Automated Vehicle Safety Consortium A Program of SAE ITC

Automated Vehicle Safety Consortium™ Best Practice

AVSC00006202103

Issued 2021-03
Superseding

AVSC Best Practice for Metrics and Methods for Assessing Safety Performance of Automated Driving Systems (ADS)

Citation: Automated Vehicle Safety Consortium. 2021. Best Practice for Metrics and Methods for Assessing Safety Performance of Automated Driving Systems (ADS). SAE Industry Technologies Consortium.

Rationale

Credible, practicable, and consistent ways to measure automated driving system (ADS) safety performance are essential to garner public trust and confidence. Safety metrics that are widely used across the automated driving systems (ADS) industry are needed to improve communication in open forums, enhance coordination with public agencies, improve public confidence in AVs, and accelerate realization of the potential safety and other benefits of automated driving system (ADS) deployments. Several approaches for assessing ADS safety performance have been put forward, but not broadly adopted.

Measuring safety outcomes in any field is a complex task that requires considerable time and exposure to achieve statistical significance. This complexity is especially true as it relates to measuring safety outcomes with respect to ADS, given the relatively low prevalence of ADS deployments in the real world. Because of this, predictive indicators of ADS safety performance are desired. Prior to and during commercial deployment, ADS operators will, and should, use metrics to track safety performance and inform risk management. Predictive metrics complement safety outcome metrics, and both serve critical roles in ADS safety performance assessment and monitoring. This effort identifies a foundational set of common, system-level metrics that can be used as part of an ADS developer's aggregate safety performance assessment. It is assumed that additional metrics will be applied in combination with these common metrics to supplement evidence of safety performance at different stages of product development and deployment.

Preface

The Automated Vehicle Safety Consortium™ (AVSC) is an industry program of SAE Industry Technologies Consortia® (SAE ITC) publishing best practices that inform and lead to industry-wide standards and advance the safe deployment of automated driving systems (ADSs). The members of this Consortium have decades of accumulated experience focused on safe, reliable, and high-quality transportation. They are committed to applying these principles to SAE level 4 and level 5 automated vehicles so that communities, government entities, and the public can be confident that these vehicles will be deployed safely.

The Consortium recognizes the need to establish best practices for the safe operation of ADS-dedicated vehicles (ADS-DVs). These technology-neutral practices provide key considerations for safely deploying ADS-DVs on public roads. Members of the AVSC intend to support the published principles and best practices to set a bar for other industry participants to meet. These best practices will enhance and expedite the development of formal industry standards at SAE International and other global standards development bodies. Effectively implementing these principles can help inform the development of sound and effective ADS regulations and safety assurance testing protocols that, in turn, will engender public confidence in the efficacy of ADS-DVs.

Comment and open discussion on the topics are welcome in appropriate industry forums. As discussion unfolds, AVSC documents will be revised as significant information and/or new approaches come to light that would increase public trust.

Introduction

ADS safety assurance is a broad, complex topic. Industry standards can support a positive risk balance when applied to the introduction of new technologies like ADS by proactively addressing aspects of safety assurance and contributing to a manufacturer's documented evidence for safe performance. Traffic safety is most directly measured in terms of the rate and severity of crashes; however, crash metrics are insufficient to manage safety risk of dynamic driving task (DDT) performance because they require crashes to occur before safety insights are gained [1]. Proactive ADS-relevant safety metrics capable of detecting potential safety issues ahead of crashes are essential. To date, efforts on this topic vary greatly in levels of maturity, specificity, and practicality. Appendix C lists some of the safety evaluation metrics and concepts considered by AVSC's team of ADS safety and development experts in the development of this best practice, including development of the metric selection criteria (4.0).

Metrics support ADS developer testing processes (e.g., Safety First for Automated Driving [2]) including scenario-based testing (e.g., NHTSA's testable cases framework [3], PEGASUS [4], SAKURA [5]) and evidence to support a company's case for the safety performance of an ADS in a given ODD (e.g. UL 4600 [6]). Consistency in the type of safety performance metrics used across the industry, across operational design domains (ODDs), and across the product development lifecycle facilitates consistent evaluation. Each developer and manufacturer will have customized metrics to test their system and provide evidence of safe performance. Combining those customized metrics with this common foundational set of metrics for monitoring safety performance can enhance communication among stakeholders and thus accelerate industry-wide continuous improvement in safety performance.

Common, foundational safety metrics that span ODD and manufacturer encompass concepts related to safety envelopes, abrupt maneuvers, and response to objects and events [7]. Safety envelope concepts generally use kinematic equations to describe motion of ADS and other objects in the environment (e.g., Responsibility Sensitive Safety (RSS) [8], Safety Force Field (SFF) [9], Criticality Metric [90], and instantaneous safety metric (ISM) [10]). Many of these metrics depend on models and assumptions about the environment and behavior of other road users that are likely to vary by location and environment. The fidelity of models – and therefore predictive power of these metrics – can also vary (i.e., whether vehicle dynamics are considered will impact accuracy). The manner in which a safety envelope is determined and acceptability of ADS responses to objects in a given environment or scenario requires flexibility based on system architecture, operational risk management protocols, and ODD-relevant assumptions.

Metrics based on data that may already be collected lend themselves to broad industry acceptance. Vehicle data already collected based on standards such as SAE J1698 and SAE J3197 have been used in event and vehicle analyses for years. The usefulness of data analysis is well established. The *AVSC Best Practice for Data Collection for Automated Driving System-Dedicated Vehicles (ADS-DVs) to Support Event Analysis* also provides a practicable starting point for data that can support ADS safety performance metrics. Infrastructure-based data may complement vehicle-based data but presents challenges to scalability and could exclude certain geographies from being able to deploy ADS-enabled vehicles if that type of data is required. At least in the near-term, the most widely implementable and reliable metrics related to ADS safety performance will come from the ADS-DV itself.

This document lays out a performance-based, technology-neutral approach for measuring and analyzing safety performance that supports long-term, socially important safety goals. ADS safety performance metrics in this document enable system-level analysis that is practical to implement for any system and does not assume a particular system architecture, nor require any subsystem-level information. ***This AVSC best practice recommends a set of safety performance metrics to be measured and analyzed and supports emerging ADS standards.***

Table of Contents

1. Scope

This *AVSC Best Practice for Metrics and Methods for Assessing Safety Performance of Automated Driving Systems (ADS)* (AVSC00006202103) recommends a set of metrics that may be used to assess ADS safety performance of the dynamic driving task (DDT)[1]. The metrics and methods described in this document are principally designed to provide evidence of safety performance for deployment of fleet-operated/managed SAE level 4 and 5 ADS-dedicated vehicles (ride-hailing or product delivery). While these metrics may also support other SAE levels of automation of ADS-equipped vehicle deployments, such as level 4 sub-trip features equipped on conventional vehicles, or dual-mode vehicles sold to private users, they are not the focus of this best practice. The recommended metrics in this document are intended for public road application; however, ADS manufacturers and developers are not precluded from adapting these metrics if appropriately analyzed for their specific use case (e.g., operating on sidewalks, warehouses, etc.). Non-safety-related performance characteristics, such as drive quality or energy efficiency, are outside the scope of this document. ODD maintenance and issues associated with an ADS exiting an ODD are also outside the scope of this document.

[1] For ADS-DVs, the ADS should perform all of the DDT on a sustained basis in an ODD (if applicable), including the DDT fallback (SAE J3016).

1.1 Purpose

The document describes how ADS-derived metrics provide evidence of safety performance for ADS-dedicated vehicles. It also describes how in the near- to mid-term metrics can serve as predictive indicators for safety outcomes and in the long-term directly support measures of safety outcomes.

The metrics and methods provided in this document are intended for use by the technical community (developers, manufacturers, testers, etc.) to aid in the development and deployment of safe ADS. This best practice may be used by ADS manufacturers and developers to evaluate and document aggregate safety performance of vehicles within the target ODD prior to and during deployment in a way that builds public understanding and acceptance. This effort identifies a foundational set of common, system-level metrics. It is expected to be considered as part of a developer's broader safety performance assessment throughout the development process.

The metrics and methods provided in this document may also be useful to public agencies and stakeholders who have interest in better understanding the safety posture of ADS deployments. To further public acceptance of automated vehicles, metrics are described to maximize understanding by a broad audience. Note that ADS technology and safety assessment methods and metrics are still evolving. It is reasonable to consider future revision or expansion based on knowledge and experience gained following wider-spread deployment of ADS technology.

2. References

2.1 Applicable Documents

The following publications were referenced during development of this document. Where appropriate, documents are cited.

2.1.1 SAE Publications

Unless otherwise indicated, the latest issue of SAE publications applies. Available from SAE International, 400 Commonwealth Drive, Warrendale, PA 15096-0001, Tel: 877-606-7323 (inside USA and Canada) or +1 724-776-4970 (outside USA), www.sae.org.

AVSC00004202009 AVSC Best Practice for Data Collection for Automated Driving System-Dedicated Vehicles (ADS-DVs) to Support Event Analysis

SAE J670_200801 Vehicle Dynamics Terminology

SAE J2944_201506 Operational Definitions of Driving Performance Measures and Statistics

SAE J3016_201806 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles

SAE J3216_202005 Taxonomy and Definitions for Terms Related to Cooperative Driving Automation for On-Road Motor Vehicles

For other referenced documents see Appendix B.

3. Definitions

3.1 [ADS-Operated Vehicle Motion Control] Maneuver

Goal-oriented vehicle motion control behavior undertaken by an ADS to achieve a specific result or outcome.

NOTE 1: The term maneuver may apply to many road users, including vehicles, pedestrians, motorcyclists, and animals. However, this document specifically references vehicle motion control actions taken by an ADS.

NOTE 2: Maneuvers are a subset of tactical and operational behaviors.

NOTE 3: ADS-operated vehicle motion control maneuvers may be undertaken in service of multiple goals, which may or may not be feature-specific.

3.2 [ADS] Situation

Adapted from IEEE paper *Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving* [11].

A particular system's interpretation of a scene, including all relevant circumstances, and intentions and predicted actions of the actors involved (traffic participants).

NOTE 1: A situation is the entirety of circumstances, which are to be considered for the selection of an appropriate tactical behavior at a particular point in time. It entails all relevant conditions, options and determinants for a tactical behavior. A situation is derived from the scene by an information selection and augmentation process based on transient (e.g., mission-specific) and, permanent goals and values.

NOTE 2: ISO 21448 [12, pp. 4-5] defines situation as "a selection of an appropriate tactical behavior pattern at a particular point of time". Elaboration is based on the need for ADS systems to predict tactical behaviors of other road users.

3.3 [Driving Automation System] Feature or Application (SAE J3016_201806)

A level 1-5 *driving automation system*'s design-specific functionality at a given level of *driving automation* within a particular *ODD*, if applicable.

NOTE 1: A given *driving automation system* may have multiple *features*, each associated with a particular level *of driving automation* and *ODD.*

NOTE 2: Each *feature* satisfies a *usage specification*.

NOTE 3: Features may be referred to by generic names (e.g., automated parking) or by proprietary names.

3.4 Dynamic Driving Task (DDT) (SAE J3016_201806)

All the real-time operational and tactical[2] functions required to *operate* a *vehicle* in on-road traffic, excluding the strategic functions such as *trip* scheduling and selection of destinations and waypoints, and including without limitation:

- Lateral vehicle motion control via steering (operational);
- Longitudinal vehicle motion control via acceleration and deceleration (operational);
- Monitoring the driving environment via object and event detection, recognition, classification, and response preparation (operational and tactical);
- Object and event response execution (operational and tactical);
- Maneuver planning (tactical); and
- Enhancing conspicuity via lighting, signaling and gesturing, etc. (tactical).

[2] Per SAE J3016: "Driving entails a variety of decisions and actions, which may or may not involve a *vehicle* being in motion, or even being in an active lane of traffic. The overall act of driving can be divided into three types of *driver* effort: Strategic, Tactical, and Operational [88]. Strategic effort involves *trip* planning, such as deciding whether, when and where to go, how to travel, best routes to take, etc. Tactical effort involves maneuvering the *vehicle* in traffic during a *trip*, including deciding whether and when to overtake another *vehicle* or change lanes, selecting an appropriate speed, checking mirrors, etc. Operational effort involves split-second reactions that can be considered pre-cognitive or innate, such as making micro- corrections to steering, braking and accelerating to maintain lane position in traffic or to avoid a sudden obstacle or hazardous event in the *vehicle's* pathway."

3.5 Metric Segmentation

Application of a metric to a specific, bounded set of conditions or characteristics to enable meaningful comparison or reduction of confounding variables.

3.6 [Operating] Scenario

A description of the temporal development through several consecutive scenes in a sequence of scenes.

NOTE 1: Every scenario starts with an initial scene. In contrast to a scene, a scenario spans a certain amount of time. Actions and events can be specified as transitions between scenes to characterize the temporal development within a scenario. Scenes in a scenario can also be augmented with goals, values, and beliefs of the traffic participants, resulting in a sequence of situations.

NOTE 2: Scenarios may be defined over varying durations. A scenario may overlap with or be completely contained within another scenario. For example, an overtaking scenario may be decomposed into three scenarios: lane change scenario, followed by lane maintenance scenario, followed by lane change scenario.

NOTE 3: Scenarios may be defined at varying levels of abstraction, ranging from individual quantitative scenarios to quantitative classes of one or more scenarios, to qualitative scenario classes with narrative descriptions [91].

NOTE 4: This document specifically references emergency and non-emergency scenarios encountered by ADS-operated vehicles.

NOTE 5: The term “operating” in this scenario definition refers to DDT performance (as opposed to, for example, a post-crash scenario in which a first responder is interacting with an ADS-equipped vehicle that is no longer performing the DDT). It also comprehends all types of operating scenarios, such as test scenarios (whether on track or in simulation), as well as scenarios encountered on-road.

3.7 Operational Design Domain (ODD) (SAE J3016_201806)

Operating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics.

3.8 Predictable Vehicle Motion Control

Predictable vehicle motion control means predictable behavior by the ADS-DV from the standpoint of other [human] road users. Predictable behavior by an ADS-DV reduces risk (e.g., a decrease in rear-end struck crashes).

3.9 Predictive [Safety] Metrics

A measurement that serves as a predictive indicator of potential [safety] outcomes.

NOTE 1: Predictive metrics attempt to estimate safety performance that aid in risk management and deployment decisions.

NOTE 2: This document provides ADS developers and manufacturers with metrics that may have predictive power of safety outcomes based on available literature and experience focused on manual driving. The predictive power of predictive metrics is ultimately demonstrated by assessing a correlation to safety outcomes. However, due to the temporal delay in measuring predictive metrics, this correlation may take time to establish. The correlation between predictive metrics and safety outcomes will be tracked and used to drive improvement of the predictive metrics and thresholds.

3.10 [Safety] Metric

A measurement used to evaluate and track safety performance [13].

NOTE 1: Safety performance can be defined by a balance of predictive metrics and safety outcomes [14].

NOTE 2: Safety metrics may be used to support key performance indicators (KPI) or safety performance indicators (SPI).

NOTE 3: Metrics in this document refer to vehicle safety. United States Code for Motor Vehicle Safety (Title 49, Chapter 301) [15] defines motor vehicle safety as “the performance of a motor vehicle or motor vehicle equipment in a way that protects the public against unreasonable risk of accidents occurring because of the design, construction, or performance of a motor vehicle, and against unreasonable risk of death or injury in an accident and includes nonoperational safety of a motor vehicle.” This document uses the term “crash” in place of “accident.”

3.11 Safety Outcome Metrics

A direct measurement of actualized outcomes or adherence to societal norms.

NOTE 1: Safety outcomes temporally lag deployment. It can take considerable time to collect a sufficient sample size to establish statistically significant measurements.

NOTE 2: Societal norms may differ by industry, geographic regions, and application. In this document, societal norms include compliance with traffic regulations.

NOTE 3: Although an ADS is expected to have benefits to other societal outcomes, such as mobility and accessibility, the focus of this work is on safety outcomes, only.

NOTE 4: ADS safety outcome metrics concern a variety of ADS market penetration rates but are generally assumed to be commercial-scale deployments, and not pilots with vehicles operated by highly trained drivers employed by ADS technology developers or small-scale demonstrations, such as very low-volume deployments, limited geographic regions, or tightly limited ODD.

3.12 Scene

A snapshot of the environment including the scenery, dynamic elements, and all actor and observer self-representations, and the relationships between those entities.

NOTE 1: This document specifically references operational scenes encountered by an ADS-DV.

NOTE 2: Only a scene representation in a simulated world can be all-encompassing (i.e., an objective scene or ground truth). In the real world, the scene is incomplete, incorrect, uncertain, and from one or several observers’ points of view (i.e., a subjective scene).

NOTE 3: A scene is a descriptive representation of the state of the world at a point in time. A scenario consists of a sequence of scenes.

EXAMPLE: At an instant in time, an ADS-DV is travelling at 35 km/h, in the right hand lane on an arterial roadway in clear conditions, while another human-operated vehicle travels in the adjacent left lane at 33 km/h with the ADS-operated vehicle located in the blind-spot of the human-operated one.

3.13 Usage Specification (SAE J3016_201806)

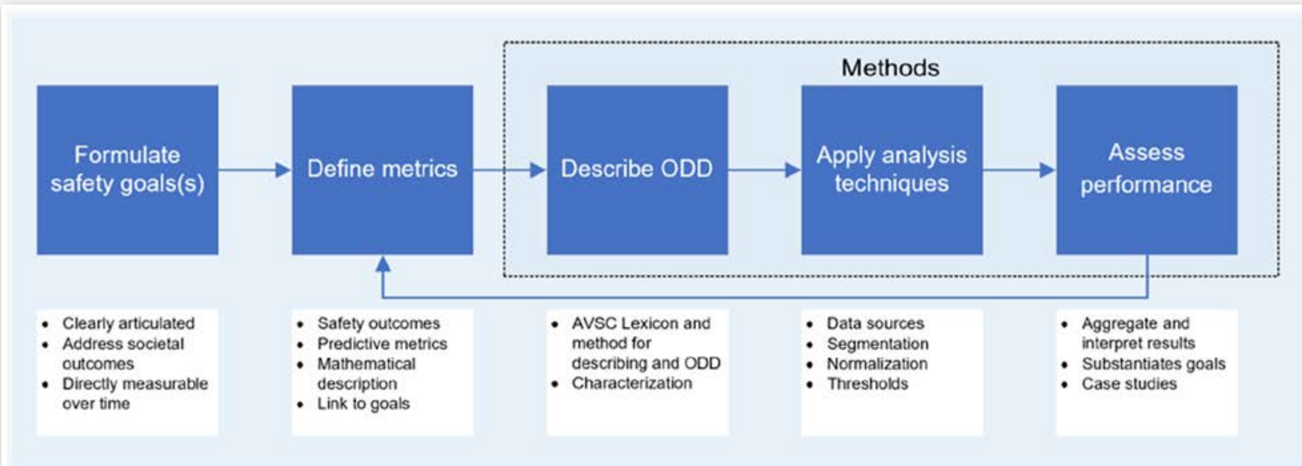
A particular level of *driving automation* within a particular *ODD*.

NOTE: Each feature satisfies a usage specification.

4. Metrics to Support ADS Safety

ADS developers and manufacturers should build evidence, including metrics, to support the argument that their ADS is acceptably safe to operate on public roads. Metrics support this argument, which is described in terms of safety goals that lend themselves to quantitative assessment[3]. Figure 1 describes a recommended process that builds evidence of safety performance beginning with defining societal-level safety goals, the use of predictive metrics and safety outcomes to support risk management and substantiate established safety goals.

FIGURE 1 Recommended process to provide evidence of safety performance.



As described in the Introduction, DDT-relevant metrics face several challenges to industry-wide adoption, including credibility, practicability, and consistency. While developing this document, “state-of-the-art” ADS system-level safety metrics and methods for assessing DDT safety performance were reviewed (Appendix C), and a set of recommended metrics was developed based on the following criteria.

Criteria for DDT-Relevant System Level Metrics

1. The metric has an established relationship to safety outcomes for human drivers.
2. Real-world ADS performance can be measured and assessed practicably by ADS developers and manufacturers.
3. The metrics and associated methods are technology neutral and can be consistently applied to different ADS usage specifications and across test venues.

The metrics and methods described in this document are meant to apply to fleet-operated ADS-DVs operated on public roads with or without human passengers. They have also been selected and described to maximize comprehensibility by a broad audience in support of public acceptance. The metrics and methods described accommodate interactions with other road users (as defined in SAE J3216) and are intended to aggregate relevant driving exposure, including events captured in the metrics where the ADS may not be at fault (e.g., other road user violates ADS right of way and creates an unavoidable hazard). As previously noted, additional metrics should be applied in combination with these metrics to provide evidence of safety performance at different stages of product development and deployment.

[3] “Above all, the safety case exists to communicate an argument. It is used to demonstrate how someone can reasonably conclude that a system is acceptably safe from the evidence available. Absolute safety is an unobtainable goal. Safety cases are there to convince someone that the system is safe enough (when compared against some definition or notion of tolerable risk).” [89, p. 1] To visualize the goal structure, organizations may consider using goal structuring notation (GSN) [82] [83].

4.1 Safety Goals

The following high-level goals for desired societal impact were used to guide the development of this best practice:

1. **Reduce the number and severity of crashes.**
2. **Perform contextually safe vehicle motion control.**

Societal harm is most directly measured by both the severity and frequency of crashes; at a global level, a positive risk balance is achieved when the combination of frequency and severity are reduced. This document focuses on DDT-relevant crashes. Crashes related to the broader driving task, e.g., driving under the influence or a tire failure, are not the focus of the document.

Contextually safe vehicle motion control is intended to convey that the AV is performing the DDT in a way that mitigates situations that have an established correlation with risk. This includes acting in a way that helps other road users to anticipate AV driving maneuvers in a way that is similar to human driving norms. The context that leads to a determination of safe vehicle motion control depends on many factors, including presence of and distance to other road users, predictability of motion (e.g., acceleration and jerk), and types or state of infrastructure, signage, and traffic control devices. Some methods (e.g., RSS, SFF, ISM) formally define an interpretation of "proper response."

The AVSC acknowledges that the UN ECE WP29 forum has chosen a safety argument, namely, "free of unreasonable (foreseeable and preventable) safety risks." [16]. This best practice establishes objective and practicable safety metrics to inform the analysis expected to support that argument.

4.2 Safety Performance Metrics

DDT-relevant safety goals should be supported by metrics**. ADS developers and manufacturers should start with the set of safety performance metrics summarized in Table 1 to support safety goals.**

TABLE 1 Recommended set of safety performance metrics for ADS developers and manufacturers.[4]

Category[5]	Safety Performance Metrics	Description
Crashes	Crash severity and frequency	Contact that the subject vehicle has with an object, either moving or fixed, at any speed resulting in fatality, injury or property damage.
Compliance with traffic regulations	Severity and frequency of citable offense	A citable violation of traffic regulations pertaining to DDT performance.
Maintain a safety envelope	Longitudinal and lateral distance (may be a function of contextual modifiers)	A violation of a kinematically defined state space around a vehicle that represents a buffer between the subject vehicle and other objects in the environment. The separation threshold may be contextually modified, e.g., based on the time to a collision (TTC) between the vehicle and other objects if they continued on their current trajectories [18]. The threshold may also be contextually modified by absolute velocity of the ADS or other road users.
Exhibit contextually safe vehicle motion control	Acceleration (longitudinal and lateral)	High acceleration events (both positive and negative) are measured based on the rate and duration of events, i.e., as the summation of all instances and duration of time that the subject vehicle accelerates above a threshold value.
	Jerk (longitudinal and lateral)	High jerk events (both positive and negative) are measured based on the rate and duration of events, same as acceleration.
Object and event detection and response (OEDR)	OEDR reaction time	The time it takes for the ADS to initiate a measurable response following the onset of an initiating event in the context of scenario-based testing.[6]

[4] It is assumed that vehicles meet or exceed required governmental regulations for safety, have been designed and tested with appropriate functional safety standards (e.g., ISO 26262), and are built with appropriate quality standards to ensure reliable operation.

[5] Measures can serve as both predictive metrics and safety outcomes. Achieving statistically significant results for safety-relevant events across the entire ODD can require a significant time delay (i.e., for safety outcomes), but even a limited number of safety-related events can provide timely insights into safety performance in particular operational contexts (i.e., thereby serving as a leading metric).

[6] This definition has been adapted from the same defined term found in SAE J2944 *Operational Definitions of Driving Performance Measures and Statistics*.

For safety goal #1 (4.1), property damage, injury severity and crash frequency metrics are directly indicative of society-level safety outcomes. The economic and social costs of vehicle crashes are well documented [17]. For safety goal #2 (4.1), compliance with motor vehicle codes and ODD-specific regulations provide a measure of contextually safe vehicle motion control.

ADS developers and manufacturers should use such metrics as key performance indicators (KPIs) to support development and deployment decisions. Metrics should effectively represent typical risks associated with a given usage specification. The usage specification includes the ODD and design-specific functionality. The usage specification is relevant to segmenting (7.3) (4.2) and normalizing (7.3) (4.2) metrics. The ODD specification can also be used as input for developing thresholds (e.g., local crash data, historical number of traffic citations).

NOTE: Guidelines for describing the ODD for an ADS is addressed in the *AVSC Best Practice Describing an Operational Design Domain: Conceptual Framework and Lexicon*.

5. Recommended Safety Outcomes

Safety performance metrics provide quantitative evidence and help developers substantiate their safety goals. The recommended safety outcomes shown in Table 2 are direct measures of the safety goals presented in Section 4.1. Compliance to motor vehicle traffic regulations is generally assumed to be less likely to cause societal harm through contextually unsafe vehicle motion control.[7]

TABLE 2 Recommended safety outcomes

Category	Metric Parameters	Definition
Crashes	Crash severity, I Frequency, f_i	Contact that the subject vehicle has with an object, either moving or fixed, at any speed resulting in injury or property damage. When making comparisons between ADS and human performance, care should be taken to ensure proper comparability. **The severity of crashes should be normalized by exposure and may be segmented (e.g., by causation), and reported with a confidence interval that reflects uncertainty in the measurement, and compared with a threshold performance set by ADS developers and manufacturers based on the ADS design and usage specification.**
Compliance with traffic regulations	Citation severity, C_i Frequency, C_f	A citable violation of traffic regulations pertaining to DDT performance, including signs, signals and road markings. When making comparisons between ADS and human performance, traffic enforcement citations should be used to ensure proper comparability. Citable offenses (i.e., committed by the ADS but not cited) may be used as a predictive indicator. **Citable offenses should be normalized by exposure, reported with a confidence interval that reflects uncertainty in the measurement, and compared with a threshold performance set by ADS developers and manufacturers based on the ADS design and usage specification.**

Safety outcomes, by their very nature, have the benefit of significant exposure in relevant settings. They are the foundation of today's traffic safety metrics and, as such, are also the international standard for characterizing traffic safety in terms of societal impact. To this end**, the safety impact of ADS deployments should be measured over time to quantify progress toward meeting safety goals using measured safety outcomes that directly capture crashes and compliance with traffic regulations (and associated severities).**

5.1 Crash Severity and Frequency

Crash severity and frequency metrics support safety outcomes and contribute to a company's evidence of safety performance. Crashes may be defined as contact that the subject vehicle has with an object, either moving or fixed, at any speed resulting in injury or property damage. This definition is consistent with the definition used in U.S. DOT National Transportation Statistics [19]. Other terms used to describe crashes, including collisions [20] and accidents [21], are included as long as they meet the criteria described above. Societal harm is the combination of severity and frequency of the total crash population for the ODD of interest, where the goal is to reduce both.

[7] It is worth noting that rule-abiding drivers are not always safer drivers.

Humans are susceptible to crashes due to impaired, emotional, or distracted driving, factors not directly applicable to an ADS; however, new factors resulting from the introduction of ADS in mixed traffic could arise as ADS deployment becomes more widespread (e.g., rear-end crashes caused by impatient human drivers). Over time, as ADS-DVs become a larger percentage of all vehicles on roads, it can be expected that crashes related to both human factors and rare ADS errors will decline as ADS developers and manufacturers analyze data to identify challenging driving scenarios to avoid crashes and identify trends (7.3.4).

ADS developers and manufacturers should collect crash data for both reported and non-reported crashes involving their ADS-DVs. Non-reported crashes will typically be lower in severity than reported crashes and therefore challenging to develop robust data collection triggers and processes. These low severity events provide an important predictive indicator of future safety outcomes.

The *AVSC Best Practice for Data Collection for Automated Driving System-Dedicated Vehicles (ADS-DVs) to Support Event Analysis* defines triggers that may be useful in identifying crashes, including delta-V,[8] deployments of non-reversible restraints, DDT-fallback, and other events that would be reported, including crashes with pedestrians and other vulnerable road users. For example, an ADS failure to detect and respond to a slower lead vehicle is DDT-relevant, and a resulting crash would be recorded as a safety outcome metric. Some events recorded based on the triggers in AVSC's Data Collection best practice are not crash events (e.g., exiting the ODD) and therefore not counted as a safety outcome metric.

In addition to tracking crashes as a safety outcome, it is recommended to monitor crashes as a triggering event for case study analysis (7.2). Each crash or violation has the potential to offer lessons learned about safety performance of an ADS, e.g., to understand precipitating and contributing factors [1]. Beyond providing potential triggers for detecting crashes, AVSC's Data Collection best practice also recommends data elements potentially useful in conducting case study analysis of crashes.

5.2 Citation Severity and Frequency

ADS-DVs should be designed to comply with normative traffic regulations applicable to its ODD, with necessary exceptions allowable in cases where a temporary deviation is needed to avoid a crash. Traffic laws are designed to help improve safety by reducing risks associated with factors known to correlate with crashes [22]. For example, in 2017, 26% of crashes were speeding related and compliance with speed limits reduces the occurrence of this type of crash [23]. Traffic regulation or motor vehicle code infractions (with no collision) are measured by severity of citable offenses, C_i, and frequency of citable offenses, C_f. **To support the safety goals, the severity and frequency of citable offenses relevant to DDT performance should be documented.**

Some traffic violations, such as a missing or damaged license plate, are not related to safety, and should not be counted toward this crash safety metric. Similarly, "violations" pertaining to rules of the road that are not codified in law, but are understood locally (e.g., Northern New Jersey's convention of allowing the first car in line to turn left across traffic to proceed when the light turns green, rather than exercising what is otherwise one's right-of-way) should not be counted toward this safety metric. The phrase "rules of the road" adds social norms and courtesies to traffic regulations, compliance with which is sometimes called "good roadmanship." Rules of the road are not included in this metric because of the difficulty measuring them consistently and lack of reliable data sources to describe them. Also, as noted at the beginning of this section, in certain situations (such as having to cross double lines on the roadway to travel safely past a broken-down vehicle on the road) human drivers may temporarily violate certain state motor vehicle driving laws [24, p. 15]. These allowable exceptions are outside the scope of this document and should not be counted as part of this metric. Finally, offenses that are not relevant to DDT performance (i.e., the only part of the broader driving task that an ADS performs) are out of scope for this metric (e.g., proper use of child restraints, snow chains, vehicle equipment violations). In addition, offenses related to aberrant human driving behavior[9] (e.g., impaired driving, reckless driving and street racing) are also out of scope because an ADS is not susceptible to such behavior by design.

[8] Delta-V, sometimes confused with relative velocity, corresponds to the change in velocity that the vehicle sensing system experienced during the recorded portion of the crash event, which takes into account the relative masses of crash partners (including objects). With respect to injury outcomes, crash Impact Direction (or Principal Direction of Force) also plays a role and should also be considered.

[9] Alcohol-impaired driving accounts for 29% of total fatalities [81]. Speeding accounts for 26% of traffic fatalities [84]. An ADS cannot drive impaired and any propensity for an ADS to speed is unlikely to be the same signifier of risk as when a human driver speeds (e.g., in order to avoid creating a hazard when prevailing traffic is also exceeding the speed limit by a significant amount).

Severity of citable offenses, C_i, means how great a risk to traffic safety is presented by a particular infraction and may be measured using a variety of scales and may differ by ODD. States employ a variety of point systems [25] which may be used as severity scales. Accumulating citations has consequences for human drivers, such as a suspended license and higher insurance premiums. Severity may also be based on the insurance rate increases for each type of citation [26]. **ADS developers should clearly define and consistently apply a severity scale for C_i to similarly enable tracking and benchmarking of DDT performance.**

Frequency of citable offenses, C_f, normalizes the number of occurrences by exposure (e.g., time, distance, scenario). For example, the number of red-light violations may be cited as a frequency based on the number of signalized intersections encountered. Citation frequency is also sensitive to consistency of enforcement (7.3.4).

The *AVSC Best Practice for Data Collection for Automated Driving System-Dedicated Vehicles (ADS-DVs) to Support Event Analysis* defines triggers that may be useful for identifying events that would be law enforcement reportable. Some events recorded by the triggers in this best practice may not be citable offenses, for example, exiting the ODD.

Each citable offense has the potential to offer lessons learned about safety performance of an ADS, e.g., through case studies (7.2). For example, a disproportionate number of impeding traffic citations may be indicative of ADS driving behaviors that are overly conservative. DDT-relevant citable “impeding traffic” offenses are correlated with crash rates, including improper turns, right-of-way errors, signal violations, stop or yield sign violations, wrong side of road, failure to signal, driving too slowly, or sudden or improper braking or stopping [27] [28]. The *AVSC Best Practice for Data Collection for Automated Driving System-Dedicated Vehicles (ADS-DVs) to Support Event Analysis* recommends data elements potentially useful in conducting case study analysis of citable offenses.

Developers may detect citable offenses by mathematically formalizing traffic laws to allow for automated monitoring of traffic law compliance and detection of violation instances [29] [30]. Implementing this method requires formalizing traffic laws which often include subjective judgements [31] [32]. In this method, any number of traffic laws could be formalized, with the frequency and severity of traffic law violations tabulated. Some governments have recognized the challenges posed in applying traffic laws to ADS-operated vehicles and have suggested collaborative activity to harmonize the interpretation of traffic laws by ADSs across developers [33] [34]. Clear and harmonized traffic law specifications would lead to more consistent interpretation of metrics based on citable offense.

6. Recommended Predictive [Safety] Metrics

Safety Outcome Metrics measure progress towards overall safety goals. Unfortunately, they may require considerable exposure (and associated time) to achieve statistical significance. Predictive indicators, often one or two steps removed from direct measurement of the safety goal, may help provide advance indication of safety outcomes [1] [35] [36].

Metrics that assess performance of the DDT can serve as predictive indicators of safety outcomes by measuring the occurrence of behaviors and events that either promote or detract from a safe driving environment. As with the safety outcome metrics previously discussed, it is important that results are based on sufficient data. Predictive indicators can be determined with higher confidence much earlier than the exposures typically needed to assess safety outcomes.

As ADS-operated vehicles become more widespread, metrics indicative of competent DDT performance such as safety envelope maintenance; predictable vehicle motion control; and the time required to detect, recognize, and react to objects and events should be considered as causal factors for crashes. There are similar measurements used to assess traffic safety performance of human-driven vehicles, including time-to-collision, hard braking, and sudden steering maneuvers. Because an SAE level 4 or 5 ADS performs the complete DDT instead of a human driver, it can be assumed that violations of a threshold for these types of metrics should be a valid indicator of safety performance.

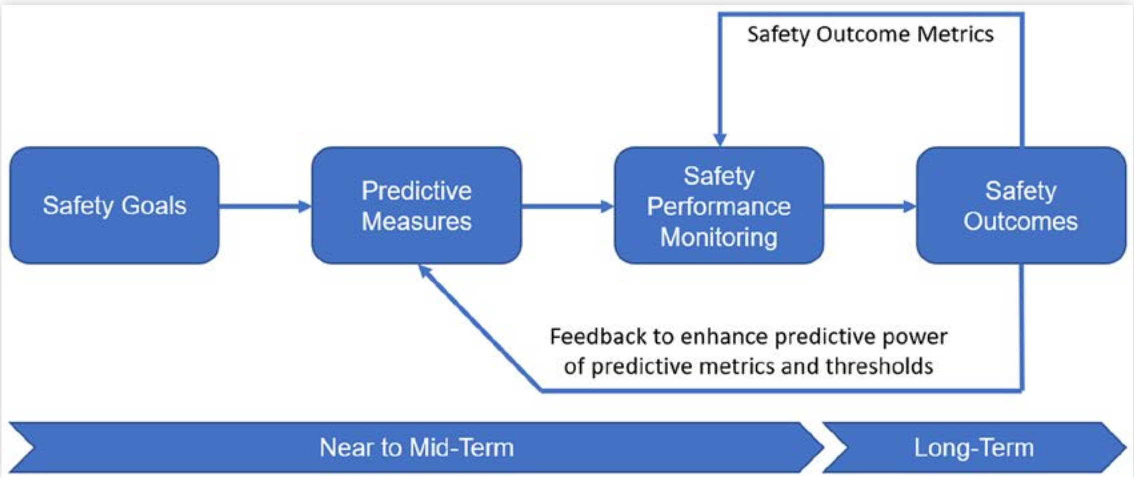
ADS developers and manufacturers should use predictive metrics summarized in Table 3, based on their correlations to safety outcomes. Predictive indicators may be assessed using various methods of data collection and analysis, e.g., using on-board and off-board measurement systems, which are also described in Section 7.

TABLE 3 Recommended predictive metrics.

Category	Predictive Metric	Potential Correlation to Safety Outcomes
Maintain a safety envelope	Longitudinal and lateral distance	Distance violation events may correlate with certain crash types, including crashes involving pedal cyclists [37], and rear-end, turning, weaving, hit objects or parked vehicle, crossing and hit pedestrian [38] [39] [40]. For example, rate of TTC violations, where TTC is one example of a contextually modified safety envelope, have been shown to be correlated with several DDT-relevant crashes including rear-end, turning/weaving, hit objects or parked vehicle, crossing and hit pedestrian [38] [39] [40]. TTC-based vehicle safety features[10] have been shown to be correlated with reduced crash rates [41] [42] [43].
Exhibit contextually safe vehicle motion control	Acceleration (longitudinal and lateral)	High acceleration events have been shown to be correlated with crash rates [23] [44] [45] [46] [47] [48] [49].
	Jerk (longitudinal and lateral)	High jerk events have been shown to be correlated with crash rates [50] [51].
OEDR	OEDR reaction time	Driver perception and reaction times have been shown to be correlated with crash rate [23] [27] [52].

As part of a robust risk management strategy, predictive metrics support decisions on a variety of timelines. Since measurements for safety outcomes require accumulation of in-use experience and events (i.e., crashes or citations), industries and industry sectors have identified *predictive metrics* helpful for predicting safety outcomes [13] [35] [53] [54]. Achieving high statistical confidence[11] that metrics are representative of the range of anticipated operating conditions will take longer for safety outcome metrics than for predictive metrics due to the frequency with which these events are experienced, i.e., crashes occur less frequently than near-crashes. Therefore, robust risk management uses both safety outcomes and predictive measures (Figure 2). Thereafter, periodic re-analysis will be needed to reestablish the validity of safety outcome metrics in view of technological advancements in the ADS-DV under development. The high level of exposure highlights the challenge of waiting for safety outcomes as safety indicators of ADS performance during pre- and early-deployment phases and motivates the need for ongoing development and continued validation of predictive metrics.

FIGURE 2 Predictive metrics and safety outcome metrics support risk management by helping predict future safety outcomes.



[10] TTC-based vehicle safety features include forward collision warning and automated emergency braking systems.

[11] There are many possible methods for determining statistical significance, such as reliability testing, confidence intervals, and power of hypothesis testing [78]. ADS developers and manufacturers may use any reasonable approach. It should be noted that for a target ODD, it may be possible to achieve a significantly higher exposure rate through different segmentation strategies (e.g., based on frequently visited intersections). Therefore, it will be able to get to a higher level of statistical significance sooner. RAND report *Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?* provides an example of how many miles may need to be driven to determine statistically significant estimation of safety outcomes [78]. The analysis used to generate this data is based on reported crashes in U.S. miles accumulated and should be representative of the entire ODD.

A baseline set of ADS safety performance indicators allow for more efficient and timely process adjustments improving long-term safety outcomes [55]. Industries ranging from public sector to chemical and aerospace use combinations of predictive metrics and safety outcomes to predict and reinforce safe performance of systems by managing risk [14]. **Metrics themselves should be assessed periodically by the ADS DV-developer to verify the correlation between predictive metrics and safety outcomes is maintained, through a feedback loop shown in Figure 2.**

6.1 Safety Envelope Maintenance

A safety envelope is a kinematically defined state space around a vehicle that represents a buffer between the subject vehicle and other objects in the environment. Maintaining a safety envelope contributes to a safer driving environment by allowing the ADS time and space to respond to the actions of nearby road users and other objects and by allowing other road users to potentially correct erroneous driving. With time and space, an ADS-DV can execute maneuvers (e.g., braking or steering) that may reduce the likelihood of a crash or the potential severity of impact compared to no action being taken (i.e., serves as surrogate measure for potential severity).

Experienced drivers know that safe driving involves adapting what is referred to as the “safety envelope” for the following factors, each of which can be addressed with the predictive metrics. That is, safe driving effectively accounts for the following:

- **Absolute velocity:** Being mindful of posted speed limits, considering current in-traffic conditions (including road surface and environmental conditions) and allowing more distance to other objects at higher vehicle speeds (i.e., lateral/longitudinal distance).
- **Relative velocity:** Timely response to a closing gap(s) which is related to the TTC to other object(s).
- **Vulnerable road users:** Allowing safe distance to pedestrians, and pedal cyclists (i.e., lateral/longitudinal distance).
- **Questionable predictability of other road user behaviors:** Allowing greater distance to other road users in scenarios where their intentions are unclear due to erratic behavior and/or poor visibility/perceptibility (i.e., lateral/longitudinal acceleration, lateral/longitudinal jerk, and OEDR).

Safety envelope violations are indicative of potentially hazardous driving conditions. The majority of NHTSA’s pre-crash scenarios would represent a safety envelope violation, if that metric were captured. Examples include scenarios with a decelerating or stopped lead vehicle, cut-in, drifting in the same direction or against oncoming vehicles, various intersections and turning scenarios, and vulnerable road user scenarios [56]. Maintaining a safety envelope is consistent with the practice of defensive driving which is generally accepted as safer driving.

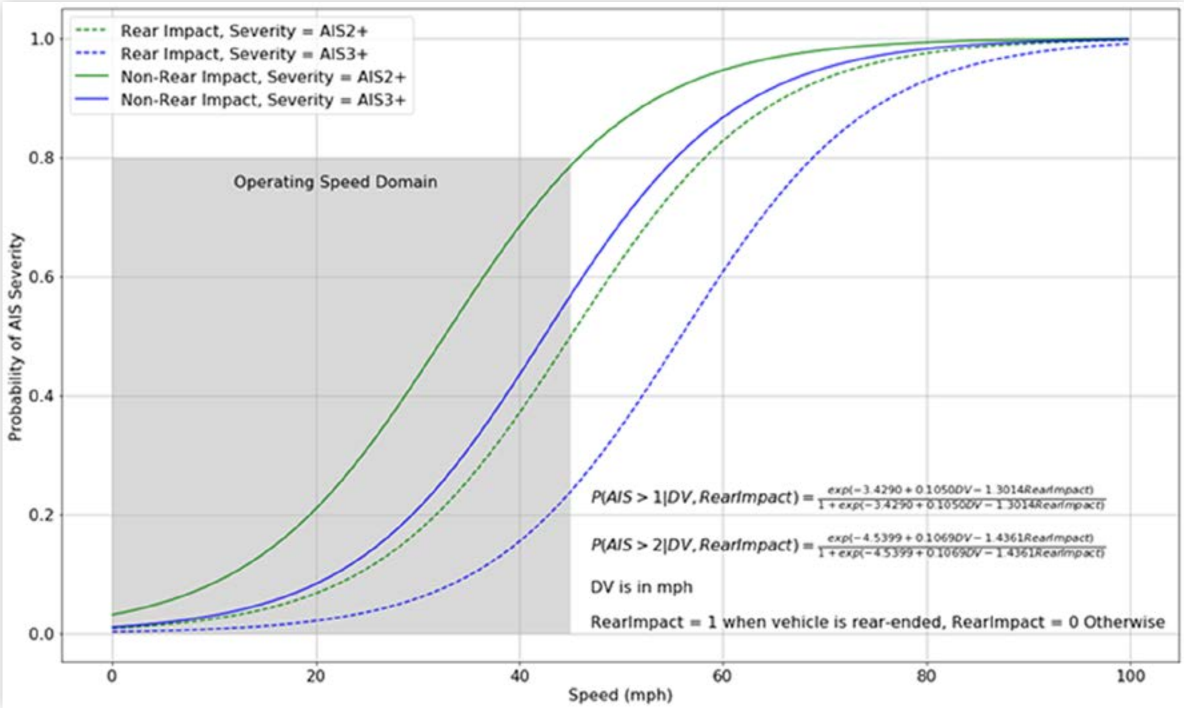
One example of a context-modified safety envelope is TTC where the context includes the dynamic relationships between actors. The rate of TTC violations have been shown to correlate with several DDT performance-relevant types of crashes, including rear-end, turning or weaving, struck object or parked vehicle, crossing and hit pedestrian [38] [39] [40]. The presence of TTC-based vehicle safety features[10] has also been shown to be correlated with reduced crash rates [41] [42] and with lower severity pedestrian crashes, because TTC-based intervention can decrease impact speeds [43]. Decreasing impact speed has been shown to decrease a pedestrian’s risk of fatality and injury [57]. The severity of rear impact crashes is shown to correlate to the change in vehicle velocity from before a crash to after the crash (delta-V), as shown in Figure 3.

6.1.1 Application of Safety Envelope Metric

Maintaining a safety envelope can be defined in terms of distance to other objects or absolute velocity of other road users (e.g., by time-to-collision). Safety envelope can be described with kinematic equations and may not include objects that are fixed to the environment (e.g., curbs). The spatial separation exists in all directions. Different coordinate systems may be useful to describe distance, e.g., right-hand Cartesian coordinate system or longitudinal and lateral directions based on the lane geometry (i.e., because vehicle motion is oriented in the longitudinal direction within traffic lanes).

Not all safety envelope violations are the fault of the ADS-DV, for example another vehicle may cut into the ADS-DV’s lane of travel. In this example, the ADS can competently perform OEDR (6.3), indicative of proper DDT performance

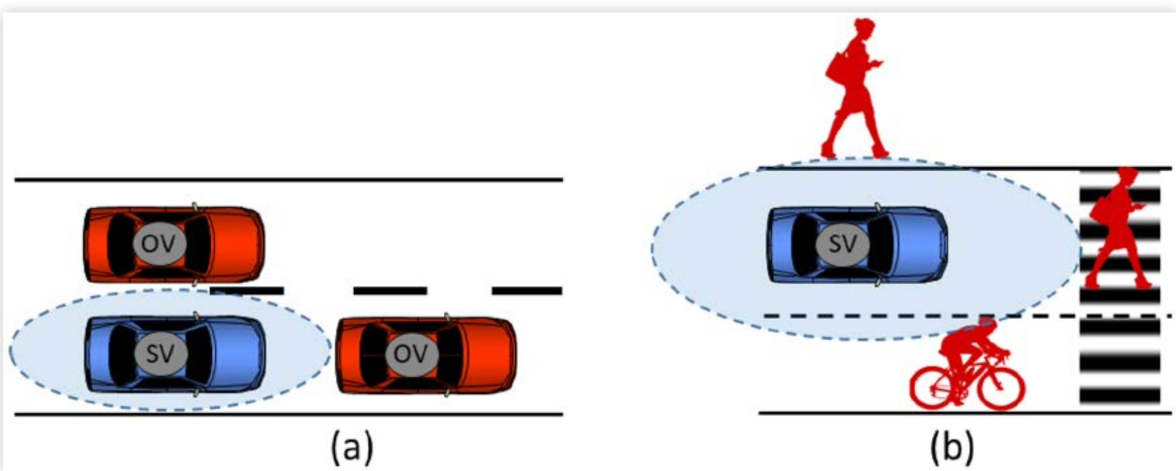
FIGURE 3 Correlation between MAIS 2+ and 3+ injuries with delta-V (based on CISS).



but still violate the safety envelope threshold. In some cases, it may be necessary to exceed a safety envelope threshold in safety critical or challenging situations as part of properly performing the DDT[12]. The prevalence of these potentially hazardous events is captured by the metric. The appropriateness of the DDT response to a specific scenario is determined *post hoc* when metric data is analyzed.

ADS developers and manufacturers should record instances when the distance between the ADS-DV and other objects violates the safety envelope. The threshold distances may vary based on several factors, including road user type (Figure 4), subject vehicle speed, and applicable traffic regulation(s). The reference frame and measurement

FIGURE 4 Depiction of longitudinal and lateral distance thresholds between a subject vehicle (SV) that can vary for (a) other vehicles (OV), and (b) vulnerable road users.

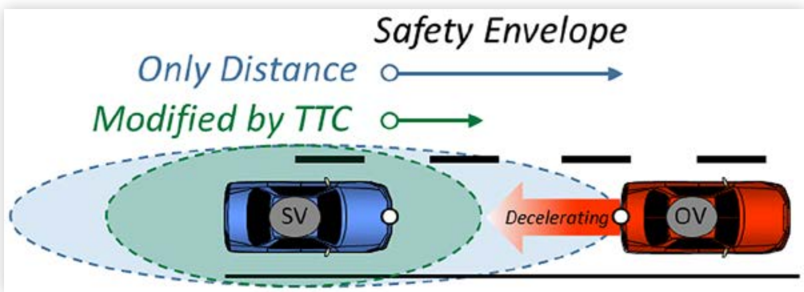


[12] Violations of the safety envelope may be indicative of a "near-miss" event or failure to maintain sufficient headway to perform an evasive maneuver if necessary. However, not all safety buffer violations may be near-misses, and not all near-misses may be safety envelope violations.

points should be defined. For example, the coordinate system may be defined in a right-handed Cartesian coordinate system (e.g., as defined in SAE J670 and ISO 8855) or road-aligned curvilinear coordinate system [58]. The origin of the coordinate system may be defined as a fixed universal inertial origin or vehicle-based origin [59]. Examples of vehicle-based origins include using the ground below the center of the rear axle or vehicle center of gravity. Objects may be represented as point masses or multi-dimensional bodies, as long as the safety envelope determination appropriately captures the offsets between the outer-most surfaces of objects that, once in contact, would represent a collision.

As noted earlier, TTC is the time-to-collision between the subject vehicle and other objects if they continue their present speeds, accelerations, and trajectories, and is a representation of the relative motion between the object and the subject vehicle. ADS developers and manufacturers should record critical TTC between the subject vehicle and salient objects in the path of the subject vehicle. TTC provides a pragmatic way to represent distance with the additional context of relative vehicle speed and acceleration. When TTC is used in the formulation of a safety envelope, a variety of context is covered. For example, distance is more effective when the subject vehicle velocity is zero, and TTC is more effective when the subject vehicle is approaching an object. Figure 5 shows this example for an ADS-DV subject vehicle approaching another vehicle, with the latter being the lead vehicle. This case leads to a simple formulation of TTC because the vehicle trajectories are co-linear (formulation given in [60]).

FIGURE 5 Depiction of how TTC creates a more context-aware representation of safety envelope by considering factors such as velocity and acceleration of the subject vehicle (SV) and other vehicle (OV).



Safety envelopes may be calculated in a variety of ways and other methods of calculation may be acceptable as long as they reliably serve the same purpose.

The concept of safety envelopes are present in various industry safety-related models, including Responsibility Sensitive Safety (RSS) [8], safety force field (SFF) [9], criticality metric [61], and instantaneous safety metric (ISM) [10]. RSS and SFF make assumptions about road users that in many cases can be expressed as a safety envelope metric and threshold. The criticality metric and ISM formulate the reachability of a safe state based on kinematics that contemplate a wide range of possible paths for the subject vehicle. To find the most critical state within the range of potential future states, these approaches require optimization of a multidimensional state space. Formulations can incorporate a variety of physics considerations and parameters, such as dynamic modeling of surface interactions and suspension systems to achieve the required level of accuracy.

6.2 Vehicle Motion Control

Contextually unsafe driving behaviors are the cause or a contributing factor in the vast majority of human-operated vehicle crashes [23]. In an ADS, this is indicative of sound OEDR performance. For the purposes of this best practice, predictable vehicle motion control is described in terms of frequency of maneuvers that fall outside the typical dynamic behavior of the vehicle.

Human-driven vehicles operated in an erratic manner or suddenly changing speeds are among the behaviors that correlated with fatal crashes [47]. There is a relationship between safety performance and predictable vehicle motion control where human drivers exhibiting more hard deceleration, acceleration, and swerve maneuvers have worse safety performance than other drivers [23] [45] [46] [47] [48]. Abrupt lane changes and sharp decelerations have been found to be correlated with accident frequency [44]. High acceleration events have been found to be correlated with crash rates for different types of vehicles and drivers, including law enforcement vehicles [44] and busses [62].

Additionally, studies have indicated that areas with high rates of jerk events have been found to be correlated with increased crash rates [50] [51]. Some situations warrant high acceleration or jerk maneuvers as the appropriate response to existing traffic conditions. For example, a high-speed stop for traffic lights in the yellow light dilemma zone may result in a violation of the acceleration threshold but be the appropriate and predictable response.

Several types of unpredictable vehicle motion control are also citable offenses in certain jurisdictions. For example, “brake checking,” or rapidly decelerating for no apparent reason, is a citable offense and puts other drivers at risk. These citations typically are described in terms of human mental state - such as road rage, careless driving, and reckless driving - because for human drivers, these are the mental states that lead to unpredictable driving behavior. Ultimately, this metric captures the types of behaviors that these citations are designed to discourage.

Hard deceleration events by an ADS can result from setting high sensitivities to potential collision risk. Developers might accept some rate of hard deceleration to ensure that the system is sensitive enough to reliably detect actual collision risks to balance high-risk false negatives. When utilizing the contextually safe vehicle motion control metric, developers should take care that changes in metric performance capture an improvement in safety rather than reduced sensitivity to collision risks.

6.2.1 Application of Contextually Safe Vehicle Motion Control Metric

Contextually safe vehicle motion control is an output of ADS OEDR manifested as a maneuver that results in a positive risk balance. Because a vehicle operating in traffic both influences and is influenced by other agents in the environment, fewer violations of contextually safe vehicle motion control should enable nearby road users to better anticipate ADS maneuvers and contribute to safer roadways.

The recommended metrics for contextually safe, predictable vehicle motion control are:

- Predictable acceleration
- Predictable jerk

These metrics capture the occurrence of abrupt maneuvers that could be indicative of poor driving behavior. For example, consider an ADS-DV navigating a straight road - maintaining acceleration and jerk below some threshold can render a vehicle’s behavior more predictable for other, proximal road users, and thus contextually safe in terms of the expectations of those other road users. An example of a vehicle driving in a contextually unsafe manner might be envisioned as an impatient driver in dense city traffic rushing to make a light. This situation is likely to elicit a higher number of extreme accelerations and jerk than safer driving. In safety-critical scenarios, it may be appropriate to undertake high jerk maneuvers, but these should be relatively rare events for any given vehicle (whether human- or ADS-operated). Thresholds identifying these occurrences in various ODDs should be set to avoid excessive false triggering for events where the ADS was competently performing the DDT.

6.2.1.1 Predictable Acceleration. Acceleration is the rate of change of velocity expressed as a unit of distance per time squared (e.g., m/s^2). A frame of reference must be established, as described in Section 6.1.1. Acceleration may be calculated as the second derivative of distance, with respect to time. Examples of high acceleration events include hard braking (negative acceleration), swerving and aggressive cornering (lateral acceleration), and poorly timed highway merging requiring excessive acceleration (positive or negative acceleration).

Acceleration in both longitudinal and lateral directions is a good indicator of abrupt actions. In scenarios where abrupt action may not be warranted this would constitute unpredictable behavior. It is recommended that both longitudinal and lateral directions be considered as leading indicators of future predictable behavior of ADS-operated vehicles, but thresholds may not necessarily be oriented in these directions. For example, the metric may be assessed using the magnitude of the acceleration vector. Alternatively, the metric may be assessed with directionally sensitive thresholds; for example, having different thresholds for longitudinal and lateral directions.

6.2.1.2 Predictable Jerk. Jerk is the rate of change of acceleration expressed as a unit of distance per time cubed (e.g., m/s^3). Jerk may be calculated as the third derivative of distance with respect to time. Lateral high jerk maneuvers would be analogous to swerving to avoid an obstacle or a lane-keeping function that starts to bounce back and forth between lanes. Longitudinal high-jerk maneuvers might be represented by an abrupt deceleration profile approaching an intersection (versus a smooth deceleration approaching the intersection). Predictable jerk may be assessed by the same approaches described in this best practice for predictable acceleration. Like acceleration, jerk in both longitudinal and lateral directions is a good indicator of abrupt actions (6.2.1.1).

6.3 OEDR Reaction Time

OEDR reaction time is the time it takes for the ADS to initiate a measurable response following the onset of an initiating event in the context of scenario-based testing in a controlled environment (e.g., track testing or simulation).[6] This is a system latency metric as well as an indirect measure of the system's capability. System latency metrics measure the amount of time it takes for a system to process a unit of work. For this latency measure, the unit of work is defined by object and event detection and response (OEDR), which is a critical aspect of DDT performance. Latency metrics are used for safety validation in other domains such as for aircraft system safety [63]. A properly functioning system will keep latency within tolerances that allow it to meet functional requirements.

Crash rates are correlated with poor driver perception and reaction times [23] [27] [52]. For example, driver recognition–related errors, such as driver's inattention, internal and external distractions, and inadequate surveillance, are associated with 41% of total crashes; driver decision-related errors, such as driving too fast for conditions or too fast for the curve, false assumption of others' actions, illegal maneuver and misjudgment of gap or others' speed, are associated with 33% of total crashes [23]. OEDR is the analogous ADS metric which similarly links to crash likelihood and severity. OEDR reaction time is also predictive of the following citable offenses: improper turn, improper lane change, improper pass, obstructing emergency vehicles, and failure to yield or stop for traffic control devices and other road users.

6.3.1 Application of OEDR Reaction Time Metric

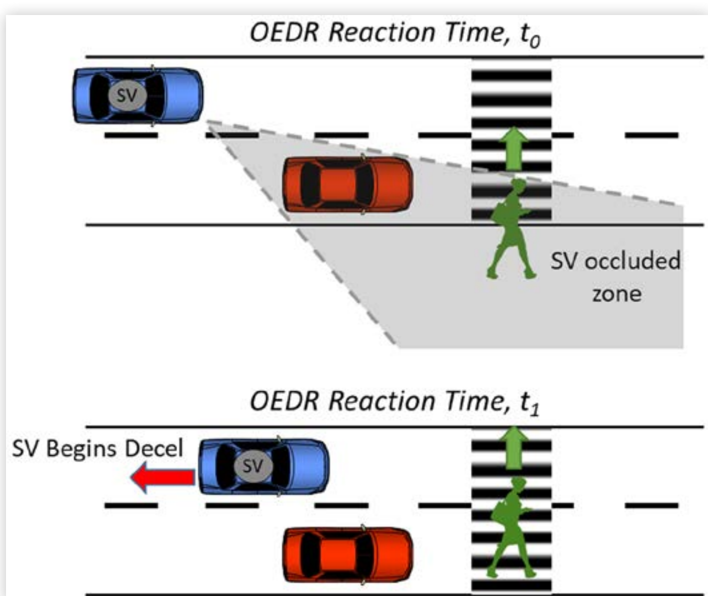
Conceptually, OEDR includes several activities, such as detection, recognition, decision, and response. The sensors must first be able to detect the presence of a salient object or event relevant to DDT performance; such detection – i.e., even prior to recognition – may result in the vehicle beginning to slow. Recognition may include classifying objects and situational context (e.g., crosswalks). Based on these cues, the ADS determines a plan and commands actuators to take longitudinal and lateral vehicle motion control actions. Actual OEDR implementations will vary among ADS developers and manufacturers, but in general, the elapsed time of all these activities constitutes the reaction time.

A technology neutral way to measure OEDR reaction time is the difference between a start time, t_0, and an end time, t_1:

$$\Delta t_{OEDR} = t_1 - t_0 \qquad \text{(Eq. 1)}$$

where the start and end time are defined using situational criteria. For example, the start time may be defined by when an object should first be reliably detected by the ADS (e.g., pedestrian steps out from behind the red vehicle as shown in Figure 6), and the end time may be defined as when the ADS first takes longitudinal or lateral control action (e.g., decelerates by braking). It is important to clearly define the start and end points for measuring this time and to measure them consistently and repeatably to ensure comparability over time.

FIGURE 6 Depiction of OEDR reaction time for a pedestrian reveal event.



ADS developers may also choose to utilize latency metrics from the subsystem components with the ADS implementation (i.e., perception, prediction, planning, actuation) in aggregate to calculate and verify the observed OEDR response time.

OEDR reaction time is complementary to safety envelope maintenance metrics and can help measure the ADS's reaction to violations by others (e.g., when another vehicle cuts into the ADS's lane of travel). If an ADS has reaction times that are too long for the scenarios experienced in the ODD, this could result in jerk or acceleration violations, safety envelope violations, or collisions due to the ADS reacting too slowly to other road users. Similar examples include NHTSA's forward collision warning test procedure for the New Car Assessment Program (NCAP) [60], which uses a longitudinal TTC metric to assess OEDR reaction time to a lead vehicle, and the Automatic Emergency Braking (AEB) pedestrian and pedal cyclist European NCAP tests [64]. One advantage of the OEDR reaction time metric is that it can be an indicator of latency-related and perception-related performance issues.

OEDR reaction time is typically evaluated in a closed course testing environment. OEDR reaction time thresholds should be determined in the context of a particular scenario or set of scenarios, including considerations for road user classification, visual obstructions, vehicle speed, road surface conditions, and other ODD elements. Data for typical human reaction time determined through OEM testing or other research, under matched conditions (to the extent possible) may serve as a reference for establishing thresholds. When available, human performance models may be used; for example, the UN regulation for automated lane keeping systems includes a human driver performance model [65].

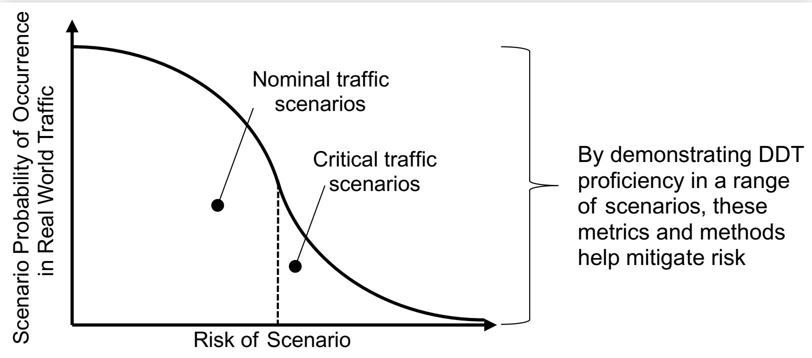
7. Methods for Assessing DDT Performance Metrics

This section discusses how to apply the safety metrics. Evidence of ADS safety performance is demonstrated using a variety of different test regimes. Applicability of the metrics to various test regimes is described in Section 7.1. The metrics can be used to evaluate specific driving scenarios (7.2) as well as can be aggregated (7.3) to establish overall performance benchmarks. A critical aspect of metrics application is establishing meaningful thresholds (7.3.3), tailored to the ADS developer's ODD and usage specification. Finally, discussion is provided on how to analyze metric data (7.3.4) factoring in exposure, segmentation, normalization, and statistical confidence in the measurements.

7.1 DDT Performance-Relevant Risk

The safety performance metrics in Table 3 can be used to assess ADS safety in a range of scenarios. One useful way to categorize scenarios is by assessing probability of scenario occurrence based on risk (Figure 7). Typical driving scenarios are encountered frequently and have relatively low risk. Critical driving scenarios, for example, re-establishing a safety buffer, occur less frequently and have relatively high risk. Edge cases, for example, an emergency response, occur rarely and are not always high risk.

FIGURE 7 Categorization of possible operating scenarios by risk.



The safety performance metrics can be monitored in various test regimes, each of which may be focused on different categories of scenarios. For example, simulation is well suited for obtaining high coverage of the scenario space (note: the meaningfulness of simulation results relies on properly validated simulators). Track testing is highly controlled and may be suitable for critical or edge case scenarios. Aggregated performance monitoring of on-road data may support evaluation of typical traffic scenarios, while on-road case studies may support evaluation of critical or edge case scenarios.

7.2 Considerations for DDT Performance Assessment Venues and Methods

The suitability of a metric to support safety goals is enhanced if the metric is useful across multiple test methods and venues. The metrics recommended in this best practice are powerful in part because they can be used consistently across simulation, track, and on-road test venues in pre-deployment test and development and post-deployment monitoring. Below are some example methods that can be used to help assess the safety performance metrics described in this best practice.

- **Simulation methods** may involve software or hardware in the loop studies that analyze the ADS performance in a simulated environment that represents a real-world environment and conditions.
- **Controlled testing** represents any testing that is completed on a closed course where the parameters can be manipulated by the experimenters.
- **A real-world scenario-based checklist** is analogous to current on-road human driving tests, and entails completing a defined set of behavioral competencies or driving conditions (e.g., left turn at signalized intersection, heavy traffic) via on-road testing within the ODD.
- **Case study and event analysis** is after-the-fact analysis of a critical traffic scenario or edge case using tools and richer data sets than aggregate on-road data collection to analyze causal and precipitating factors. The *AVSC Best Practice for Data Collection for Automated Driving System-Dedicated Vehicles (ADS-DVs) to Support Event Analysis* describes data collection for case study analysis of deployment of non-reversible restraints, exceeding the specified delta-V threshold, DDT fallback event, or other events related to driving rules and regulations.
- **Aggregate on-road data collection** represents collecting data while performing operational tests or commercial deployments that may or may not involve an in-vehicle fallback test driver (i.e., naturalistic and developmental tests), and is intended to capture the range of typical traffic scenarios within an ODD.

Individual metrics may have strengths and limitations that make them more-or-less applicable to different test methods and venues. For example, the OEDR reaction time is well suited for simulation, real-world scenario-based checklists, or controlled testing. However, there are practical resource-intensive challenges with aggregate collection of on-road measurements, including effectively classifying situations and objectively identifying a start and end time for the measurement for large amounts of test data.

7.3 Aggregation Method for On-Road DDT Performance

Aggregated metrics from on-road ADS operations can provide insights across a wide range of conditions, including typical, critical, and edge case scenarios. The strengths of aggregated analysis are the real-world fidelity and range of conditions. Effective analysis of metrics includes collecting data with sufficient exposure for typical scenarios, normalizing results by exposure, accounting for categories of risks (i.e., segmentation), establishing target thresholds, and achieving statistical confidence in results. Additionally, this type of testing may identify rare critical and edge case scenarios.

7.3.1 Application of Metrics for Aggregated On-Road DDT Performance Method

Safety may be assessed based on the rate and duration of safety-relevant events experienced by an ADS during on-road testing and operations (Table 4). ***Safety envelope*** **and contextually safe** ***vehicle motion control*** **metrics should be assessed using both methods,** where safety-relevant events are identified as occurrences in which kinematically-defined thresholds are exceeded (refer to 7.3.3 for discussion of thresholds). Threshold violations are expected to be infrequent and short in duration. ***Crashes and citable offenses*** **should be assessed using rate of occurrences.**

TABLE 4 Application of metrics for aggregate on-road DDT performance method

Representation of Exposure	Safety Envelope Violations	Contextually Safe Vehicle Motion Control Violations	Crashes	Citable Offenses
Rate of Occurrences	X	X	X	X
Percent of Travel Time of Occurrences	X	X		

Rate of Occurrences: Assess the frequency by which the observed parameter, p_i, is greater than a threshold, p_{ref}. This can serve as a good indicator of a significant event (e.g., crash, citation, safety envelope violation or unpredictable vehicle motion control). The rate of occurrences, f_p, is counted by:

$$f_p = \frac{1}{n}\sum_{i}^{n} k_i, \quad \text{where } k_i = \begin{cases} 1, \text{if } p_i \geq p_{ref} \text{ and } p_{i-1} < p_{ref} \\ 0, \text{else} \end{cases} \qquad \text{(Eq. 2)}$$

where n is the sample size to account for exposure. Alternatively, the threshold may be based on the deviation from the mean of a sample that is representative of risk exposure within the ODD. In this case, a threshold may be specified in terms of the number, C, of standard deviations, σ_p, from the sample mean, $\bar{p}$:

$$p_{ref} = \bar{p} + C\sigma_p, \quad \text{where } \bar{p} = \frac{1}{n}\sum_{i}^{n} p_i, \sigma_p = \sqrt{\frac{\sum_i^n (p_i - \bar{p})^2}{n-1}} \qquad \text{(Eq. 3)}$$

The acceptance criteria can be described by whether the rate of occurrences, f_p, exceeds the threshold, $f_{ref,p}$:

$$f_p < f_{ref,p} \qquad \text{(Eq. 4)}$$

Note that occurrences may be further analyzed based on the severity of the occurrence (e.g., using crash or citation severity scales described in Section 5).

Percent of Travel Time During Occurrences: Assess the percent of travel time, Δt, over which the observed parameter, p_i, is greater than a threshold, p_{ref},:

$$\Delta t_p = \frac{1}{n}\sum_{i}^{n} k_i, \quad \text{where } k_i = \begin{cases} dt, \text{if } p_i \geq p_{ref} \\ 0, \text{else} \end{cases} \qquad \text{(Eq. 5)}$$

where dt is the duration of the time step and n is the sample size to account for exposure. Alternatively, the threshold may be based on the deviation from the mean, $\bar{p}$, in terms of the number, C, of standard deviations, σ_p, from the mean, as discussed in Equation 3. As an example, aggregating longitudinal safety envelope violations by percent travel time can serve as a good indicator of the time spent following too closely. It should be noted that metrics are context-dependent and should be appropriately segmented to account for circumstantial variables.

The acceptance criteria can be described by whether the cumulative duration, Δt_p, exceeds the threshold, $\Delta t_{ref,\,p}$:

$$\Delta t_p < \Delta t_{ref,p} \qquad \text{(Eq. 6)}$$

7.3.2 Data Sources and Collection

Relevant data sources vary according to the metric, analysis method(s) used, and the applicable safety goal(s). Data sources may include public crash databases, law enforcement data, data from off-board devices (e.g., automated license plate readers), and data collected from onboard systems with appropriate verification.

For safety outcomes, data sources will include off-board sources, such as crash statistics, naturalistic driving data, and state safety data, specific to the ODD among others. Crash and citation data may be collected from publicly available databases, directly by ADS developers and manufacturers or in collaboration with local law enforcement [66] [67]. In the case of cited traffic offenses, data sources vary by jurisdiction and ODD. There are a variety of publicly available databases of citations, but obtaining higher geographic resolution is challenging and may require working with local law enforcement agencies for the target ODD. Citation frequency is also sensitive to consistency of enforcement, which can vary by characteristics independent of actual fault such as time, region, driver demographic, and level of traffic enforcement activity [22] [68] [69].

Care should be taken when using disparate data sources to ensure that recorded parameters are consistent and translatable for comparison purposes. Any known limitations and assumptions relevant to the analysis should be noted. In addition, the ODD and usage specification for any given ADS-DV fleet application should be factored into the metric segmentation (7.3.4.2), as well as any data sources used for establishing thresholds. For example, crashes attributed to snow/ice-covered roadways are not relevant for an ADS-DV fleet designed to operate exclusively in warm climates, and the ability to negotiate freeway exit- and on-ramps is not relevant for a usage specification that is restricted to urban streets. Some existing data sources contain ODD-relevant information (e.g., location, time of day, atmospheric conditions, lighting, sequence of events, relation to junction, manner of collision) that can be used to tailor a data set to a given ODD and usage specification. It should be noted, however, that in some cases, a dearth of existing, ODD-relevant crash data may require that additional crash data is accumulated following deployment in order to achieve a level of data granularity and quality for a given ADS-DV fleet application to support an appropriately-segmented and statistically-significant analysis.

Data should be relevant to the subject ODD in order to determine the appropriate baseline for human performance in the same environment and under the same operating conditions, and may potentially be collected from vehicles, fleets, ride sharing services, and/or infrastructure**.** Kinematic vehicle data may be collected from ADS-DV operations or available human driving data (e.g., crash databases, state or local databases, and naturalistic driving databases). This data is readily available and easily collected and may be used to establish product-specific thresholds for a given ODD. Additional data sources (Appendix D) may include appropriately matched naturalistic driving data sets, as long as these sources support calculations of desired metrics. Data for establishing thresholds for predictable vehicle motion control may be collected from naturalistic driving data, telematics data, commercially deployed vehicles, fleets [70], or other suitable source. These data could be used by ADS developers and manufacturers to determine an acceleration and jerk baseline for human performance within an ODD. Human performance data may serve as a reference and be combined with other data to establish thresholds (7.3.3). **Thresholds should be established that consider and support federal, state, and local laws (e.g., following distance).**

Care should be taken to ensure that the quality of any direct comparison of ADS safety performance to baseline safety performance of human drivers is valid in terms of performance matching, exposure, and context. For example, the U.S. DOT estimates that one third to one half of all human-operated vehicle crashes go unreported; by contrast a higher percentage of ADS-operated vehicle crashes are expected to be reported (for example, reporting of all crashes involving ADS is currently required in certain states). As a result of this disparity in crash reporting between human-operated vehicle crashes and ADS-operated vehicle crashes, comparing all human-operated vehicle crashes with all ADS-operated vehicle crashes would inaccurately skew the comparison to falsely-indicate “superior” human-operated vehicle safety.

Data collected from on-board vehicle sources may be useful for assessing distance, safety envelope, acceleration, jerk, crashes, citable offenses, and OEDR reaction time. An advantage of on-board data sources is its consistency over time. Data may be collected by an ADS-DV during development, controlled testing, field operational trials (FOT), and when fully deployed. The practicality of measuring metrics on board a vehicle in the field has been demonstrated [38] [39] [49] [66]. Fusion techniques may be used to aggregate data from multiple sources to obtain a more accurate determination of kinematics, contextual relevance, and overall safety performance.

Data collected during deployed operations should include parameters or classifications necessary for later analysis or segmentation, as these may be difficult to determine after the fact. Examples include documenting the key variables for analysis such as object type or road user and whether an event was caused by the ADS or another factor.

A list of example data sources, along with related reference materials are included in Appendix D.

7.3.3 Thresholds

Thresholds are the level or value at which a measurement is taken. Events that surpass thresholds trigger the recording or tallying of occurrences of safety relevant events. For aggregated metrics, thresholds can be used to delineate reasonably safe versus unacceptable performance of the DDT.

ADS developers and manufacturers should establish practicable ADS safety performance thresholds for metrics specific to their ADS design, ODD, and usage specification.[1] Thresholds should be established that support federal, state, and local laws (e.g., following distance). Threshold setting may be done as part of a parameterized model or test procedure-specific criteria with specified scenarios. Thresholds may depend on a variety of factors known to correlate with risk including the presence of different types of road users, and the status of traffic control devices. Section 7.3.2 provides some examples of relevant data sets, but additional data granularity may be needed. State and local agencies or other (non-ADS) fleet operators (e.g., transportation network companies) may need to be engaged to gather more granular data for an ODD.

To advance the goals in Section 4.1, thresholds for crash severity and frequency should be set such that they improve upon matched baseline aggregate human driver safety performance of the transportation system within the ODD where human performance data is available. Threshold values for an ODD may also be based on ADS-DV performance indicators (e.g., maximum longitudinal and lateral accelerations), risk factors (e.g., correlation to crash data), or societal norms (e.g., compliance with traffic regulations).

Some degree of false positives (i.e., triggered events that are not indicative of negative ADS safety outcomes) should be expected (i.e., target event rate of occurrences may be non-zero). This acknowledges that the ADS will operate in mixed driving environments where the ADS may – correctly – make an evasive maneuver to respond to another road user behaving unpredictably. Reducing non-crash related passenger injuries (e.g., whiplash) due to unpredictable driving behavior is not within the scope of this document.

Determining an appropriate threshold depends on the context of the driving environment and capabilities of the subject vehicle. Interpreting the safety or appropriateness of maneuvers that exceed thresholds is context dependent, but should be conservatively set. By way of example, the threshold for safety envelope may be a function of several parameters such as vehicle capabilities (e.g., $a_{max,decel}$), road user type, legal requirements in the ODD, and speed of the subject vehicle. Larger lateral thresholds may apply to bicyclists and pedestrians than to vehicles and larger lateral distances may apply when the ADS is operating at higher speeds. The lateral distance to a pedestrian may depend on whether the pedestrian is on a sidewalk or not. Even within an ODD, thresholds may vary from scenario to scenario. Thresholds for what constitutes contextually safe vehicle maneuvers may be sensitive to environmental characteristics such as traffic congestion, speed limit, number of lanes, managed access facilities, areas with more VRUs, etc.[13] In some cases, minimum values set by regulation will establish threshold values. For example, the minimum lateral distance between a vehicle and a bicyclist varies in the U.S. by state[14]. The threshold may differ based on intent of the ADS and other road users; for example, motorcycles performing lane-splitting. Restrictive or overly conservative thresholds could result in worse safety performance than a human driver by negatively impacting driving conditions and impeding traffic flow. (Note: applicable measurement tolerances should be accounted for in setting thresholds to prevent a compounding effect.)

Thresholds should be comprehensible by safety stakeholders. Human driving performance data can be collected from naturalistic driving studies (NDS) and simulations to provide an understanding of human driving behavior in an ODD [71]. Human performance data may be useful for setting thresholds, but should match the use and purpose of the ADS to the extent possible. Thresholds using human driving performance data from NDS and simulation may be established statistically using standard deviations from average human performance (e.g., errors per million opportunities used in Six Sigma [72]). In cases like this, the threshold is dependent on the data set used, which means it may not be easily scalable to encompass multiple ODDs. Other human-derived performance thresholds can be described based on cognitive models; for example, the time it takes to perceive and classify a scenario, decide on a maneuver, and take action [71]. Proposed models have different levels of complexity and calibration requirements and vary in practicality and generalizability to different traffic scenarios.

7.3.4 Analysis and Context

Without the right context, safety metrics have little meaning and no single metric provides a complete view of ADS safety. Therefore, **ADS manufacturers and developers should utilize the results of several metrics, evaluate them in context, and use the findings in combination with other elements of their overall safety performance evaluation (e.g., risk management plans, systems engineering verification and validation) in their deployment decisions.**

[13] TTC thresholds have been developed for many applications; for example, assessing urban intersection design [18] and developing infrastructure safety models [85]. Care must be taken in assessing the relevance of these metrics as they differ in purpose from the intent of this best practice.

[14] For example, thirty-two states have at least a three-foot passing requirement for motorists [86].

Citation metrics can be collected and analyzed in multiple ways, including for direct comparison to baseline safety of human drivers and for analysis of events for which no baseline exists. The ability to compare ADS performance to human drivers depends on data availability and completeness of the data. Law enforcement issued citations can provide a comparison to human driver performance but are subject to biases (7.3.2). In some cases, more granular data may be obtained through coordination with law enforcement agencies. ADS developers and manufacturers may also collect citable offenses in cases where no traffic citation was issued.

7.3.4.1 Exposure. Metrics and measurement data must represent the range of conditions and risks that can be reasonably anticipated within a manufacturer's usage specification. Initial ADS-DV fleets will accumulate far fewer vehicle miles traveled than all conventional vehicles and will be limited in their operational domains. Meaningful analysis can be completed by considering other measures of exposure, even for deployments in limited operational domains.

Manufacturers should document their sampling strategy and describe how it accounts for significant exposure of DDT performance events and ODD characteristics[15]. For example, logging a high number of vehicle miles travelled (VMT) in highway scenarios is not representative of ADS performance in urban scenarios. This presents a limitation to using factors like VMT and hours of operation[16] without relevant context for risk exposure. These limitations can be overcome by using other exposure factors such as number of interactions with other road users and types of typical and critical scenarios experienced such as cut-in/out, lane changes, unprotected turns, etc. Measuring exposure based on types of interactions with other road users or key infrastructure elements (e.g., complex junctions, traffic circles, or weaving sections) can help capture the complexity of an ODD and enhance scalability across domains. Segmenting by type of driving scenario can capture how frequently the ADS is required to demonstrate specific behavioral competencies and lead to higher confidence in a relatively short amount of time. For example, even though vehicle miles traveled in an ODD may be relatively low, an ADS-DV may experience a particular scenario many times on a single trip (such as turning across the path of oncoming traffic, a scenario associated with a considerable number of crashes in the U.S.) [56]. The number of times a vehicle had to take a left hand turn across path captures exposure to a known hazard, which is correlated to the confidence that this risk has been mitigated. This best practice recommends that contextually relevant measures of risk exposure should be used, while not specifying a universal exposure factor that is generalizable to all ODDs and usage specifications.

7.3.4.2 Segmentation. Segmentation refers to partitioning data into discrete categories to align normalized data with types of risks. ADS developers and manufacturers may segment data by events caused by the ADS versus events caused by other road users (e.g., ADS safety envelope violation caused by a lead vehicle cut-in or analysis of crash severity). If this segmentation is used, criteria for adjudicating the causation of the violation should be described.

Safety relevant events may be segmented by exposure factors, e.g., interactions with other road users or scenario families, and normalized by the number of times the vehicle experiences those types of interaction. Some segmentation examples that may be useful for safety analyses include:

- Type of road user (e.g., vehicle, vulnerable road user);
- Time of day
- ODD characteristics (e.g., urban versus rural, roadway type, weather conditions);
- Operating characteristics (e.g., longitudinal reference vehicle speed); and
- Crash type (e.g., rear-end, cut-ins) and predisposing and precipitating factors[17].

[15] Exposure can be normalized by many factors, such as vehicle miles traveled (VMT), hours of operation, number of interactions with other road users and infrastructure elements and number of driving scenarios encountered.

[16] "In situations of low exposure, events can be leveraged as case studies. For example, exposure on the order of millions or billions of vehicle miles traveled would be needed to detect even modest differences with statistical confidence in crash rates between conventional vehicles and AVs on public roads. It is extremely unlikely that within "development, public roads" there would be enough exposure to calculate such a difference within a reasonable amount of time [1, p. 38].

[17] Predisposing factors are specific characteristics corresponding to each system element that have some influence on whether a driving/walking/biking task will be carried out successfully (or unsuccessfully). Precipitating factors are the types and nature of events and event sequences prior to a crash that start with a "collision course" and ultimately end with either a collision or collision avoidance [87].

Contextually safe vehicle motion control data may be segmented by roadway geometries, traffic patterns, and types of road users, especially the presence of vulnerable road users. **The data analysis process should include criteria for classifying segmentation categories.**

Citable offense data can be segmented by type of offense, type of road user involved in the citation event (e.g., pedestrian, vehicle), and driving conditions (e.g., traffic level, time of day). Metrics should include cases where an in-vehicle fallback test driver takeover prevents a safety envelope violation. This type of analysis lends itself to the use of simulation.

Finally, the sample size for the segmented population's risk exposure, *n*, should be reported in all cases and should be representative of the risk associated with the ADS usage specification, including foreseeable interactions with other road users.

7.3.4.3 Normalization. Normalizing results refers to adjusting data with different measurement scales and risk levels to enable comparison across data sets and correct for known limitations. For example, two safety envelope violations per 100,000 pedestrian encounters or four citable offenses per 10,000 intersection traversals. Metrics in this document are framed in terms of the count or duration of occurrences of safety relevant events. When determining event frequency, measurements should be normalized by risk exposure to adjust data with different scales [73] [74]. Exposure normalization improves the comparability of results across different data sets by creating a common scale and improves interpretability and portability of results. For example, the overall number of crashes does not enable an assessment of safety unless the exposure is also known (e.g., number of pedestrian interactions).

Mathematically, normalization can be approached in many ways. The appropriate approach depends on the sample size and nature of the data. In the near-term with small sample sizes, descriptive statistics (i.e., mean, median, mode, range, and quartile deviation) can be used. Larger data sets may be normalized (e.g., Z score values) based on distribution of the occurrence and providing an indication of exposure to the event (e.g., Poisson, binomial, negative binomial distributions). One approach to normalize measurements and obtain the mean rate of occurrence, f, is to divide by the sample size (i.e., the risk exposure factor), n:

$$f = \frac{\text{measurement}}{n} \qquad \text{(Eq. 7)}$$

Sampling strategies should determine an appropriate time horizon that is resilient to temporary changes in performance. Larger data sets yield more significant and stable results. At the same time, crash risk is known to vary over time for humans (e.g., based on experience [69] [75]) and similarly, ADS performance is anticipated to improve over time (i.e., as machine learning algorithms improve with richer data sets and as AVs become a larger proportion of the on-road vehicle fleet). The time horizon for the sampling strategy should consider these factors.

Normalization by risk exposure allows for meaningful comparison between ADS-DVs and conventional vehicle baselines. While many public crash databases are best equipped to measure exposure in terms of VMT, some types of data sources can express exposure in terms of events or specific locations. Naturalistic driving data can support in-depth analysis of situation classification and normalization when there is enough data that can be appropriately matched with respect to a target ODD [76].

There can be considerable variation in the richness and format of crash data. Factors for risk exposure may be more challenging to obtain in certain ODDs than in others, for example based on data reporting practices. Examples of categories for crash data analysis are identified in Appendix C. **To normalize data for risk exposure, this best practice recommends segmenting crash data by the type of service offered:**

- If the service is passenger transport, then taxi crash rates can be relevant.
- If the service is goods delivery, then package delivery service crash rates can be relevant.
- If the service is long-haul trucking, then crash rates on city roads may be excluded from the analysis (in 2017, the Federal Highway Administration (FHWA) determined that there were 3,212 billion VMT, 70% of which was in urban geography [77]).

7.3.4.4 Considerations for Statistical Confidence and Trends. Statistical confidence is what allows conclusions to be drawn or inferred based on the evidence collected. Statistical confidence quantitatively describes the uncertainty associated with the likelihood of encountering an event again over a certain measurement period (such as time, VMT, exposure, etc.).

Confidence that crash metrics are representative of aggregate system performance can be calculated in several ways. One such method is outlined in the RAND methodology [78]. Confidence in the stability of ADS safety performance vis-à-vis a metric may be assessed using a moving average that measures performance over time. There are several ways to calculate a moving average. One example calculation for parameter, p, is provided below for the moving average of acceleration, $\bar{p}$:

$$\bar{p}_i = \frac{1}{n}\sum_{i=1}^{n} p_i \qquad \text{(Eq. 8)}$$

where n specifies the sample size (e.g., the time horizon), which is defined by an ADS developer or manufacturer, i is a measurement point between 1 and n; and p_i is acceleration at point i. The gradient is assessed as the difference between a moving average at the current time and a historical moving average over the time, and compared to a threshold, $\Delta\bar{p}_{ref}$:

$$\frac{\bar{p}_2 - \bar{p}_1}{t_2 - t_1} > \Delta\bar{p}_{ref} \qquad \text{(Eq. 9)}$$

Transportation and many metrics related to transportation safety are influenced by national and regional trends. Factors such as vehicle ownership, land use, economic fluctuations, and an aging population may be reflected in national crash data metrics [79]. Trends associated with factors such as these need to be monitored because they can influence baselines that may be used for comparison to ADS safety performance.

8. Summary

The safety performance metrics described in this document are intended to be a foundational set of common, system-level metrics that can contribute to the assessment of aggregate safety performance of the dynamic driving task (DDT) by an ADS. It is assumed that additional metrics will be applied in combination with these common metrics to supplement evidence of safety performance at different stages of product development and deployment. Consistent with other AVSC best practices, this document supports industry-led, voluntary approaches in the standards development community and is expected to evolve as technology matures.

Combining customized metrics with this common foundational set of metrics can enhance communication of an ADS-DV safety posture among stakeholders and accelerate public acceptance. Public agencies may use this document to better understand the safety posture of ADS deployments. In addition to the technical development community, other audiences considered in the development of this best practice include standards bodies, public agencies, and other decision-makers that may influence the deployment of ADS-DVs.

9. About Automated Vehicle Safety Consortium™

The objective of the Automated Vehicle Safety Consortium™ is to provide a safety framework around which automated vehicle technology can responsibly evolve in advance of the broad use of commercialized vehicles. The consortium will leverage the expertise of its current and future members and engage government and industry groups to establish safety principles and best practices. These technology-neutral principles are key considerations for deploying SAE level 4 and level 5 automated vehicles on public roads.

AVSC Vision:

Public acceptance of SAE level 4 and level 5 automated driving systems as a safe and beneficial component of transportation through industry consensus.

AVSC Mission:

The mission of the Automated Vehicle Safety Consortium™ (AVSC) is to quickly establish safety principles, common terminology, and best safety practices, leading to standards to engender public confidence in the safe operation of SAE level 4 and level 5 light-duty passenger and cargo on-road vehicles ahead of their widespread deployment.

The AVSC will:

- Develop and prioritize a roadmap of pre-competitive topics;
- Establish working groups to address each of the topics;
- Engage the expertise of external stakeholders;
- Share output/information with the global community;
- Initially focus on fleet service applications.

10. Contact Information

To learn more about the Automated Vehicle Safety Consortium™, please visit https://avsc.sae-itc.org.

Contact: AVSCinfo@sae-itc.org

11. Acknowledgements

The Automated Vehicle Safety Consortium™ would like to acknowledge the contributions of the member organizations during the development of this document.

Aurora Innovations, Ford, Honda, Lyft, Motional, Toyota, and VW.

12. Abbreviations

ADS - Automated Driving Systems
ADS-DV - ADS-Dedicated Vehicle
AEB - Automatic Emergency Braking
AV - Automated Vehicle
AVSC - Automated Vehicle Safety Consortium™
CISS - Crash Investigation Sampling System
DDT - Dynamic Driving Task
FARS - Fatality Analysis Reporting System
FHWA - Federal Highway Administration
FOT - Field Operational Trial
GSN - Goal Structuring Notation
IAM - Institute for Automated Mobility
ISM - Instantaneous Safety Metric
KPI - Key Performance Indicator
MPrISM - Model Predictive Instantaneous Safety Metric
NDS - Naturalistic Driving Study
NHTSA - National Highway Traffic Safety Administration
ODD - Operational Design Domain
OEDR - Object and Event Detection and Response
OV - Other Vehicles
RSS - Responsibility Sensitive Safety
SAE ITC - SAE Industry Technologies Consortia®
SFF - Safety Force Field
SPI - Safety Performance Indicators
TTC - Time-to-Collision
VMT - Vehicle Miles Travelled

Appendix A. Best Practice Quick Look

AVSC Best Practice for Metrics and Methods for Assessing Safety Performance of Automated Driving Systems (ADS)

Metrics to Support ADS Safety (4.0). ADS developers and manufacturers should build evidence, including metrics, to support the argument that their ADS is acceptably safe to operate on public roads. Figure 1 recommends a process that provides evidence of safety performance.

- **Safety Goals (4.1).** Two high-level goals.
  - Reduce number and severity of crashes; and
  - Perform contextually safe vehicle motion control.
- **Safety Performance Metrics (4.2).** ADS developers and manufacturers should start with the set of safety performance metrics (Table 1) to support safety goals.

Recommended Safety Outcomes (5.0). Recommended safety outcomes are crashes and compliance with motor vehicle traffic regulations (Table 2).

- **Crash Severity and Frequency (5.1).** Crash severity and frequency metrics support safety outcomes and contribute to evidence of safety performance. Over time, as ADS-DVs become a larger percentage of all vehicles on roads, it can be expected that crashes related to both human factors and ADS perception errors to decline as ADS developers and manufacturers analyze data to identify challenging driving scenarios to avoid crashes and identify trends.
  - ADS developers and manufacturers should collect crash data for both reported and non-reported crashes.
  - Monitor crashes as a triggering event for case study analysis.
- **Citation Severity and Frequency (5.2).** To support the safety goals, the severity and frequency of citable offenses relevant to DDT performance should be documented. ADS developers should clearly define and consistently apply a severity scale for C_i to similarly enable tracking and benchmarking of DDT performance.

Recommended Predictive [Safety] Metrics (6.0). ADS developers and manufacturers should use predictive metrics (Table 3). Metrics should be assessed periodically to verify the correlation between predictive metrics and safety outcomes is maintained, through a feedback loop (Figure 2).

Safety Envelope Maintenance (6.1). Safety envelope violations are indicative of potentially hazardous driving conditions.

- **Application of Safety Envelope Metric (6.1.1).** ADS developers and manufacturers should record instances when the distance between the ADS-DV and other objects violates the safety envelope. The reference frame and measurement points should be defined.
- **Contextually Safe Vehicle Motion Control (6.2).** Unpredictable vehicle motion control generally (i.e., except when necessary to avoid a crash) may create an unsafe situation vis-à-vis other road users.
  - **Application of Contextually Safe Vehicle Motion Control Metric (6.2.1).** The recommended metrics for predictable vehicle motion control are:
    - Predictable acceleration; and
    - Predictable jerk.
      - **Predictable Acceleration (6.2.1.1).** A frame of reference should be established, as described in Section 6.1.1. It is recommended that both longitudinal and lateral directions be considered as leading indicators of future predictable behavior of ADS-operated vehicles.
      - **Predictable Jerk (6.2.1.2).** Predictable jerk may be assessed by the same approaches described for predictable acceleration **(6.2.1.1)**. Like acceleration, jerk in both longitudinal and lateral directions is a good indicator of abrupt actions.

OEDR Reaction Time (6.3). Competent OEDR demonstrates proper recognition and decision behaviors that are directly linked to crash likelihood and severity and is also indicative of compliance with traffic rules.

- **Application of OEDR Reaction Time Metric (6.3.1).** OEDR includes several activities, such as detection, recognition, decision, and response. OEDR reaction time thresholds should be determined in the context of a particular scenario, including considerations for road user classification, visual obstructions, vehicle speed, and road surface conditions.

Methods for Assessing DDT Performance Metrics (7.0). Evidence of ADS safety performance is demonstrated using a variety of different test regimes.

- **DDT Performance-Relevant Risk (7.1).** Categorize scenarios by assessing probability of scenario occurrence based on risk (Figure 7).
- **Considerations for DDT Performance Assessment Venues and Methods (7.2).** Example methods that can be used to implement the recommended set of safety metrics.
  - Simulation method;
  - Controlled testing;
  - A real-world scenario-based checklist;
  - Case study and event analysis; and
  - Aggregate on-road data collection.
- **Aggregation Method for On-Road DDT Performance (7.3).** Aggregated metrics from on-road ADS operations can provide insights across a wide range of conditions, including typical, critical, and edge case scenarios.
  - **Application of Metrics for Aggregated On-Road DDT Performance Method (7.3.1).** Recommended evaluation methods to assess aggregate on-road DDT performance (Table 4). Safety envelope and contextually safe vehicle motion control metrics should be assessed using both rate of occurrences and percent of travel time during occurrences. Crashes and citable offenses should be assessed using rate of occurrences.
  - **Data Sources and Collection (7.3.2).** Relevant data sources vary according to the metric, analysis method(s) used, and the applicable safety goal(s). Care should be taken when using disparate data sources to ensure that recorded parameters are consistent and translatable for comparison purposes. Any known limitations and assumptions relevant to the analysis should be noted. In addition, the ODD and usage specification for any given ADS-DV fleet application should be factored into the metric segmentation (7.3.4.2). Data should be relevant to the subject ODD in order to determine the appropriate baseline for human performance in the same environment and under the same operating conditions. Data collected during deployed operations should include parameters or classifications necessary for later analysis or segmentation.
  - **Thresholds (7.3.3).** ADS developers and manufacturers should establish practicable ADS safety performance thresholds for metrics specific to their ADS design, ODD, and usage specification. Thresholds should:
    - Support federal, state, and local laws (e.g., following distance);
    - Improve upon baseline aggregate human driver safety performance of the transportation system within the ODD; and
    - Be comprehensible by safety stakeholders.
  - **Analysis and Context (7.3.4).** ADS manufacturers and developers should utilize the results of several metrics, evaluate them in context, and use the findings in combination with other elements of their overall safety performance evaluation (e.g., risk management plans) in their deployment decisions.
    - **Exposure (7.3.4.1).** Manufacturers should document their sampling strategy and describe how it accounts for significant exposure of DDT performance events and ODD characteristics.
    - **Segmentation (7.3.4.2).** The data analysis process should include criteria for classifying segmentation categories. If segmentation is used, criteria for adjudicating the cause of the violation should be described. The sample size for the segmented population’s risk exposure, *n*, should be reported in all cases and should be representative of the risk associated with the ADS usage specification, including foreseeable interactions with other road users.

- **Normalization (7.3.4.3).** Exposure normalization improves the comparability of results across different data sets by creating a common scale and improves interpretability and portability of results. The appropriate approach depends on the sample size and nature of the data. To normalize data for risk exposure, it is recommended to segment crash data by the type of service offered:
  - If the service is passenger transport, then taxi crash rates can be relevant.
  - If the service is goods delivery, then package delivery service crash rates can be relevant.
  - If the service is long-haul trucking, then crash rates on city roads may be excluded from the analysis.
- **Considerations for Statistical Confidence and Trends (7.3.4.4).** Confidence that crash metrics are representative of aggregate system performance can be calculated in several ways.

Appendix B. References

Other Documents

ISO 26262-1:2018 Road vehicles – Functional safety

ISO PAS 21448:2019 Safety of the Intended Functionality

ISO 8855:2011 Road vehicles – Vehicle dynamics and road handling ability – Vocabulary

1. Fraade-Blanar, L., Blumenthal, M.S., Anderson, J.M., and Kalra, N., *Measuring Automated Vehicle Safety: Forging a Framework* (Santa Monica, CA: RAND Corporation, 2018).
2. Aptiv Services US, LLC; AUDI AG; Bayrische Motoren Werke AG; Beijing Baidu Netcom Science Technology Co., Ltd; Continental Teves AG & Co oHG; Daimler AG; FCA US LLC; HERE Global B.V.; Infineon Technologies AG; Intel; Volkswagen AG, "Safety First for Automated Driving," 2019, [Online]. Available: https://www.daimler.com/documents/innovation/other/safety-first-for-automated-driving.pdf. [Accessed 15 June 2020].
3. Thorn, E., Kimmel, S., and Chaka, M., "A Framework for Automated Driving Systems Testable Cases and Scenarios," National Highway Traffic Safety Administration, 2018.
4. Mazzega, J., "PEGASUS Research Project," 2019. [Online]. Available: https://www.pegasusprojekt.de/en/about-PEGASUS. [Accessed 1 September 2020].
5. Safety Assurance KUdos fro Reliable Autonomous Vehicles, "Safety Assurance KUdos fro Reliable Autonomous Vehicles (SAKURA) Research Project," 2020. [Online]. Available: https://www.sakura-prj.jp/project_info/. [Accessed 1 September 2020].
6. Underwriters Laboratories, "UL 4600: Standard for Safety for the Evaluation of Autonomous Products," Underwriters Laboratories, 2020.
7. Wishart, J., Como, S., Russo, B., Weast, J., Altekar, N., and James, E., "Driving Safety Performance Assessment Metrics for ADS-Equipped Vehicles," *SAE International Journal Advances and Current Practices in Mobility* 2(5):2881-2899, 2020.
8. Shalev-Shwartz, S., Shammah, S., and Shashua, A., "On a Formal Model of Safe and Scalable Self-driving Cars," 2017.
9. Nister, D., Lee, H.-L., Ng, J., and Wang, Y., *An Introduction to the Safety Force Field* (Santa Clara: NVIDIA Corporation, 2019).
10. Every, J., Barickman, F., Martin, J., Rao, S., Schnelle, S., and Weng, B., "A Novel Method to Evaluate the Safety of Highly Automated Vehicles," in *25th International Technical Conference on Enhanced Safety of Vehicles*, Detroit, 2017.
11. Ulbrich, S., Menzel, T., Reschka, A., Schuldt, F., and Maurer, M., "Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving," in *IEEE 18th International Conference on Intelligent Transportation Systems*, Las Palmas, 2015.

12. International Organization for Standardization, "ISO/PAS 21448: 2019 Road Vehicles - Safety of the Intended Functionality," January 2019. [Online]. Available: https://www.iso.org/standard/70939.html. [Accessed 15 June 2020].
13. Occupational Safety and Health Administration, "The Use of Metrics in Process Safety Management Facilities," U.S. Department of Labor, Washington, DC, 2016.
14. Esposito, P., "Safety Metrics: Corporate & Site-Level Scorecards," *Professional Safety Journal* 30-33, June 2018.
15. "United States Code, 2012 Edition, Supplement 3, Title 49 - TRANSPORTATION," 3 January 2016. [Online]. Available: https://www.govinfo.gov/app/details/USCODE-2015-title49/USCODE-2015-title49-subtitleVI-partA-chap301. [Accessed 14 January 2021].
16. UN ECE, "Draft Annex 4 on Audit/CEL to the New UN Regulation on Automated Lane Keeping Systems (ALKS)," 4 March 2020. [Online]. Available: https://unece.org/fileadmin/DAM/trans/doc/2020/wp29grva/GRVA-06-10e.pdf. [Accessed 15 January 2021].
17. Blincoe, L., Miller, T.R., Zaloshnja, E., and Lawrence, B.A., *The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised)* (Washington, D.C.: NHTSA, 2015).
18. Hayward, J.C., "Near-Miss Determination through Use of a Scale of Danger," in *51st Annual Meeting of the Highway Research Board*, Washington, DC, 1972.
19. US DOT, "National Transportation Statistics," 2018.
20. Transport Canada, "Statistics and Data - Road and Motor Vehicle Safety - Road Transportation," 2018. [Online]. Available: https://tc.canada.ca/en/road-transportation/motor-vehicle-safety/statistics-data. [Accessed 15 June 2020].
21. World Health Organization, "World Report on Road Traffic Injury Prevention," World Health Organization, Geneva, 2004.
22. Blais, E., and Dupont, B., "Assessing the Capability of Intensive Police Programmes to Prevent Severe Road Accidents A Systematic Review," *British Journal of Intensive* 45(6):914-937, 2005.
23. NHTSA National Center for Statistics and Analysis, "Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey DOT HS 812 506," U.S. Department of Transportation, Washington D.C., 2018.
24. NHTSA, "Automated Driving Systems: A Vision for Safety 2.0," National Highway Traffic Safety Administration (NHTSA), Washington, DC, 2017.
25. FindLaw, "Driver's License Points by State," Thomson Reuters, 2020. [Online]. Available: https://traffic.findlaw.com/traffic-tickets/state-specific-points-systems.html. [Accessed 20 September 2020].
26. Megna, M., "Insurance.com," 5 May 2020. [Online]. Available: https://www.insurance.com/auto-insurance/traffic-ticket-calculator.aspx. [Accessed 10 June 2020].
27. Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., and Knipling, R.R., "The 100-Car Naturalistic Driving Study: Phase II - Results of the 100-Car Field Experiment," National Highway Traffic Safety Administration, Washington, DC, 2006.
28. Luca, D.L., "Do Traffic Tickets Reduce Motor Vehicle Accidents? Evidence from a Natural Experiment," *Journal of Policy Analysis and Management* 34(Winter 2015):85-106, 2014.
29. Rizaldi, A., Keinholz, J., Huber, M., Feldle, J., Immler, F., Althoff, M., Hilgendorf, E., and Nipkow, T., "Formalising and Monitoring Traffic Rules for Autonomous Vehicles in Isabelle/HOL," in *Integrated Formal Methods 2017*, 2017.
30. Xiao, W., Mehdipour, N., Collin, A., Bin-Nun, A., Frazzoli, E., Tebbens, R.D., and Belta, C., "Rule-Based Optimal Control for Autonomous Driving," in *12th ACM/IEEE international Conference on Cyber-Physical Systems*, Nashville, 2021.
31. Prakken, H., "On Making Autonomous Vehicles Respect Traffic Law: A Case Study for Dutch Law," in *Proceedings of the 16th Edition of the International Conference on Artificial Intelligence and Law*, 2017.
32. Censi, A., Slutsky, K., Wongpiromsarn, T., Yershov, D., Pendleton, S., Fu, J., and Frazzoli, E., "Liability, Ethics, and Culture-Aware Behavior Specification Using Rulebooks," in *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, 2019.

33. U.K. Law Commission and Scottish Law Commission, “Automated Vehicles Consultation Paper 3 - A Regulatory Framework for Automated Vehicles: A Joint Consultation Paper,” Crown, 2020.
34. Singapore Standards Council, “Auonomous Vehicles - Part 1: Basic Behavior,” Enterprise Singapore, Singapore, 2019.
35. Macpherson, M., “Performance Measurement in Not-for-Profit and Public-Sector Organisations,” *Measuring Business Excellence* 5(2):13-17, 2001.
36. Bin-Nun, A., Panasci, A., and Tebbens, R., “Heinrich’s Triangle, Heavy-Tailed Distributions, and Autonomous Vehicle Safety,” January 2020. [Online]. Available: https://annualmeeting.mytrb.org/OnlineProgramArchive/Details/13633. [Accessed 12 January 2021].
37. Liggett, R., Cooper, J., Huff, H., Taylor-Gratzer, R., Wong, N., Benitez, D., Douglas, T., Howe, J., Griswold, J., Amos, D., and Proulx, F., “Bicycle Crash Risk: How Does It Vary, and Why?,” UCLA Institute of Transportation Studies, Los Angeles, 2016.
38. Dingus, T.A., Guo, F., Lee, S., Antin, J., Perez, M., Buchanan-King, M., and Hankey, J., “Driver Crash Risk Factors and Prevalence Evaluation Using Naturalistic Driving Data,” *Proceedings of the National Academy of Sciences* 113(10):2636-2641, 2016.
39. Garber, N.J. and Gousios, S., “Relationship between Time to Collision Conflicts and Crashes on Interstate Highways Subjected to Truck Lane Restrictions,” in *Proceedings of the 88th Annual Meeting of the Transportation Research Board*, Washington, DC, 2009.
40. Mahmud, S.S., Ferreira, L., Hoque, M.S., and Hojati, A.T., “Application of Proximal Surrogate Indicators for Safety Evaluation: A Review of Recent Developments and Research Needs,” *International Association of Traffic and Safety Sciences Research* 41(4):153-163, 2017.
41. Insurance Institute for Highway Safety, Highway Loss Data Institute, “Real-World Benefits of Crash Avoidance Technologies,” June 2019. [Online]. Available: https://www.iihs.org/media/259e5bbd-f859-42a7-bd54-3888f7a2d3ef/e9boUQ/Topics/ADVANCED%20DRIVER%20ASSISTANCE/IIHS-real-world-CA-benefits.pdf. [Accessed 15 May 2020].
42. Leslie, A.J., Kiefer, R.J., Meitzner, M.R., and Flannagan, C.A., “Analysis of the Field Effectiveness of General Motors Production Active Safety and Advanced Headlight Systems,” University of Michigan Transportation Research Institute (UMTRI), Ann Arbor, 2019.
43. Lubbe, N. and Kullgren, A., “Assessment of Integrated Pedestrian Protection Systems with Forward Collision Warning and Automated Emergency Braking,” in *IRCOBI Conference 2015*, Lyon, France, 2015.
44. Gully, S.M., Whitney, D.J., and Vanosdall, F.E., “Prediction of Police Officers’ Traffic Accident Involvement Using Behavioral Observations,” *Accident Analysis and Prevention* 27(3):355-362, 1995.
45. Winlaw, M., Steiner, S.H., MacKay, R.J., and Hilal, A.R., “Using Telematics Data to Find Risky Driver Behaviour,” *Accident Analysis & Prevention* 131:131-136, 2019.
46. Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., and Ramsey, D.J., “Comparing Real-World Behaviors of Drivers with High versus Low Rates of Crashes and Near-Crashes,” NHTSA, Washington, DC, 2009.
47. Insurance Information Institute, “Facts + Statistics: Aggressive Driving,” 2009. [Online]. Available: https://www.iii.org/fact-statistic/facts-statistics-aggressive-driving. [Accessed 15 May 2020].
48. Allstate, “Allstate America’s Best Drivers Report 2019,” 2019. [Online]. Available: https://www.allstate.com/resources/allstate/attachments/best-drivers/2019bestdriversmediashareablebestdriversreport.pdf. [Accessed 15 May 2020].
49. Zendrive, “Preventing Collisions and Reducing Fleet Costs While Using the Zendrive Dashboard,” 2017. [Online]. Available: https://zendrive.com/wp-content/uploads/2019/04/Zendrive_White_Paper_Loss_Control.pdf. [Accessed 15 May 2020].
50. Wolshon, B., Parr, S.A., and Mousavi, S.M., “Identifying High-Risk Roadways for Infrastructure Investment Using Naturalistic Driving Data,” Maritime Transportation Research and Education Center, Fayetteville, AR, 2015.
51. Mousavi, S.M., Parr, S.A., Pande, A., Wolshon, B., and Zhang, Z., “Identifying High Crash Risk Highway Segments Using Jerk-Cluster Analysis,” in *International Conference on Transportation and Development 2019*, 2019.
52. Kuang, Y., Qu, X., and Etemad-Shahidi, A., *PLoS ONE* 10(9):1, 2015.

53. American Petroleum Institute, "Fact Sheet: ANSI/API Recommended Practice 754, Process Safety Performance Indicators for the Refining and Petrochemical Industries," API, Washington, DC, 2010.
54. Inouye, J., "Practical Guide to Leading Indicators: Metrics, Case Studies & Metrics," Campbell Institute, National Safety Council, 2015.
55. Center for Chemical Process Safety, "Process Safety Leading and Lagging Metrics," January 2011. [Online]. Available: https://www.aiche.org/sites/default/files/docs/pages/CCPS_ProcessSafety_Lagging_2011_2-24.pdf. [Accessed 15 May 2020].
56. NHTSA, "Description of Light-Vehicle Pre-Crash Scenarios for Safety Applications Based on Vehicle-to-Vehicle Communications," U.S. Department of Transportation, Washington, DC, 2013.
57. Rosén, E.A.O., and Sander, U., "Pedestrian Fatality Risk as a Function of Car Impact Speed," *Accident Analysis & Prevention* 41(3):536-542, 2009.
58. Kim, J., Kichun, J., Wonteak, L., Minchul, L., and Myoungho, S., "Curvilinear-Coordinate-Based Object and Situational Assessment for Highly Automated Vehciles," *IEEE Transactions on Intelligent Transportation Systems* 16(10.119/TITS.2014.236937):1-17, 2014.
59. MathWorks, "Coordinate Systems for Unreal Engine Simulation in Automated Driving Toolbox, R2020b Documentation," [Online]. Available: https://www.mathworks.com/help/driving/ug/coordinate-systems-for-3d-simulation-in-automated-driving-toolbox.html. [Accessed 15 August 2020].
60. NHTSA, "Forward Collision Warning System Confirmation Test," Office of Vehicle Safety, Office of Crash] Avoidance Standards, Washington, DC, 2013.
61. Junietz, P., Bonakdar, F., Klamann, B., and Winner, H., "Criticality Metric for the Safety Validation of Automated Driving Using Model Predictive Trajectory Optimization," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, 2018.
62. af Wåhlberg, A., "Long-Term Prediction of Traffic Accident Record from Bus Driver Celeration Behavior," *International Journal of Occupational Safety and Ergonomics: JOSE* 13(2):159-171, 2007.
63. Federal Aviation Administration, "Safety and Certification Approaches for Ethernet-Based Aviation Databuses," DOT/FAA/AR-05/52, Washington, DC, 2005.
64. EuroNCAP, "Test Protocol - AEB VRU Systems," February 2019. [Online]. Available: https://cdn.euroncap.com/media/43371/euro-ncap-aeb-vru-test-protocol-v301.pdf. [Accessed 15 May 2020].
65. World Forum for Harmonization of Vehicle Regulations (WP.29), "Proposal for a New UN Regulation on Uniform Provisions Concerning the Approval of Vehicles with Regards to Automated Lane Keeping System," June 2020. [Online]. Available: https://undocs.org/ECE/TRANS/WP.29/2020/81. [Accessed 9 November 2020].
66. Benmimoun, M., Pütz, A., Zlocki, A., and Eckstein, L., "euroFOT: Field Operational Test and Impact Assessment of Advanced Driver Assistance Systems: Final Results," in *Proceedings of the FISITA 2012 World Automotive Congress*, 2012.
67. Dingus, T.A., Hankey, J.M., Antin, J.F., Lee, S.E., Elchelberger, L., Stulce, K.E., McGraw, D., Perex, M., and Stowe, L., "Naturalistic Driving Study: Technical Coordination and Quality Control," Transportation Researcch Board, Washington, DC, 2015.
68. Makowsky, M.D.S.T., "More Tickets, Fewer Accidents: How Cash-Strapped Towns Make for Safer Roads," *The Journal of Law and Economics* 54(4):863-888, 2011.
69. Levy, D.T., "Youth and Traffic Safety: The Effects of Driving Age, Experience, and Education," *Accident Analysis and Prevention* 22:327-334, 1990.
70. Arai, Y., Nishimoto, T., Ezaka, Y. and Yoshimoto, K., "Accidents and Near-Misses Analysis by Using Video Drive-Recorders in a Fleet Test," in *17th Internatinal Technical Conference on the Enhanced Safety Vehicles*, Amsterdam, 2015.
71. Rösener, C., Harth, M., Josten, J., Weber, H., and Eckstein, L., "Modelling Human Driver Performance for Safety Assessment of Road Vehicle Automation," in *21st International Conference on Intelligent Transportation Systems (ITSC)*, Hawaii, 2018.
72. Kubiak, T.M., and Benbow, D.W., *The Certified Six Sigma Black Belt Handbook* Second Edition (2009).
73. Dodge, Y., *The Oxford Dictionary of Statistical Terms* (Oxford University Press, The International Statistical Institute, 2003).

74. Federal Motor Carrier Safety Administration, "CSA: Introduciton to the Safety Measurement System Version 3.0," March 2013. [Online]. Available: https://csa.fmcsa.dot.gov/Documents/IntroductiontoSMS3_0.pdf. [Accessed 15 May 2020].
75. Brorsson, B., Rydgren, H., and Ifver, J., "Single-Vehicle Accidents in Sweden: A Comparative Study of Risk and Riskfactors by Age," *Journal of Safety Research* 24(1):55-65, 1993.
76. Blanco, M., Atwood, J., Russell, S., Trimble, T., McClafferty, J., and Perez, M., "Automated Vehicle Crash Rate Comparison Using Naturalistic Data," Virginia Tech Transportation Institute, 2016.
77. B. o. T. Statistics, "Bureau of Transportation Statistics: Highway," 2020. [Online]. Available: https://www.bts.gov/topics/highway. [Accessed 15 June 2020].
78. Kalra, N., and Paddock, S., *Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?* (RAND Corporation, 2016).
79. NHTSA, "Traffic Safety Facts 2017 Data, Rural/Urban Comparison of Traffic Fatalities DOT HS 812 741," U.S. Department of Transportation, Washington, DC, 2019.
80. U. Laboratories, "UL 4600 Proposed First Edition of the Standard for Safety for the Evaluation of Autonomous Products," Underwriters Laboratories, LLC, October 2, 2019.
81. NHTSA, "2018 Fatal Motor Vehicle Crashes," DOT HS 812 826, 2019.
82. Rich, K., Blanchard, H., and McCloskey, J., "The Use of Goal Structuring Notation as a Method for Ensuring that Human Factors Is Represented in a Safety Case," in *2007 2nd Institution of Engineering and Technology International Conference on System Safety*, 2007.
83. ACWG, The Assurance Case Working Group, Goal Structuring Notation Community Standard Version 2, 2018.
84. NHTSA, "Traffic Safety Facts," April 2020. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812932.
85. Archer, J., "Indicators for Traffic Safety Assessment and Prediction and Their Application in Micro-Simulation Modelling: A Study of Urban and Suburban Intersections," in Royal Institute of Technology, Sweden, 2005.
86. National Conference of State Legislatures (NCSL), "Safely Passing Bicyclists Chart," National Conference of State Legislatures, 28 April 2020. [Online]. Available: http://www.ncsl.org/research/transportation/safely-passing-bicyclists.aspx. [Accessed 28 May 2020].
87. Porter, R., Himes, S., Musunuru, A., Le, T., Eccles, K., Peach, K., Tasic, I., Zlatkovic, M., Tatineni, K., and Duffy, B., "Understanding the Causative, Precipitating, and Predisposing Factors in Rural Two-Lane Crashes," Federal Highway Administration, McLean VA, 2018.
88. Michon, J., "A Critical View of Driver Behavior Models: What Do We Know, What Should We Do?," in *Human Behavior and Traffic Safety*, Evans, L. and Schwing, R.C. (Eds.), Plenum Press, 1985, 485-520.
89. Tim Kelly, R.W., Kelly, T., and Weaver, R., "The Goal Structuring Notation - A Safety Argument Notation," 2004.
90. Pegasus Projekt, "PEGASUS Symposium 2019," 14 May 2019. [Online]. Available: https://www.pegasusprojekt.de/files/tmpl/Pegasus-Abschlussveranstaltung/08_Criticality_Metric.pdf. [Accessed 15 June 2020].
91. Pegasus Projekt, "PEGASUS Symposium 2019," 14 May 2019. [Online]. Available: https://www.pegasusprojekt.de/files/tmpl/PDF-Symposium/04_Scenario-Description.pdf. [Accessed 15 June 2020].

Appendix C. Safety Concepts and Metrics

There are a number of proposed ways to measure or promote ADS safety performance, with new recommendations continually promoted as experience is gained, new research published and as technology matures. The metrics and safety concepts below in Tables A.1 and A.2 contributed to the development of the criteria for DDT-Relevant System Level Metrics (4.0).

TABLE A.1 Safety-related metrics reviewed during the development of *AVSC Best Practice for Metrics and Methods for Assessing Safety Performance of Automated Driving Systems (ADS)*

Metrics	Source	Description
Criticality metric	Junietz, P., Bonkadar, F., Klamann, B., and Winner, H. 2018. "Criticality Metric for the Safety Validation of Automated Driving using Model Predictive Trajectory Optimization." 21st International Conference on Intelligent Transportation Systems (ITSC). https://ieeexplore.ieee.org/document/8569326	The criticality metric, developed as part of the PEGASUS project, uses an algorithm similar to those used by vehicle path planning algorithms to find an optimal trajectory. The computation of the proposed metric uses elements of model predictive control using an objective function that contains four elements that describe the difficulty of the driving task. The severity of an imminent collision may be predicted at the critical point.
Disengagements	California Title 13, Division 1, Chapter 1, Article 3.7 - Testing of Autonomous Vehicles https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/ https://www.dmv.ca.gov/portal/uploads/2020/06/Adopted-Regulatory-Text-2019-1.pdf	The state of California requires manufacturers who conduct ADS testing in the state to report on all unscheduled disengagements of the ADS that occur during testing. The protocols for disengagements may vary among ADS developers. Not all disengagements may result from safety-relevant events.
Minimum Following Distance and Lane Intrusion	UNECE ALKS, April 2020. "Proposal for a new UN Regulation on uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping System" https://undocs.org/ECE/TRANS/WP.29/2020/81	Specific embodiments of safety envelope metrics used by UNECE to evaluate Automated Lane Keeping Assist features. Minimum following distance is calculated as the current velocity of the ADS vehicle multiplied by the minimum time gap in seconds between the ADS vehicle and a leading vehicle in front, as summarized in a table. The minimum lane intrusion is a TTC value determined based on when the outside of the tire of the intruding vehicle's front wheel closest to the lane markings crosses a line 0.3 meters beyond the outside edge of the visible lane marking to which the intruding vehicle is being drifted, captured by the equation : $TTC\ Lane\ Intrusion > vrel/(2 \cdot 6 m/s^2) + 0.35$*seconds, where *vrel* is the relative velocity between both vehicles, positive for vehicle being faster than the cutting in vehicle.
Model predictive instantaneous safety metric (MPrISM)	Weng, B., Rao, S., Deosthale, E., Schnelle, S., Barickman, F. 2020. "Model Predictive Instantaneous Safety Metric for Evaluation of Automated Driving Systems." https://arxiv.org/abs/2005.09999	MPrISM determines the safety status of an ADS-DV considering a worst-case scenario for a given traffic scene. This metric determines the time to collision of an ADS-DV under the worst-case scenario by considering the various decisions and actions for the ADS and evaluating them in simulation. The algorithm provides an ultimate score for the ADS depending on the results using a methodic algorithm and has also been utilized in real world tests. Although the metric was derived assuming deterministic behavior of a vehicle's dynamic control and actuation constraints, the algorithm can be extended to use in probability and risk assessments.
Post-encroachment time	Peesapati, L. N., Hunter, M. P., and Rodgers, M. O. 2018. "Can post encroachment time substitute intersection characteristics in crash prediction models?" *Journal of Safety Research 66, pp.* 205-211.	Post encroachment time a measure that represents the time difference between a vehicle leaving the area of encroachment and a conflicting vehicle entering the same area. For example, a scenario in which post-encroachment time offers insights into potential safety-relevant events is a left-turning vehicle at an intersection is crossing in front of the path of opposing through vehicles.

TABLE A.2 Safety-related concepts reviewed during the development of *AVSC Best Practice for Metrics and Methods for Assessing Safety Performance of Automated Driving Systems (ADS)*

Concepts	Source	Description
Adherence to system policies	ANSI/UL 4600: Standard for Evaluation of Autonomous Produces https://www.shopulstandards.com/ProductDetail.aspx?productid=UL4600	This standard covers the safety principles, risk mitigation, tools, techniques, and lifecycle processes for building and evaluating a safety argument for vehicles that can operate in an autonomous mode." An ADS should adhere to system developer set policies, assumptions, design goals, and conclusions, such as speed limitation, lane boundary departures, offsets in lane, and road departures. The approach to measuring compliance with these policies is left up to ADS developers and manufacturers.
Culture	ANSI/UL 4600: Standard for Evaluation of Autonomous Produces https://www.shopulstandards.com/ProductDetail.aspx?productid=UL4600	Safety culture is an organizational approach to improve safety through processes, policies, mitigations, and trainings. Some example elements of safety culture are provided below for reference: • Process adherence • Training and skill validation • Fraction of field identified defects traced back to deviations from development and validation processes
Drivability	Gil Gómez, G.L., Nybacka, M., Bakker, E., and Drugge, L. 2016. "Objective metrics for vehicle handling and steering and their correlations with subjective assessments." *International Journal of Automotive Technology, Vol. 17,* No. 5, pp. 777-794.	Generally speaking, drivability is defined as the smoothness of vehicle's motion control under all operating conditions. This pertains to vehicle maneuvers, such as speed maintenance, stop and go braking, lane centering, and weaving. Drivability most directly relates to passenger satisfaction and is only tangentially related to safety.
Dynamic driving task (DDT) fallback executed correctly		The ability of the ADS to achieve a minimal risk condition - 1) after occurrence of a DDT performance-relevant system failure(s), or 2) upon operational design domain (ODD) exit.
Extreme value theorem, e.g., criticality metric, instantaneous metric (ISM)	Farah, H. and Azevedo, C. L. 2017. "Safety analysis of passing maneuvers using extreme value theory." *International Association of Traffic and Safety Sciences Research, Volume 41, Issue 1,* pp. 12-21. https://www.sciencedirect.com/journal/iatss-research/vol/41/issue/1 Every, J., Barickman, F., Martin, J., Rao, S., Schnelle, S., and Weng, B. 2017. "A Novel Method to Evaluate the Safety of Highly Automated Vehicles." *25th International Technical Conference on Enhanced Safety of Vehicles.* NHTSA. https://trid.trb.org/view/1485370	Extreme value theorem states that under certain cases there is a critical point (extreme value) of a function across a closed interval. When that function is representative of driving risk, then the extreme value can be used to produce a safety metric. There are many proposed safety metrics that fall within the construct of the extreme value theorem, such as the criticality metric and the model predictive instantaneous safety metric (MPrISM). The criticality metric and ISM formulate the reachability of a safe state based on kinematics that contemplate a wide range of possible paths for the subject vehicle. To find the most critical state within the range of potential future states, these approaches require optimization of a multidimensional state space.
Formal safety models, e.g., Responsibility Sensitive Safety (RSS), safety force field (SFF), IEEE P2846 Assumptions for Models in Safety-Related Automated Vehicle Behavior	Shalev-Shwartz, S., Shammah, S., and Shashua, A. 2017. "On a Formal Model of Safe and Scalable Self-driving Cars." https://arxiv.org/abs/1708.06374 Nister, D., Lee, H. L., Ng, J., and Wang, Y. 2019. *The Safety Force Field.* NVIDIA https://www.nvidia.com/content/dam/en-zz/Solutions/self-driving-cars/safety-force-field/the-safety-force-field.pdf IEEE WG: VT/ITS/AV Decision Making P2846 https://sagroups.ieee.org/2846/	Responsibility Sensitive Safety (RSS) and safety force field (SFF) provide frameworks for formalizing safety models and assumptions. RSS and SFF seek to formalize assumptions about road users, which for example may be expressed as a safety envelope metric and threshold. For example, RSS suggests a minimal safe longitudinal distance can be established based on kinematic equations which can be expressed as a longitudinal time to collision (TTC) safety envelope violation. IEEE P2846 Assumptions for Models in Safety-Related Automated Vehicle Behavior. IEEE committee developing standard describing the minimum set of reasonable assumptions used in foreseeable scenarios to be considered for road vehicles in the development of safety-related models that are part of automated driving systems (ADS). The standard includes consideration of rules of the road and their regional and/or temporal dependencies.

TABLE A.2 (Continued) Safety-related concepts reviewed during the development of *AVSC Best Practice for Metrics and Methods for Assessing Safety Performance of Automated Driving Systems (ADS)*

Concepts	Source	Description
Near Miss	Hayward, C., 1972. "Near-Miss Determination Through Use of a Scale of Danger." http://onlinepubs.trb.org/Onlinepubs/hrr/1972/384/384-004.pdf	Near miss generally captures an event where a collision is narrowly averted, but there is no universally accepted quantitative definition for this term. ADS developers and manufacturers could capture "near-miss" data using the metrics identified in this best practice (6.0), as defined for the specific ODD in which the ADS is operating.
Reduction in simulated collisions [for disengagements]	Gaspar, J. G. and Noyce, D. A. 2017. "Using Naturalistic Data to Develop Simulator Scenarios." A Report on Research Sponsored by SAFER-Sim. http://safersim.nads-sc.uiowa.edu/final_reports/UWI-1-Y2_Report.pdf	In the case of a safety-critical disengagement, ADS developers and manufacturers may be able to simulate the outcome of the situation had the disengagement not occurred. Of particular interest is whether a crash would have resulted if the in-vehicle fallback test driver did not take over control. Should simulations be conducted, one potential threshold to determine whether a collision would have occurred is a safety envelope violation. The data from on-board systems can be used to re-create the scenario in a simulated environment to understand how the ADS would have responded and predict whether any collisions may have potentially occurred.
Reduction in system errors	ANSI/UL 4600: Standard for Evaluation of Autonomous Products https://www.shopulstandards.com/ProductDetail.aspx?productid=UL4600	ADS developers and manufacturers continually seek to reduce system errors as part of system engineering processes. System errors may be monitored using the metrics defined in this best practice. Some additional examples of system error reduction concepts are: • Data quality for sensing may have accuracy, precision, and resolution requirements that are monitored. • Object classification errors (e.g., misclassification rates for classification algorithms, false positive and false negative rates, prediction error rates, correlated fault and failure rates, repeatability). • Lifecycle metrics (e.g., post-deployment safety related software defect rates, field failure rates of safety related preventive and periodic maintenance items, field failure rates of items that are subject to inspection). • Performance and quality measures (e.g., software execution fault rate, real time performance fault rates). • Hazard rate occurrence, including mitigated and unmitigated or partially mitigated hazards. • Detected hardware and software component failures. • Mission success rate, including successful and unsuccessful activations of failsafes and instances of DDT fallback performance. • Rate of encountering surprises (e.g., detection of work zones, temporary lane shifts, and new situations not included in training data). • Rate of rework to safety related systems. • Model-based anomaly detection (e.g., violating controllability, maneuverability, situational awareness models or object models). • Error detection and mitigation approaches vary by vehicle architecture and may be assessed at a subsystem level.
Right-of-way	Shalev-Shwartz, S., Shammah, S., and Shashua, A. 2017. "On a Formal Model of Safe and Scalable Self-driving Cars." https://arxiv.org/abs/1708.06374 Nister, D., Lee, H. L., Ng, J., and Wang, Y. 2019. *The Safety Force Field.* NVIDIA https://www.nvidia.com/content/dam/en-zz/Solutions/self-driving-cars/safety-force-field/the-safety-force-field.pdf	Right-of-way generally describes yielding behavior between road users. ADS should obey the universal norm that the right-of-way is to be given, not taken. Initial efforts to define this concept include the Responsibility Sensitive Safety and safety force field. Driving norms must first be described in closed-form solutions that permit quantitative metrics.

TABLE A.2 (Continued) Safety-related concepts reviewed during the development of *AVSC Best Practice for Metrics and Methods for Assessing Safety Performance of Automated Driving Systems (ADS)*

Concepts	Source	Description
Roadmanship	Fraade-Blanar, L., Blumenthal, M.S., Anderson, J.M. and Kalra, N. 2018. "*Measuring Automated Vehicle Safety: Forging a framework.*" RAND Corporation. https://www.rand.org/pubs/research_reports/RR2662.html	"Roadmanship" captures the ability to drive on the road safely without creating hazards and responding well (regardless of legality) to the hazards created by others. The concept centers on whether the vehicle "plays well with others," even if others are not around.
Safety envelope variability	Innamaa, Satu; Kuisma, Salla, "Key performance indicators for assessing the impacts of automation in road transportation Results of the Trilateral key performance indicator survey," Research Report : VTT-R-01054-18, 2018 https://cris.vtt.fi/en/publications/key-performance-indicators-for-assessing-the-impacts-of-automatio	A measure of the dispersion in safety envelope values over time, i.e., high variability indicates a lack of consistency. For example, this could be in terms of the time-headway to the vehicle in front in car following situations.
Scenario completeness	Binyamini, Z. 2020. "What counts as a valid measurement in autonomous vehicle development?" *Automotive Testing Technology International.* https://www.automotivetestingtechnologyinternational.com/industry-opinion/what-counts-as-a-valid-measurement-in-autonomous-vehicle-development.html	Completeness is a metric that describes the degree to which one set ("A") covers another set ("B"). This may also be referred to as coverage. Black box testing for completeness may be described in terms of requirements coverage [80]. Driving scenarios provide a measure of requirements coverage. For example, set "A" may be the scenarios driven by an ADS-DV fleet, and set "B" is the theoretical total possible scenarios in the ODD. While the total permutations of scenarios within a given ODD may be infinite, scenarios can be organized into a finite number of scenario families. These scenario families may be described by characteristics, such as environment, roadways, maneuvers, and other road users. If a scenario is defined by its (a) stage, (b) actors (ADS and others), and (c) blocking, then stage and ADS actor and intended ADS blocking is static for a scenario, the quantity and blocking of other actors that interfere with ADS blocking is what makes scenarios differ.
System latency	ANSI/UL 4600: Standard for Evaluation of Autonomous Products https://www.shopulstandards.com/ProductDetail.aspx?productid=UL4600	Latency is the time it takes to process a unit of work. Similar to human reaction speed, machines take time to detect and respond. For example, an ADS may respond slower when processing a situation with a multitude of pedestrians that require processing time to detect. Latency may be measured many ways. For example, this may be measured as the amount of time between an object coming within sensor range and identification, classification, response planning, and response executed by the vehicle system. A system level measure of latency is described by Object and Event Detection and Response (OEDR) time. Subsystem level latency metrics are part of system engineering processes.

Appendix D. Safety Metric Data Sources and Reference Materials

Many data sources are available for potential use in ADS safety assessments. It is likely that multiple data sources will be needed to provide a robust measure of ADS safety performance and its impact on societal outcomes. It should be noted that it will remain a challenge to mine data that is specific to a particular ADS-DV's ODD and that developers will therefore need to continue to seek out new data sources and analysis techniques that support metric validity.

Tables A.3 and A.4 list examples of data sources and reference materials relevant to the set of metrics recommended in this best practice. This is not an exhaustive list.

TABLE A.3 Data sources and their relevance to AVSC Best Practices

Data Source	Description	Relevance to Metrics in AVSC Best Practices
The 100-Car Naturalistic Driving Study Data https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/100carmain.pdf	The 100-Car Naturalistic Driving Study is the first instrumented-vehicle study undertaken with the primary purpose of collecting large-scale, naturalistic driving data.	Relative positions of objects, acceleration, jerk, time-to-collision front and rear, crash factors
Bureau of Transportation Statistics (BTS) https://www.bts.gov/topics/highway	Databases, including National Transportation Atlas Database – Highway, and seasonally -adjusted vehicle miles traveled data.	Crash frequency and severity, exposure, vehicle miles traveled
Crash Investigation Sampling System (CISS) https://www.nhtsa.gov/crash-data-systems/crash-investigation-sampling-system	NHTSA's crash data collection program consists of the Crash Investigation Sampling System (CISS), the Fatality Analysis Reporting System (FARS), the Crash Report Sampling System (CRSS), Special Crash Investigations (SCI), Non-Traffic Surveillance, the Crash Injury Research and Engineering Network, and special studies conducted to address various safety topics. The CISS builds on the retiring, long running National Automotive Sampling System Crashworthiness Data System; maintained by NHTSA.	Crash frequency and severity; includes EDR data that can be used to determine acceleration and jerk
Crash Report Sampling System (CRSS) https://www.nhtsa.gov/crash-data-systems/crash-report-sampling-system	Sample of reported crashes involving all types of motor vehicles, pedestrians, and cyclists, ranging from property-damage-only crashes to those that result in fatalities. CRSS replaces the National Automotive Sampling System General Estimates System; maintained by NHTSA.	Crash frequency and severity, legally cited violations, crash factors
Fatality Analysis Reporting System (FARS) https://www.nhtsa.gov/crash-data-systems/fatality-analysis-reporting-system	Contains annual data on the census of fatal traffic crashes (U.S.) within the 50 states, the District of Columbia, and Puerto Rico. To be included in FARS, a crash must involve a motor vehicle traveling on a traffic way customarily open to the public and must result in the death of a vehicle occupant or a nonoccupant within 30 days of the crash.	Crash occurrence (fatality only), legally cited violations, crash factors
National Center for Statistics and Analysis (NCSA) https://www.nhtsa.gov/research-data/national-center-statistics-and-analysis-ncsa	Provides a wide range of analytical and statistical support to NHTSA and the highway safety community, including areas of human, vehicle, environmental, and roadway characteristics, as they relate to crash frequency and injuries, effectiveness of crash avoidance efforts, and quantifying the benefits resulting from proposed agency rules.	Crash frequency and severity
National Motor Vehicle Crash Causation Survey (NMVCCS) https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812506	Conducted from 2005 to 2007, was aimed at collecting on-scene information about the events and associated factors leading up to crashes involving light vehicles.	Crash frequency, legally cited violations, crash factors
Highway Safety information System (HSIS) https://www.hsisinfo.org/	Multistate database that contains crash, roadway inventory, and traffic volume data for a select group of states (U.S.A).	Crash frequency and severity, crash factors
Special Crash Investigation (SCI) https://www.nhtsa.gov/research-data/special-crash-investigations-sci	Data contained in routine police and insurance crash reports to comprehensive data from special reports by professional crash investigation teams. Hundreds of data elements relevant to the vehicle, occupants, injury mechanisms, roadway, and safety systems are collected for each of the over 100 crashes designated for study annually.	Crash factors, including severity (delta-V), acceleration, jerk, relative positions of objects, traffic control device state

TABLE A.4 Reference Materials and their relevance to AVSC Best Practices

Reference Materials	Description	Relevance to Metrics in AVSC Best Practices
Automated Driving System Data Logger SAE J3197_202004 https://saemobilus.sae.org/content/J3197_202004/	SAE Recommended Practice that defines data elements to supplement the J1698-1 defined EDR in order to facilitate the determination of the background and events leading up to a collision in an ADS operated-vehicle.	Crash frequency and severity, acceleration, jerk
Event Data Recorder (EDR) J1698_201703 https://www.sae.org/standards/content/j1698_201703/	SAE Recommended Practice describes common definitions and operational elements of Event Data Recorders. Series of documents: • SAE J1698-1 - Event Data Recorder - Output Data Definition • SAE J1698-2 - Event Data Recorder - Retrieval Tool Protocol • SAE J1698-3 - Event Data Recorder - Compliance Assessment	Crash frequency and severity, acceleration, jerk
Integrated Vehicle-Based Safety Systems (IVBSS) program https://www.its.dot.gov/research_archives/ivbss/index.htm https://rosap.ntl.bts.gov/view/dot/3358	Presents results from the light-vehicle and heavy-truck field operational tests performed as part of the Integrated Vehicle-Based Safety Systems (IVBSS) program. The findings are the result of analyses performed by the University of Michigan Transportation Research Institute to examine the effect of a prototype integrated crash warning system on driver behavior and driver acceptance.	Safety envelope, OEDR reaction time
On-board ADS Data Logger AVSC Best Practice for Data Collection for ADS-DVs to Support Event Analysis (AVSC00004202009) https://avsc.sae-itc.org/#roadmap SAE J3197_202004 https://www.sae.org/standards/content/j3197_202004/	SAE Recommended Practices and AVSC best practices provide common data output formats and definitions for a variety of data elements that may be useful for analyzing the performance of an automated driving system (ADS).	Presence of crash event; object and event data for citable offense, TTC, and OEDR response time determination; acceleration and jerk.
Safety Pilot https://www.its.dot.gov/factsheets/pdf/SafetyPilotModelDeployment.pdf	Research initiative featuring real-world implementation of connected vehicle safety technologies, applications, and systems in everyday vehicles and multimodal driving conditions.	Acceleration, jerk
Second Strategic Highway Research Program 2 (SHRP 2) http://www.trb.org/Publications/PubsSHRP2Publications.aspx http://www.trb.org/Publications/Blurbs/170935.aspx	TRB's second Strategic Highway Research Program (SHRP 2) Report S2-S06-RW-1: Naturalistic Driving Study: Technical Coordination and Quality Control documents the coordination and oversight of participant- and vehicle-based operations for an in-vehicle driving behavior field study collected from naturalistic driving data and associated participant, vehicle, and crash-related data.	Relative positions of objects, acceleration, jerk, crash factors