Trucking Fleet Concept of Operations for Automated Driving System-equipped Commercial Motor Vehicles

Chapter 5.6 ADS-equipped Truck Safety Metrics/Variables

Authors: Miller, A., Pugliese, B., Krum, A., Stojanovski, O.

Partners: Pronto.ai

VIRGINIA TECH TRANSPORTATION INSTITUTE

PRONTO

July 2024

Abstract

Automated Driving Systems (ADS) are set to revolutionize the transportation system. In this project, the research team led by the Virginia Tech Transportation Institute developed and documented a concept of operations (CONOPS) that informs the trucking industry, government agencies, and non-government associations on the benefits of ADS and the best practices for implementing this technology into fleet operations.

The sections of Chapter 5 provide guidance on a range of topics for fleets to consider and apply when preparing to deploy ADS-equipped CMVs in their fleet. The topics cover fleet-derived specifications, ADS installation and maintenance, ADS inspection procedures, driver-monitor alertness management, insuring ADS-equipped trucks, identification of ADS safety metrics/variables, ADS road assessment, and data security/transfer protocol and cybersecurity best practices.

Traditional safety metrics, including crashes and moving violations, may be insufficient for monitoring ADS-equipped trucks' performance or convincing the public of their safety. This effort describes potential variables and collection strategies for fleet decision-makers and the public to evaluate ADS safety toward the development of guidelines for ADS-equipped truck safety metrics. This effort also describes necessary data to assess and monitor ADS safety before and after deployment. Our findings categorize safety metrics into lagging and leading metrics. Lagging metrics, such as incidents per vehicle count and incidents per million miles, measure system safety after deployment but are poor for incident prevention. Leading metrics, like near-crash events and disengagements, are proactive indicators of future safety performance. We identified the application of both metrics to inform policymaking.

This report may be useful to fleets and ADS developers that support and operate ADS-equipped trucks to develop safety management plans prior to deployment. It may also be useful to public agencies and standards development groups that seek to identify and benchmark heavy vehicle safety performance criteria for human and ADS operations.

The following chapter has been extracted from the final report. For access to the full report, see this link: https://www.vtti.vt.edu/PDFs/conops/VTTI_ADS-Trucking_CONOPS_Final-Report.pdf

5. GUIDELINES

5.6 ADS-EQUIPPED TRUCK SAFETY METRICS/VARIABLES

5.6.1 Background

Surface transportation in the United States has become the primary means of transporting goods, with a heavy reliance on large trucks. Trucks affect every U.S. citizen regardless of personal mode of transportation, as nearly all consumer goods are delivered by trucks at some point in the delivery cycle.(1) There are approximately 3.5 million commercial truck driving licenses in active use,(2) and approximately 1.8 million of these licenses are used by drivers operating heavy and tractor-trailer trucks.(3) Trucks hauled 11.4 billion tons of freight in 2015, valued at more than $13 billion in 2012 dollars.(4) Following the 2008 recession, demand for freight services has steadily increased as the economy has grown, and truck drivers have needed to move more goods. As of 2015, there were 551,150 interstate motor carriers actively operating in the United States.(5) The trucking industry contributes significantly to the nation's economic portfolio, hauling 61% of the total freight transported in the United States by value in 2016(6) and contributing an estimated 3.5% of the nation's gross domestic product.(7)

Contrary to the transportation system's gradual evolution, vehicle technology is undergoing rapid changes that could affect all types of road transportation, and its effects on trucking could have a particularly important effect on society. Increasing demand for consumer goods and just-in-time inventory strategies (i.e., receiving goods only as needed) place a significant demand on truck drivers and the U.S. highway system as more and more goods are delivered by trucks. According to the Bureau of Labor Statistics, while the heavy and tractor-trailer truck driver workforce will only grow by slightly over 100,000 individuals from 2016 to 2026, with the level of expected retirements there will be openings for over 210,000 drivers per year over this period.(8) This may amount to over 100% turnover in some segments of the industry. In addition, the trucking industry has been aware of a truck driver shortage for some time,(9) and industry surveys of member firms show that turnover rates in an important industry segment (long distance truckload) have been persistently high for decades.(10)

Traffic congestion is one of the most critical challenges compromising the efficiency of the transportation system. The annual cost to the U.S. economy of travel delays caused by traffic congestion amounts to $160 billion, or $960 per commuter; each year, delays keep travelers stuck in their vehicles for 7 billion extra hours, corresponding to 42 hours per commuter, and waste 3 billion gallons of fuel.(11) In addition, traffic congestion leads to higher crash rates and negative environmental impacts resulting from increased CO_2 emissions and noise. These effects degrade the public's quality of life.

Beyond the costs associated with reduced efficiency and pollution, trucks represent a safety concern. Large truck and bus crashes place an estimated $112 billion burden on the U.S. economy, including costs related to lost productivity, property damage, medical treatment and rehabilitation, travel delays, legal services, emergency services, insurance, and costs to employers.(12) Although large trucks have lower rates of involvement in property-damage-only crashes and injury crashes compared to passenger cars, due to their size and weight, large truck crashes are more likely to result in severe consequences and costs. In fact, over two thirds of fatal truck crashes, which usually involve a passenger vehicle, result in the death of the other vehicle's driver. In 2014, there were 326,000 property-damage-only crashes, 3,424 fatal crashes, and 82,000 injury crashes involving large trucks.(13) Compared to the general U.S. working population, heavy truck drivers are 12 times more likely to die on the job(14) and 3 times more likely to suffer an injury involving time off work.(15)

It is for all these reasons (demand for goods, better safety, reduced congestion, environmental concerns, and lower driver costs) that OEMs and technology firms are pouring funds into the development of ADS. The introduction of ADS is expected to bring about a major change in the transportation system. By 2050, 80% of vehicles sold and contributing to miles traveled will likely be ADS-equipped.(16) This is expected to result in an estimated 21,700 lives saved and 4.2 million fewer crashes each year, as well as reduced traffic congestion, increased fuel efficiency, and increased productivity.(17) As a disruptive yet beneficial technology, ADS will also profoundly affect the U.S. economy.

Fleet personnel will need data on the safety of an ADS before implementing ADS-equipped vehicles into their operations. They will also need data to monitor how the ADS performs while deployed in their operations. The public will require data on the safety efficacy of ADS-equipped trucks to ensure they feel comfortable sharing the road with these vehicles. However, traditional safety metrics, such as crashes and moving violations, may be inadequate for monitoring the efficacy of ADS-equipped trucks once they are deployed or for convincing the public of the safety of these technologies. New safety metrics must be explored and must provide objective measures of ADS safety, but these metrics should also be informative to the end users (i.e., fleet decision-makers and the driving public). In a similar vein, an operational definition of truck ADS safety must also be developed (i.e., what is the minimum level of ADS-equipped truck safety required to deploy these vehicles and to maintain deployment?).

5.6.2 Objective

What are appropriate safety metrics and variables for ADS-equipped trucks? The goal of this section is to outline potential variables that might be used by fleet decision-makers and the public to evaluate the safety of the ADS. Existing metrics (e.g., miles driven, disengagements, crashes) used by ADS technology vendors may fall short of the industry's stated safety aspirations. For example, it is common to report the total miles driven to tout technological progress and imply greater safety. However, progress in ADS development does not equate to overall safety; thus, these two criteria should be viewed separately. If total miles driven are allowed to stand as a proxy for safety, ADS technology developers might be incentivized to put the public at additional risk by driving more than necessary (e.g., it appears Uber ATG was driving an unjustifiable number of miles to accumulate more "total miles driven," which likely contributed to the fatal crash in Arizona in 2018).(18)

Similarly, ADS technology developers have historically been good at avoiding at-fault crashes but do a poor job of avoiding preventable crashes. The ADS-equipped shuttle crash in Las Vegas in 2017 is a perfect example. A large truck backed into the ADS-equipped shuttle, which resulted in a crash where the truck driver was legally at fault. This crash was in fact preventable, but the ADS did not make any effort to avoid the crash in the manner that a human driver would have (i.e., by backing out of the way).[19] To date, there has not been a systematic evaluation of the safety practices and metrics used by the industry.

5.6.3 Criteria for Safety Metrics

The VTTI team systematically identified and quantified the shortcomings and misaligned incentives of traditional ADS safety metrics. Then, we investigated alternative metrics that may be better indicators of ADS safety and that better align with incentives to develop and deploy ADSs in a prudent manner. (See references [20], [21], [22], [23], [24], and [25].) As a final step, the VTTI team consulted FMCSA, other policymakers, safety advocates, and ADS developers on which variables will be used to evaluate the safety of the ADS to get their feedback on the suggested safety metrics. There will likely not be any single metric and resulting "magic number" that indicates an ADS is safe to deploy. Also, the safety evaluation of an ADS should not be considered a one-time event (e.g., certification the ADS is safe), but should rather be a continual process given there will be new software upgrades, vehicle platforms, ODDs, etc.

5.6.4 Characteristics of Safety Metrics

An ideal metric to track ADS vehicle safety must be valid, reliable, feasible, non-manipulatable, and informative to the end user.[26] For a metric to be valid, it must directly measure the characteristic being tested (as opposed to a proxy variable). Reliability here means the safety metric is well-defined and consistent. Feasible means the metric can be easily tracked, considering time and resources. Non-manipulatable means it is not possible to "game" the data. End-user comprehension means the safety metric provides useful or interesting information to end users (i.e., fleet decision-makers and the public); thus, the safety metrics must be something end users can easily interpret and understand (otherwise, additional education and training are necessary).

Measurement Stage: The goal of the CONOPS project is to develop a living, comprehensive document that describes the ADS characteristics from the viewpoint of the truck fleets that will use ADS technology. This CONOPS will provide the trucking industry with clear guidelines on how to safely implement, and benefit from, ADS-equipped trucks. Thus, the safety metrics will focus on ADS truck deployment in fleets rather than development (establish and improve the ADS) or demonstration (exhibit ADS functionality).[27,28] Given the focus is on deployment, the safety metrics will focus on the ADS rather than any ADS subsystems or specific components or subcomponents.

ODD: Safety metrics, where possible, should be stratified by ODD.[29,30,31] The ODD specifies where the ADS can operate. Stratifying safety metrics by ODD will inform where the ADS performs better or worse, which is useful for driver training and education (if a driver is present) and route planning. These metrics will also be useful for refining the ADS through continued development and demonstration. The ODD can be defined by many different factors, including road geometry (straight vs. curved, incline vs. level, etc.), weather, time of day, road lighting, road surface (wet vs. dry, dry vs. snow, etc.), level of service (i.e., traffic density), road classification (see MAF/TIGER Feature Class Code definitions),[32] etc.

Tactical Maneuvers: Further stratification of safety metrics may occur through identification of performed tactical maneuvers. These tactical maneuvers are presented as control-related tasks of the ADS-equipped vehicle,[33] and the safety metrics would demonstrate outcomes across instances of tactical maneuvers. Metrics calculated within tactical maneuvers reflect a more precise means to define behavioral outcomes that serve as a comparison to other ADS-equipped vehicles, human performance, simulated models, or predetermined safety performance thresholds. For example, lane changes performed by the ADS-equipped vehicles can be parsed from the data, and safety metrics can be calculated across lane changes to determine the safety performance of the vehicle. Though the metrics calculated within tactical maneuvers are expected to be very informative as a comparison and a benchmark against safety thresholds, the ability of a fleet to define, parse, and stratify data pertaining to maneuvers remains burdensome. Future efforts investigating means for extracting tactical maneuvers may alleviate the burden on fleets by constraining or limiting needed parameters for calculations.

5.6.4.1 Summary of Exposure

This section provides an overview of the potential segmentation of safety metrics of a single or multiple ADS-equipped vehicle. Table 37 provides a list of the exposure characteristics (feasible and specificity) and the rating levels (high, medium, low) for each characteristic.

Table 1. Exposure characteristics and rating levels.

Characteristic	High	Medium	Low
Feasible	Collected with limited effort (e.g., time, cost, resources).	Collected with moderate difficulty.	Exposure is difficult to capture.
Specificity	Exposure provides insight into metrics at a granular level.	Exposure provides some insight into metrics.	Exposure produces high-level or overview metrics.

Table 38 provides a list of the potential exposures available used to segment safety metrics and the rating level for each characteristic, along with examples. As indicated in Table 40, fleet decision-makers are familiar with lagging safety metrics, as these metrics are currently used to evaluate their fleets' safety performance.

Table 2. Potential ADS exposure calculations and rating characteristics.

Exposure Type	Feasibility	Specificity	Examples
Organization	High	Low	Entire organization, miles driven, hours driven
Site	High	Low	Site location, yard
Vehicle Type	High	Low	OEM-specific
ADS Version	High	Low	AV release version
Operation Type	High	Low	Hub-to-hub, port drayage
Trips	High	Medium	Specific trips
ODD: routes	High	Medium	Interstate, exit-to-exit
ODD: conditions	Medium	Medium	Weather, work zones, time of day
ADS Mode	Medium	Medium	ADS engaged or disengaged
Events	Medium	High	Crashes, near-crashes
Tactical Maneuvers	Low	High	Lane change, backing, vehicle cut-in

5.6.4.2 Operational Definition of Safety

What is an acceptable level of safety in an ADS (how safe is safe)? The CONOPS project does not answer this question. However, the authors would like to acknowledge its importance with respect to safety metrics, which can be measured against this level and/or the relative difference. Fraade-Blanar et al.(34) provides a good summary on this topic—surprisingly, there is no agreed upon operational definition of safety. Some have argued that ADS should be compared against the behavior of human drivers.(35,36,37) However, this method also poses challenges in terms of what the appropriate human comparison would be (truck driver, teen driver, drunk driver, crash-free driver, attentive driver). Even if an ADS were as safe as the average driver, 50% of the driving population would be riding in a vehicle that was less safe compared to their own driving.

Although achieving zero crashes is the vision in ADS implementation, it is likely that crashes will still occur. An appropriate human reference is an important benchmark for evaluating ADS. Krum et al.(38) provided baseline driving performance from 3.12 million miles of naturalistic truck driving data. These data were stratified by ODD and six maneuvers—speed behavior, longitudinal deceleration, following distance, lateral acceleration, lane deviation, and lane stability—which provide a human reference of driving performance in a particular ODD. Also included in that study is a public-use data tool for querying event rates based on a range of selectable parameters. These data provide baseline safety performance measures from human-operated trucks.

5.6.5 Safety Metrics

The safety metrics noted below are grouped as lagging or leading metrics with respect to ADS safety. Lagging safety metrics measure ADS "incidents" in the form of prior safety statistics. As they are lagging indicators, they are a poor measure for preventing safety incidents. These are the most commonly used safety metrics, including incidents per vehicle count, incidents per million miles, incidents per division or business unit, year-over-year number of vehicle crashes, and on-road injuries per 200,000 hours worked (aligned to Occupational Safety and Health Administration reporting). A leading safety metric precedes or indicates a future event and measures activities carried out to prevent and control safety incidents. These metrics are proactive and provide information on how the ADS is performing on a regular basis.(39) As indicated above, the safety metrics (regardless of lagging or leading) should be calculated for each specific ODD, as should accounting for exposure (using a denominator to obtain a rate, such as vehicle miles traveled, driving hours, per ADS-equipped truck, trips, events, etc.). Of these measures of exposure, the gold standard has been to calculate using vehicle miles traveled or driving hours. However, recent efforts have discussed evaluating safety metrics within incidents of events or scenarios, describing vehicle behaviors at a refined level. Applying that concept of leading indicators, an example safety metric would be the distance to all other vehicles when the ADS-equipped vehicle is performing a turn, or the speed and headway adjustment after the ADS-equipped vehicle experiences a cut-in by another vehicle.

5.6.5.1 Lagging Metrics

In this section, we provide a high-level overview of the suggested lagging indicators. Table 39 provides a list of the safety metric characteristics (valid, reliable, feasible, non-manipulatable, and informative) and the rating levels (high, medium, low) for each characteristic. As indicated in Table 40, fleet decision-makers are familiar with these metrics, as they are currently used to evaluate their fleets' safety performance with their current (human-driven) power units. These data is available for fleet decision-makers with little input from the ADS developer.

Table 3. Safety metric characteristics and rating level (adapted from Fraade-Blanar et al.[40]).

Characteristic	High	Medium	Low
Valid	Directly measures ADS safety.	Somewhat measures ADS safety.	Indirectly measures ADS safety.
Reliable	Safety metric is well defined and quantitative.	Safety metric is somewhat defined and quantitative.	Safety metric is qualitative, subjective, anecdotal.
Feasible	Collected with limited effort (e.g., time, cost, resources).	Collected with moderate difficulty.	Safety metric is difficult to capture.
Non-manipulatable	Limited opportunity to manipulate this safety metric.	Moderate opportunity to manipulate this safety metric.	Easy to manipulate.
End User Comprehension	Well understood by the end user, use is common practice.	Used by some end users.	Not currently used by end users.

Table 40 provides a list of the potential lagging safety metrics (described below) and the rating level (high, medium, low) for each characteristic (valid, reliable, feasible, non-manipulatable, and informative). Fleet decision-makers are familiar with lagging safety metrics, as they are currently used to evaluate their fleets' safety performance.

Table 4. Potential ADS lagging safety metrics and rating characteristics.

Safety Metric	Temporal	Valid	Reliable	Feasible	Non-Manipulatable	End-User Comprehension
Crash	Lagging	Medium	Medium	High	High	High
FMCSA-reportable	Lagging	Medium	High	Medium	High	High
Preventable Crash	Lagging	High	High	Medium	High	High
Non-preventable Crash	Lagging	Low	High	Medium	High	High
Injury Crash	Lagging	Medium	Medium	Low	High	High
Fatal Crash	Lagging	High	High	Medium	High	High
Tow-away Crash	Lagging	Medium	Medium	High	High	High

Miles Driven: Miles driven refers to the total miles driven under control of the ADS. These can be subdivided by specific ODD. Although miles driven is an important measure of exposure, which should be included as a denominator in the safety metrics, it does little to reflect the ADS's safety.[41]

Crashes: Crashes are the most widely used safety metric. They are defined as the ego vehicle contacting another vehicle, pedestrian, animal, road debris, other stationary object, or a road departure. Crashes can be further divided based on their severity (see KABCO Injury Classification)[42] and/or cost. Below are the most commonly used crash metrics.

FMCSA-reportable Crashes: FMCSA-reportable crashes must be reported to FMCSA. These crashes involve a fatality, injury that requires immediate medical treatment away from the crash scene, or a vehicle that is disabled as a result of the crash and must be transported away by a tow truck or other vehicle.[43]

Preventable Crashes: Each fleet has their own operational definition of a preventable crash. The National Safety Council defines a preventable crash as one in which the driver failed to exercise every reasonable precaution to prevent the accident. This is irrespective of whether there was property damage or personal injury, the extent of the loss or injury, to whom it occurred, and the location of the crash.[44] An example of this type of crash would be a vehicle hitting another vehicle that was stopped at an intersection facing the direction of travel.

Non-Preventable Crashes: By definition, a non-preventable crash is any crash that was not determined to be a preventable crash. Non-preventable means any crash in which everything that could have been reasonably done to prevent it was done and the crash still occurred. For example, a vehicle stopped at an intersection facing the direction of travel is struck from behind by another vehicle.

Fatality: A crash that results in one or more fatalities. Death is recorded within a period after the crash (e.g., 30 days).

Injury: A crash that results in one or more injuries. These can be nested based on the severity of the injuries (e.g., incapacitating injury, non-incapacitating injury, possible injury). Injury is recorded within a period after the crash (e.g., 30 days).

Tow-away Crash: A crash that results in a vehicle that is disabled and must be transported away by a tow truck or other vehicle.

5.6.5.2 Leading Metrics

This section provides a high-level overview of the suggested leading indicators. As indicated in Table 41, most fleet decision-makers are unfamiliar with these metrics; thus, training and education are needed to increase awareness. Most of the leading safety metrics described below are not readily available to fleet decision-makers and require input from ADS developers.

Table 41 provides a list of the potential leading safety metrics (described below) and the rating level (high, medium, low) for each characteristic (valid, reliable, feasible, non-manipulatable, and informative).

Table 5. Potential ADS leading safety metrics and rating characteristics.

Safety Metric	Temporal	Valid	Reliable	Feasible	Non-Manipulatable	End-User Comprehension
Near-crash	Leading	Medium	Medium	High	High	Medium
Traffic Violation	Leading	Low	Low	Medium	Low	Medium
Disengagement	Leading	Low	Medium	High	Low	Low
Simulated Manual Disengagement	Leading	Medium	High	Medium	Low	Low

Safety Metric	Temporal	Valid	Reliable	Feasible	Non-Manipulatable	End-User Comprehension
Conventional Indicators	Leading	High	High	High	Medium	Medium
Perception-based Indicators	Leading	High	High	Low	Medium	High
Safety Envelop Violation	Leading	High	Medium	Low	Medium	Low
Fleet Integration	Leading	Medium	Medium	Medium	Low	Medium
Confidence and Accuracy	Leading	Medium	Medium	Medium	Medium	Low

Near-crashes: Until recently, lagging metrics were the only widely available metric for fleet decision-makers. Although the use of near-crashes is a relatively new safety metric in trucking, near miss reporting has been used successfully in the aviation industry for many decades. Near-crashes are non-crash events (a subjective judgement on the potential for a crash); however, there is no standardized operational definition for these events. Hankey et al.(45) defined a near-crash as "any circumstance that requires a rapid evasive maneuver by the subject vehicle, or any other vehicle, pedestrian, cyclist, or animal, to avoid a crash is considered a near-crash. A rapid evasive maneuver is defined as steering, braking, accelerating, or any combination of control inputs."

Traffic Violations: A traffic violation is a State or Federal (in the case of FMCSA) law that regulates the operation of trucks on streets and highways. These laws vary by State. Traffic violations can be moving (i.e., vehicle is in motion) or non-moving (i.e., vehicle is not in motion). Moving violations include speeding, failure to yield, turning into the wrong lane, etc., whereas non-moving violations are usually reflective of parking violations (e.g., parking in front of a fire hydrant, parking in a no-parking zone). Most of the latter are unrelated to safety per se.(46)

Disengagements: A disengagement is when the ADS-equipped vehicle is in automated mode and control of the vehicle is returned to the human driver. There are two types of disengagements: (1) automatic and (2) manual. An automatic disengagement is when an ADS-equipped vehicle exits the automation mode through an error or kickout, or, if able, when the system requests a human driver to take over the dynamic driving task (as depicted in SAE Level 3 automation). A manual disengagement is when the human driver is not confident with the ADS (e.g., discomfort, adverse weather conditions, heavy traffic, poor infrastructure, potential adverse situation) and takes control of the vehicle from the ADS.(47,48) The relationship between disengagements and safety is unclear, as fewer disengagements may not necessarily reflect better safety.(49) One potential option to increase the validity of manual disengagements is to simulate the ADS's behavior (and the behavior of other actors) had it not been disengaged by the driver. This could be a useful solution to determine if the disengagement was warranted.(50) Thus, manual disengagements could be subdivided into those where the ADS would have functioned safely and those where the ADS was unsafe. A subset of manual disengagements includes both disengaging the system for test-related or normal operations, such as exiting automation to take an exit to refuel or leaving the parameters of the ODD testing area, and accidental disengagements from the safety operator. These disengagements would be irrelevant to the functional safety of the vehicle and would not be counted as part of a safety metric.

Conventional Indicators: Traditional metrics used to demonstrate the capabilities of a manned vehicle have been represented across numerous studies, both experimental and naturalistic. These metrics typically relate to specific outcomes of the vehicle and are reflective of immediately comprehensible vehicle parameters. The calculation of these indicators will typically rely on sensors (non-visual) equipped on the vehicle as well as kinematic data and other information coming from the CAN bus's J1939 protocol. Metrics can be summarized through typical statistical methods, including creating averages, ranges, minimums and maximums, or standard deviations of data across some exposure level. Examples include yaw rates, acceleration, and speed, each of which can be calculated across trips, ODDs, tactical maneuvers, or another meaningful stratification method.

Perception-based Indicators: Similar safety metrics can be calculated using processed visual sensors in conjunction with conventional indicators to determine safe operation of the ADS-equipped vehicle in relation to roadway elements and other traffic actors. The inclusion of perception sensors allows for a more real-world understanding of the vehicle's position in relation to all elements, static or dynamic, on the roadway. Further, perception-based safety metrics include the placement of the ADS-equipped vehicle in lane and the relative proximity and velocity of other road users. Examples of metrics include lane tracking and lane centering, car following, and distance to other vehicles or objects. These metrics can also be stratified within vehicle or system, or across trips, ODDs, or tactical maneuvers. One further example is to parse out tactical maneuvers in which the ADS-equipped vehicle merges or changes lanes in front of another vehicle and calculate the average minimum distance to that following vehicle across every instance of the maneuver. This average minimum would serve as an easily understood metric of safety that can be compared to other ADS technologies or against a human baseline of performance. These metrics are typically presented as lower-order metrics that combine to create higher-order metrics representing the safety envelope of the vehicle.

Safety-Envelope (Risk-based) Violations: Fraade-Blanar et al.[51] termed safety-envelope violations "roadmanship" (i.e., the ability of the ADS to drive safely without creating hazards and/or responding to other hazards). These violations can be counted and defined by an initiator and a responder so the violation can be attributed to the ego vehicle or the other road user. These violations are likely to vary by ODD and ADS developer unless there are standards or regulations. These safety envelopes could be defined based on the safe lateral and longitudinal distance to another vehicle, defensive driving, quickness to give right-of-way, and infrastructure limitations. Fraade-Blanar et al.[52] envision a series of boundaries, each with a more extreme evasive response from the ADS. See the Underwriters Laboratories[53] standard, UL 4600, for specific safety-envelope violations. Practical applications of risk-based metrics include Responsible Sensitive Safety and NHTSA's Model Predictive Instantaneous Safety Metric. These metrics attempt to define the safety status of the ADS-equipped vehicle.

Fleet Integration: The introduction of ADS-equipped vehicles requires continuous evaluation of metrics related to the efficient implementation of the technology into the existing organizational structures. As ADS technologies are first introduced, a close relationship between the ADS developer and the incorporating fleet is required for mixed-fleet operations. This relationship should produce an implementation plan that will safely incorporate ADS technologies into the existing system. While integration metrics may include operational (e.g., number of trucks involved, tasks assigned) or monetary (e.g., efficiency) components, the metrics related to safe implementation are critical. These metrics may include lagging (e.g., crashes) or leading (e.g., conventional metrics) indicators as described above, but could also include non-traditional metrics relating to the training of personnel (e.g., safety operators, support team, maintenance), implementation of operational policies (e.g., coaching, culture) and protocols (e.g., communications, interactions), and tracking of individual behaviors surrounding ADS operations.

Confidence and Accuracy: Across each decision made during the motion or path planning of the ADS-equipped vehicle, the system is expected to produce an internal go/no-go for each choice the ADS makes. These decisions are dependent on many parameters, and the integration of each relevant factor will ultimately dictate the behavior of the vehicle. A potential option to evaluate vehicle behaviors during edge case events is to insert the ADS into a simulated situation and record which behaviors the vehicle is most likely to execute, along with confidence or similar outputs that dictate the choice of behavior selected by the ADS. Other opportunities for evaluation using naturalistic data may provide similar insight into the likely behavioral competencies of the ADS during scenarios that are not often encountered on the roadway.

---

[1] Federal Highway Administration. (2020). *2016 Freight Quick Facts Report*. https://ops.fhwa.dot.gov/publications/fhwahop16083/ch1.htm

[2] Federal Motor Carrier Safety Administration. (2017). *2017 Pocket Guide to Large Truck and Bus Statistics*. U.S. Department of Transportation.

[3] Bureau of Labor Statistics. (2018). Heavy and Tractor-Trailer Truck Drivers (SOC code 53-3032). *Occupational Outlook Handbook*. U.S. Department of Labor. http://www.bls.gov/oes/current/oes533032.htm

[4] Bureau of Labor Statistics. (2018).

[5] Federal Motor Carrier Safety Administration. (2016). *2016 Pocket Guide to Large Truck and Bus Statistics*. U.S. Department of Transportation. https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/2016_Pocket_Guide_to_Large_Truck_and_Bus_Statistics.pdf

[6] Bureau of Labor Statistics. (2018).

[7] Bureau of Transportation Statistics. (2017b). *Freight Facts and Figures 2017*. U.S. Department of Transportation.

[8] Bureau of Labor Statistics. (2017a). Table 1.8 2016–26 Industry-occupation matrix data, by occupation. U.S. Bureau of Labor. *Occupational Employment Projections*. https://www.bls.gov/emp/tables/industry-occupation-matrix-occupation.htm

[9] Costello, B. (2017). *Truck Driver Shortage Analysis 2017*. American Trucking Associations.

[10] ATA Economics Department. (2017). *Driver Turnover Historic Database TL*. American Trucking Associations.

[11] Schrank, D., Eisele, B., Lomax, T., & Bak, J. (2016). 2015 Urban Mobility Scorecard. https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-scorecard-2015.pdf

[12] Blincoe, L., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Luchter, S., & Spicer, R. (2002). *The Economic Impact of Motor Vehicle Crashes, 2000* (Rep. DOT HS 809 446). National Highway Traffic Safety Administration.

[13] Federal Motor Carrier Safety Administration. (2017).

[14] Chen, G.X., Amandus, H.E., & Wu, N. (2014). Occupational fatalities of truck driver and driver/sales workers in the United States, 2003–2008. *American Journal of Industrial Medicine, 57*, 800–809.

[15] Bureau of Labor Statistics. (2014). *Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work, 2013.* U.S Department of Labor. http://www.bls.gov/news.release/pdf/osh2.pdf

[16] Lavasani, M., Jin, X., & Du, Y. (2016). Market penetration model for autonomous vehicles on the basis of earlier technology adoption experience. *Transportation Research Record: Journal of the Transportation Research Board, 2597*, 67-74.

[17] Flagnant, D.J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A, 77*, 167-181.

[18] National Transportation Safety Board. (2019). *Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian: Tempe, Arizona March 18, 2018* (NTSB/HAR-19/03, PB2019-101402). National Transportation Safety Board.

[19] National Transportation Safety Board. (2019). *Low-Speed Collision Between Truck-Tractor and Autonomous Shuttle: Las Vegas, Nevada, November 8, 2017* (HWY18FH001). National Transportation Safety Board.

[20] Underwriters Laboratories. (2020). *Standard for Evaluation for Autonomous Products: UL 4600* (1st ed.).

[21] Greer, C., Griffor, E., & Wollman, D. (2019). *Workshop Report: Consensus Safety Measurement Methodologies for Automated Driving System-Equipped Vehicles* (NIST Special Publication 1900-320). National Institute of Standards and Technology.

[22] Fraade-Blanar, L., Blumenthal, M.S., Anderson, J.M., & Kalra, N. (2018). *Measuring Automated Vehicle Safety*. RAND Corporation.

[23] Webb, N., Smith, D., Ludwick, C., Victor, T.W., Hommes, Q., Favarò F., Ivanov, G., & Daniel, T. (2020). *Waymo's Safety Methodologies and Safety Readiness Determinations*. www.waymo.com/safety

[24] Schwall, M., Daniel, T., Victor, T., Favaro, F., & Hohnhold, H. (2020). *Waymo Public Road Safety Performance Data*. www.waymo.com/safety

[25] Aptiv. (2019). *Safety First for Automated Driving*. https://www.daimler.com/documents/innovation/other/safety-first-for-automated-driving.pdf

[26] Fraade-Blanar et al. (2018).

[27] Greer et al. (2019).

[28] Fraade-Blanar et al. (2018).

[29] Aptive. (2019).

[30] Greer et al. (2019).

[31] Krum, A., Miller, A., Sarkar, A., Engstrom, J., Soccolich, S., Grove, K., Hickman, J. Hanowski, R., & Ali, GF. (2021). *Naturalistic Driving Data Baseline for Highly Automated Commercial Motor Vehicle Applications*. Federal Motor Carrier Safety Administration.

[32] U.S. Census Bureau. (2009). MAF/TIGER Feature Class Code (MTFCC) Definitions. https://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2009/TGRSHP09AF.pdf

[33] Thorn, E., Kimmel, S., & Chaka, M. (2018). *A Framework for Automated Driving System Testable Cases and Scenarios*. National Highway Traffic Safety Administration.

[34] Fraade-Blanar et al. (2018).

[35] Krum et al. (under Agency review).

[36] Greer et al. (2019).

[37] Krum et al. (under Agency review).

[38] Krum et al. (under Agency review).

[39] Geller, E. S. 2001. *The Psychology of Safety Handbook.* CRC Press.

[40] Fraade-Blanar et al. (2018).

[41] Kalra, N., & Paddock, S.M. (2016). *How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?* RAND Corporation.

[42] National Highway Traffic Safety Administration. (2012). *Model Minimum Uniform Crash Criteria* (4th ed. DOT-HS-811-631). U.S. Department of Transportation.

[43] Federal Motor Carrier Safety Administration. (n.d.). *Truck and Bus Crashes Reportable to FMCSA*. https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/Truck_and_Bus_Crashes_Reportable_to_FMCSA.pdf

[44] National Safety Council. (n.d.). *A Guide to Determine Motor Vehicle Collision Preventability*.

[45] Hankey, J.M., McClafferty, J.A., & Perez, M.A. (2016). *Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Events*. Transportation Research Board of the National Academies

[46] Fraade-Blanar et al. (2018).

[47] Dixit, V.V., Chand, S., & Nair, D.J. (2016). Autonomous vehicles: Disengagements, accidents and reaction times. *PLoS One, 11*(12).

[48] Favarò, F.M., Nader, N., Eurich, S.O., Tripp, M., & Varadaraju, N. (2017). Examining accident reports involving autonomous vehicles in California. *PLoS One, 12*(9).

[49] Fraade-Blanar et al. (2018).

[50] Schwall et al. (2020).

[51] Fraade-Blanar et al. (2018).

[52] Fraade-Blanar et al. (2018).

[53] Underwriters Laboratories. (2020).