Research Institute (UMTRI) over the past decade found substantial underreporting by many states, with some reporting less than 20% of qualifying crashes. UMTRI's most recent analyses (e.g., those evaluating state performance in the past five years) found some states reported only 30-40% of qualifying crashes.³¹

Second, the contention that FMCSA has data on carriers involved in 90 percent of reported crashes is an artifact of the composition of the industry. While a small carrier may only have one or two crashes each year, a large carrier will have dozens, if not hundreds. By having data on most large carriers (comprising less than 10% of all motor carriers), the agency can contend to have information on "carriers involved in most crashes." However, FMCSA probably cannot claim to have data on "most carriers involved in crashes."

Regional Enforcement Disparities

One problem that impacts the reliability of carriers' SMS scores is the presence of tremendous disparity in enforcement practices between states. Some states conduct more robust enforcement of certain laws and regulations (e.g., speeding, seat belt use), which skews the SMS comparative scores. Carriers operating in these states may be perceived to be less safe not because they necessarily commit these violations more frequently, but because they are far more likely to be cited for such violations in the jurisdictions in which they operate.

A recent analysis conducted by Vigillo, a data and service provider to the trucking industry, found that ten states (including Indiana) account for almost half of all commercial motor vehicle

²⁷ *American Trucking Trends*, 2013, American Trucking Associations, Arlington, VA, page 4.

²⁸ *Motor Carrier Safety: More Assessment and Transparency Could Enhance Benefits of New Oversight Program*. Washington, DC: U.S. Government Accountability Office. Retrieved from <http://www.gao.gov/new.items/d11858.pdf>, 2011.

²⁹ Available at <http://www.trucking.org/>

³⁰ FMCSA Presentation at The American Trucking Association's Management Conference and Exhibition, October 21, 2013.

³¹ FMCSA-Sponsored MCMIS Evaluation Reports prepared by UMTRI available at <http://www.umtri.umich.edu/divisionPage.php?pageID=308>, see evaluations for Florida, Mississippi, New York and New Mexico.

speeding violations nationwide.³² Commercial Carrier Journal (CCJ) documented these disparate state enforcement practices as well. CCJ found that although moving violations represented 8% of all violations cited against motor carriers nationally, they represented 29% of violations cited in Indiana and Delaware. By comparison, only 1.4% of violations cited in Mississippi were moving violations.³³ It is difficult to imagine that carriers in Indiana are that much more likely to commit moving violations, especially considering that moving violations only represented 7.5% of violations cited in neighboring Ohio. Such anomalies may help explain the presence of “exceptions,” as discussed earlier in this document. In other words, otherwise safe carriers may have high scores in some BASICS because they operate in states with comparatively more targeted enforcement practices.

Crash Accountability

A much criticized element of the CSA Safety Measurement System is how carriers’ scores in the *Crash Indicator* BASIC are assigned. Currently, carriers’ *Crash Indicator* BASIC scores are based on their frequency of involvement in crashes meeting certain thresholds, regardless of fault or preventability. In other words, the SMS considers all crashes equally, whether or not the truck driver caused the crash or could have done nothing to prevent it. A crash in which a motor carrier was rear-ended while stopped at an intersection bears the same weight as one in which its truck crashed into a parked car.

This flaw is significant since it paints all crash-involved carriers as being equally culpable. For instance, a carrier that *causes* three crashes is perceived as being just as unsafe as one *involved* in three crashes that it neither caused nor could have prevented. This is a source of frustration for motor carriers, particularly those involved in fatal crashes (which bear substantial weight in the SMS scoring), since research suggests that car drivers are principally at-fault in about three-quarters (70-75%) of fatal car-truck crashes.³⁴

FMCSA contends its approach to scoring carriers in the *Crash Indicator* is logical and appropriate because past crash involvement, regardless of fault, is a strong predictor of future crash involvement. An important distinction, however, is that FMCSA refers to crash “involvement” not “fault” in this context. Greater crash involvement (frequency) does not necessarily mean a carrier is more likely to have caused more crashes. Higher crash frequency often reflects the fact that a carrier operates in an urban environment, characterized by elevated exposure, and is more likely to be *involved* in crashes, but not necessarily more likely to *cause* them.

FMCSA’s current safety rating methodology acknowledges the role exposure plays in crash risk and applies a higher acceptable threshold for crash rates to carriers operating in urban environments. Specifically, FMCSA sets a higher threshold for acceptable crash rates for those carriers operating in urban environments. The language in the safety rating methodology reads as follows:

*Experience has shown that urban carriers, those motor carriers operating primarily within a radius of less than 100 air miles (normally in urban areas), have a higher exposure to accident situations because of their environment and normally have higher accident rates.*³⁵

³² *CSA Speeding Study*, Vigillo, Inc., June 2013, available at <http://www.trucking.org>

³³ *State Inspection Intensity*, Commercial Carrier Journal, March 2013.

³⁴ American Trucking Associations, *Relative Contribution/Fault in Car-Truck Crashes*, February 2013, page 5. Available at <http://www.trucking.org>

³⁵ See 49 C.F.R Part 385 *Appendix B- Explanation of Safety Rating Process, B. Accident Factor.*

For most carriers, FMCSA has established a threshold of 1.5 crashes per million miles as acceptable performance. Carriers with crash rates above that threshold are assigned a rating of “*Unsatisfactory*” in the accident factor of the safety rating methodology and, as a result, are unable to obtain an overall safety rating better than “*Conditional*.” However, for urban carriers the acceptable threshold (for measuring safe performance) is 1.7 crashes per million miles.

Violation Severity Weights

Perhaps the single largest factor affecting the correlation between carriers’ SMS scores and crash risk is how violations are weighted in the system. The SMS methodology assigns each violation a severity weight on a scale of 1 – 10 which is intended to reflect its relative crash risk. Of note, however, is that in this context “crash risk” is defined as the risk of crash involvement *and* of greater consequences resulting from a crash. As a result, some violations are assigned higher weights not because they make a crash more likely but because they may increase crash severity. For instance, a seat belt violation carries a weight of seven points not because failing to wear a seat belt is likely to cause a crash, but because doing so potentially increases crash consequences.

Another problem stems from how severity weights were originally assigned. Violations of similar types (e.g., lights, tires) were first placed into groups. Then, each violation in the group was assigned the same severity weight based on the presumed crash risk of the violation group. These weights were applied even if violations within a group had somewhat disparate relationships to crash risk. For instance, a “noncompliant fog lamp” was assigned the same weight as a stop lamp violation, which is considerably more likely to contribute to a crash. As a result, each individual violation was not assigned a weight that necessarily reflected its specific relationship to crash risk but rather the perceived risk of most violations within its group. Also, after weights were assigned to groups based on statistical crash risk, FMCSA modified the weights based on subjective input from “subject matter experts.” Doing so further blurred the statistical relationships between individual violations and crash risk.

A review of the various violations used to develop SMS scores reveals that some bear little or no apparent relationship to safety at all. Indeed, though the predecessor system to CSA (called SafeStat) focused almost exclusively on violations resulting in the declaration of out of service (OOS) orders, the SMS includes nearly all violations, even minor ones. By definition, violations resulting in out-of-service orders are those that “would likely cause an accident or a breakdown,”³⁶ which suggests that non-OOS violations, on their own, are not likely to cause a crash. For example, the SMS includes violations such as having an oil or grease leak and failing to carry spare fuses.

Scores Based on Comparative Performance

The use of SMS scores to draw conclusions about fleet safety performance raises questions about the meaning of these scores. A fleet’s safety posture is measured relative to the performance of other carriers with similar exposure. While scores reflect that other similarly situated fleets may have performed better or worse, the question is whether or not the fleet’s performance can be regarded as “safe” or “unsafe” as a result. To draw an analogy, the assignment of CSA scores is like grading on a curve. If most students taking a test scored 100% but several got the answers to just a few questions wrong, using the SMS methodology the latter group would get very poor scores. This would hold true even if, by conventional standards, they “passed” the test (e.g., got enough answers right).

³⁶ 49 C.F.R., Section 396.9 (c).

This process of measuring against carriers of similar exposure also affects the reliability and consistency of fleets' scores. In most measurement categories, the process involves comparing fleets to those with a similar number of relevant inspections by placing them into safety event groups (e.g., those with between 5 and 10 inspections in the prior 24 months, 11 – 20 inspections, and so on). However, a fleet's score can vary dramatically simply by moving from one safety event group to another because the point of reference changes. For instance, a fleet with 10 inspections and a relatively good score in a given BASIC can experience a dramatic change in its score upon receiving an 11th inspection and then being compared to a different safety event group (e.g., carriers with 11-20 inspections).

Conclusion

Researchers have arrived at mixed conclusions with respect to the reliability of SMS scores in identifying unsafe (crash prone) motor carriers. Some found virtually no correlation between scores and crash rates in any of the measurement categories. However, using a better predictive modeling tool, ATRI found a positive relationship between scores and crash risk in three of the publicly available measurement categories (BASICs) but found that scores in two others bear an inverse relationship to crash risk. Of the non-publicly available categories, scores in one (the *Crash Indicator* BASIC) likely correspond well to future crash involvement,³⁷ but scores in the other (the *HM Compliance* BASIC) do not.

ATRI also pointed out, however, that the number of alerts a carrier has been assigned is a strong indicator of crash risk. For instance, on average, a carrier with a single "Alert" in a BASIC will, on average, have a crash rate 1.24 times higher than a carrier with no alerts (but at least one BASIC score). Presumably, however, this varies depending on the BASIC in which the carrier has an alert, since scores in some BASICs are more strongly correlated with crash risk than those in others.

In sum, at least three of the system's seven measurement categories hold poor predictive value with respect to fleet safety. An important consideration, however, is that even in those BASICs that have a positive statistical relationship to crash risk generally, this correlation often does not hold true for individual carriers. In almost all measurement categories there are thousands of fleets with high scores but low crash rates or vice-versa.

The relationship between scores and crash risk is impacted by a number of data and methodology problems that plague the system. A substantial lack of data, particularly on small carriers who comprise the bulk of the industry, hinders the system's ability to render meaningful scores of comparative performance. Regional enforcement disparities likely cause fleets of all sizes operating in jurisdictions with robust enforcement practices to be perceived as less safe than those operating in other regions. Also, the questionable assignment of severity weights to individual violations can skew carriers' scores. Finally, the underreporting of crashes by states, the use of crashes that were not caused by motor carriers, and the increased exposure to crashes experienced by carriers operating in urban environments, all affect the significance of *Crash Indicator* BASIC scores.

³⁷ Note that crash involvement does not necessarily imply cause. See discussion of cause vs. involvement in the *Crash Accountability* section of this document.

Summary

Though there appears to be a statistical correlation between SMS scores in certain categories and crash risk, as well as the total number of alerts assigned and crash risk, the information can often be unreliable and inaccurate. While there is a general relationship between scores and crash risk in four measurement categories, in at least three of the seven measurement categories scores do not bear a positive statistical relationship to crash risk. Further, even in the categories that correspond to crash risk generally, the sheer number of “exceptions” (i.e. carriers with high scores but low crash rates and vice-versa) leads to the conclusion that SMS scores alone as a measure of an individual carrier’s safety performance are, at a minimum, unreliable. In all categories, data quality, data sufficiency and methodology problems hinder the system’s ability to produce dependable, consistent reflections of safety performance.

For more information contact Rob Abbott, Vice President of Safety Policy, at rabbott@trucking.org or P. Sean Garney, Manager of Safety Policy, at sgarney@trucking.org .