A METHOD TO COMPARE THE SAFETY OF AUTONOMOUS VEHICLES TO HUMAN-DRIVEN NON-EXEMPT MOTOR VEHICLES IN THE UNITED STATES

Lisa Gavin
U.S. Department of Transportation
United States of America

Paper Number 19-0145

ABSTRACT

The objective of this technical paper is to present a method that characterizes autonomous vehicle (AV) safety performance through the application of risk-based validation that leverages existing crash incidence and severity data, physics based model and simulation, and U.S. Federal Motor Vehicle Safety Standard (FMVSS) benchmark metrics. The output of the proposed risk-based methodology is a framework that organizes the number and type of physical tests and model/simulation runs necessary to provide meaningful evidence of AV safety performance statistically equivalent to human-driven non-exempt motor vehicles.

INTRODUCTION

The lure of AVs promises elimination of vehicle crashes, injuries and fatalities. For consumers, passengers, and other road users to embrace AV technology, AVs must perform safely and reliably. A formidable challenge is measuring and quantifying the levels of safety offered by AVs. The debate of how safe is safe enough for AVs has been structured around a mindset of billions of miles traveled.

Common sense dictates the starting benchmark is the current level of safety performance; but how best is this measured? Traditionally, mechanical and physical safety have been measured in terms of compliance with government safety standards. AVs venture beyond this template in that the vehicles are loaded with complex sensor technologies and controlled by software. This paper proposes application of a risk-based methodology that leverages existing knowledge of vehicle performance characteristics and crash problem data with safety metrics to build a framework that compares an overall safety level between nonexempt vehicles and AVs.

Foundational Concepts

Defining AV safety metrics that are recognized and accepted industry wide by all stakeholders is an important first step. Four foundational concepts are in play: AVs must provide an overall safety level at least equal to the overall safety level of nonexempt motor vehicles; the Haddon Matrix, which is the most commonly used paradigm in the injury prevention field; all motor vehicles, including AVs, present as a system of systems; and safety is not reliability.

When determining if a vehicle presents an unreasonable risk to safety, probability of failure, consequence of failure, occurrence and severity of injury are the primary factors to consider. In a probabilistic risk assessment, there is a close relationship between safety and reliability. Yet, safety cannot generally be achieved through component or system reliability alone. The Federal Highway Administration (FHWA) version of the Haddon Matrix [3] illustrates the relationship between four factors of injury (human, vehicle/equipment, physical environment, and socioeconomic) and the phase of injury (pre-crash, crash, and post-crash).

If we consider the vehicle as a system of systems, one system would be the ‘driver,’ either in human form or in digital technology form. Vehicle systems such as powertrain, steering, braking, suspension, tires, fuel, occupant protection, and exterior lighting exist on all vehicles, whether AV or human driven. The primary differences between AVs and nonexempt vehicles will likely emerge in the driver system, in visibility and glazing systems, and interior human-machine interface (HMI) systems.

A Safety Network can be defined as shown in Equation 1.

$$Safety\ Network = Environment + Vehicle + Driver + Unknowns \qquad \text{(Equation 1)}$$

Reliability Concepts

Reliability is considered the absence of failures, and is predicated on how failure is defined. In the context of motor vehicle safety, we can describe a failure rate as both the crash rate and a function of system performance. Adopting the Advanced Product Quality Planning (APQP) and Control Plan manual definitions for reliability and confidence level supports analysis using key risk metrics. Reliability is defined as the probability that an item (i.e., vehicle) will continue to function at customer (i.e., roadway user) expectation levels at a measurement point, under specified environmental and duty cycle conditions. Confidence level refers to the percentage of all possible samples that can be expected to include the true population parameter. Additional reliability concepts include selecting an appropriate reliability distribution, sufficient sample size, and consideration of non-critical failures in the reliability analysis.

Several reliability distributions [7] appear to mirror the crash problem data, such as the binomial distribution, the exponential distribution, the Poisson distribution, and the Pareto distribution. Additionally, the bathtub distribution holds true for AVs in that sensor and camera initialization increases crash risk at the beginning of vehicle deployment. If certain crash avoidance data is collected from AVs, the normal and logistic continuous distributions offer the possibility to include the rate of crash events avoided, plotted as negative severity values. For any reliability distribution, key parameters (e.g., shape, scale, location) need to be confirmed. The distribution most appropriate for AVs may depend on the sample. In this paper, the sample was organized by vehicle classification, and the binomial distribution (with the assumption of replacement) was applied to count the number of successes (i.e., no crash) in a number of independent trials (i.e., VMT); if a crash occurs, then severity outcome is measured as no damage, property damage, injury, or fatality. The number of observations or trials must be sufficiently large.

Crash Problem

Utilizing the U.S. Department of Transportation data collections, databases and published statistical analysis, the crash problem on U.S. roadways in 2016 [4] was reported as 34,439 fatal police-reported crashes, 2,177,000 injury police reported crashes, and 5,065,000 property damage only (PDO) police-reported crashes, 37,461 people killed, 3,144,000 people injured, and 3,174 billion vehicle miles traveled. Additional crashes occurred that were not reported to the police; in 2010, the National Highway Traffic Safety Administration (NHTSA) estimated these unreported crashes as a 59.7% increase in PDO and a 39.7% increase in injury crashes [2]. Figure 1 shows a plot of the 2010 crash incidence versus severity. Ideally, future safety network analyses would be founded on the combined number of police-reported plus unreported crashes.

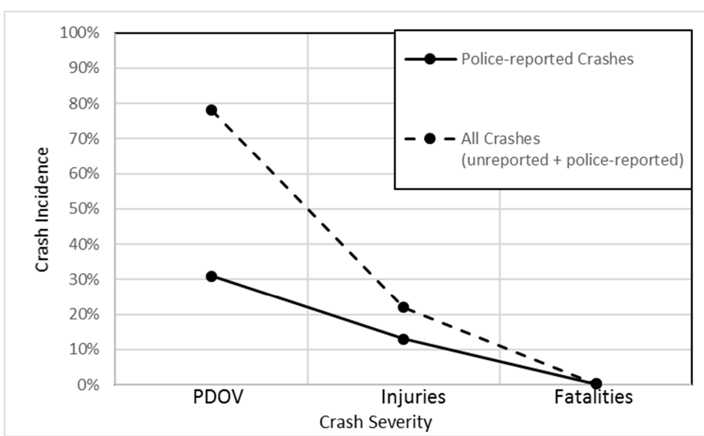

Figure 1. Crash Incidence versus Severity (2010 data; source: DOT HS 812 013)

Table 1 lists the police-reported crash incidence rates using 2016 data. Note that the crash per VMT rate for passenger cars, light trucks and buses are similar in magnitude as the total average. The police-reported crash incidence rate for large trucks is significantly lower, possibly due to these vehicles being operated by trained professional drivers.

Table 1.
Crash Incidence by Vehicle Classification (2016 data; sources: DOT HS 812 580 and [10])

Vehicle Classification	Police-reported Crashes (Fatal + Injury + PDO)	VMT (millions)	1 crash per (VMT)
Passenger Car	7,198,839	1,440,228	200,064
Light Truck	5,010,069	1,409,490	281,331
Large Truck	502,213	287,895	573,253
Motorcycle	129,421	20,445	157,973
Bus	71,227	16,350	229,548
Other/unknown	21,462	Not reported	-
	Total = 12,933,231	Total = 3,174,408	Total = 245,446

In 2015, NHTSA published findings from a statistical analysis of the National Motor Vehicle Crash Causation Survey (NMVCCS) [5], which collected on-scene information about the events and associated factors leading up to crashes involving light vehicles. NMVCCS is a weighted sample of 5,470 crashes, which represents an estimated 2,189,000 crashes nationwide. NHTSA found that the critical reason, which is the last event in the crash causal chain, was assigned to the driver in 94 percent (±2.2%) of the crashes. In about 2 percent (±0.7%) of the crashes, NHTSA found that the critical reason was assigned to a vehicle component's failure or degradation, and in 2 percent (±1.3%) of crashes, it was attributed to the environment (slick roads, weather, etc.). Among an estimated 2,046,000 drivers who were assigned critical reasons, NHTSA found recognition errors accounted for about 41 percent (±2.1%), decision errors 33 percent (±3.7%), and performance errors 11 percent (±2.7%) of the crashes.

METHODOLOGY

A technology-neutral approach to AVs would focus on safety aspects and system safety performance. Shifting the mindset to a system of systems construct with an emphasis on test and evaluation supports quantifying safety in terms of risk and performance. A test and evaluation strategy would include physical testing, modeling, simulation, verification, validation, and accreditation. Stakeholders would use this methodology to generate a sample size of test scenarios to which manufacturers would demonstrate the level of safety. Large statistical sample sizes will only be achieved through physical testing and modeling/simulation. The benchmark would be comprised of a combination of test trials plus simulation runs that vary key performance factors. Reliability theories were developed for aircraft components under a metric of flight time hours and are adapted here to VMT. This method describes how to statistically estimate the level of AV safety without billions of on-road demonstration miles. There exists tremendous opportunity to leverage modeling and simulation along with targeted testing to characterize AV safety performance in terms of a reliability distribution. Modeling and simulation supports enhancing the physical test scenarios through iterations that vary speed ranges, travel direction, traffic density, etc.

Key steps in this analysis are:

- Identify the most relevant set of risk metrics. For example, incidence (number of crashes, severity in terms of fatalities, injuries, and property damage), and vehicle miles traveled (VMT).
- Identify the data needed to support a risk-based analysis. For example, 2,967 billion VMT (2010 data) divided by 6,077,362 police-reported crashes (2010 data) results in 1 crash per 488,205 VMT. The average of 11,866 VMT per registered vehicle (2010 data) multiplied by an estimated average vehicle age of 10.8 years (passenger car vehicles, 2010 data) results in an available test time T_{test} of 128,153.
- Adapt the parametric binomial reliability distribution test by replacing the random variable of Time with vehicle miles traveled (VMT).
- Select the % reliability to be demonstrated. For example, '85% reliable.'
- Select the % confidence level. For example, 'with 90% confidence.'

- Select the number of test failures that can occur in the sample. For example, '1 failure allowed.'
- Calculate the sample size based on VMT and T_{test} to which AV manufactures would demonstrate the AV level of safety.
- Choose trials (e.g., tests and model/simulation runs) that characterize vehicle performance in steering, accelerating, braking, sensor recognition, causes of vehicle control loss, visibility, etc.

In this analysis, the independent test trials correspond to VMT, regardless of vehicle maneuver, speed, etc. An example of the typical resulting output is in the form: a sample size of 11 pedestrian detection system test trials with 0 failures occurring will demonstrate a reliability of 80% at the 90% confidence level. In other words, if the item reliability is < 80%, the chances of passing this test are < 10%.

Test Sample Size

Table 2 is a representative test sample size matrix which was populated by exercising a parametric binomial reliability demonstration test calculator [8] with mission time equal to 1 crash per 488,205 VMT (based on 2010 data) and the available test time equal to 128,153 hours (based on 2010 data). Setting the reliability and confidence levels is a subjective decision. If the current level of safety for nonexempt passenger cars and light trucks is estimated at 85% reliability with 95% confidence, then a test series for equivalent AV safety performance would require a sample size of 425 tests that allows one failure. An alternate approach is to conduct testing until one failure occurs, and then estimate the reliability and confidence level.

Table 2.
Example of a Test Sample Size Calculated using the Parametric Binomial Reliability Distribution

Passenger Cars & Light Trucks	Confidence with 1 Failure Allowed					
Reliability	**80%**	**85%**	**90%**	**95%**	**98%**	**99%**
99%	4,325	4,871	5,618	6,851	8,425	9,587
98%	2,152	2,423	2,795	3,409	4,192	4,770
95%	848	955	1,101	1,343	1,652	1,879
90%	413	465	537	654	804	915
85%	268	302	348	425	522	594
80%	196	220	254	309	380	433

Once the test sample size is determined, the test and evaluation strategy can be developed that describes the test scenarios and corresponding specific number of physical tests. Initially, test scenarios can be derived as a mix of existing FMVSS tests and AV sensor suite edge or challenging cases. Examination of sensor algorithms would assist in prioritizing tests and test scenarios that score high in risk assessment parameters probability of failure and consequence of failure. For example, low sun angle is a challenge for camera technologies. Therefore, of the 425 tests, a proportion representative of the risk would be allocated to low sun angle conditions for which the camera technology significantly contributes to vehicle control. Finally, the test and evaluation strategy can be tailored to a specific geographical region, such as state, city, county, geofenced area, or national level.

Building the Safety Framework

Consider building the safety framework by vehicle classification. The U.S. follows a self-certification system of compliance, in which vehicle and equipment manufacturers certify that their products meet applicable standards. Additionally, the manufacturer determines the vehicle classification – e.g., passenger car, large truck, bus, etc. Historically, stakeholders have considered self-certification to be demonstrated through physical test. AVs will likely propel stakeholders toward a new era of targeted physical testing supplemented with extensive modeling and simulation to demonstrate safety.

The values shown in Table 3 represent an example of a safety framework for nonexempt passenger cars and light trucks. The crash risk is derived from reference [9]. The reliability distributions were selected to reflect the network element risk. For example, U.S. DOT data shows high motor vehicle reliability with high confidence as demonstrated through the low number of crashes caused by vehicle failure, likely buttressed by NHTSA recall authority. The environment – e.g., roadway surface, markings, traffic communications (signage, lighting), etc. - has high reliability with high confidence, however, atmospheric conditions may contribute uncertainty and adversely impact reliability and confidence. Estimating the current level of safety for human drivers of nonexempt vehicles to be 85% reliability with 90% confidence reflects the NVMCCS analysis attributing 94% of passenger car and light truck crashes due to driver error. The unknown/uncertainty element gives stakeholders flexibility to examine competencies and scenarios of interest; for this paper, the reliability distribution of 80% / 85% was assigned as a minimum value.

Table 3.
Example of a Safety Framework for Passenger Cars & Light Trucks

Network Element	Crash Risk [Ref. #]	FMVSS	Reliability Distribution	Number of Tests: Parameter(s)
Environment	2%	301, 302, 303, 304, 305	90% / 95%	109 tests: rain, ice, snow
Vehicle	2%	All Standards	95% / 95%	220 tests: brakes, steering, occupant protection, etc.
Driver	94%	Recognition: 101, 103, 104, 108, 111, 113, 123, 125, 131, 138, 205 Decision: 102, 108, 124, 135, 209, 210, 213, 225, 401 Performance: 105, 106, 109, 110, 116, 117, 118 Non-Performance: 114 Other:	85% / 90%	17 tests: low sun angle 28 tests: speed, curves, intersections 7 tests: lane management, LTAP 3 tests: maneuvers near a "taco truck" with pedestrians 4 tests: see NMVCCS data
Unknown/ Uncertainty	2%	Varies	80% / 85%	37 tests: double parked, orange cone, etc.
				Total = 425 tests

Overall safety by vehicle class allows for differentiation in safety levels. A safety framework for large trucks would vary from Table 3 in the crash risk and reliability distributions, requiring additional data analysis. It is likely that trained professional drivers would be associated with a lower crash risk and a higher reliability distribution for the Driver element, and a safety level higher than passenger cars.

Table 3 shows data parsed into the NMVCCS categories which aligns better with the construct of "level of safety" for AVs because it treats the common vehicle systems (e.g., powertrain, steering, braking, etc.) separate from the Driver, and also allows for refinement of the driver behaviors (e.g., recognition, decision, performance, non-performance, and other). However, the NMVCCS sample is only light vehicles and this distribution cannot be projected directly onto NHTSA GES or CDS estimates for other vehicle classifications. An alternate option could be to parse the test sample by pre-crash scenario (Rear End, Crossing Paths, Road Departure, Pedestrian, Cyclist, etc.) which is a good fit for tracking how crashes occur and factoring system effectiveness.

Successful implementation depends on reaching consensus on the metrics, collecting and sharing relevant AV characterization data, and revisiting at regularly defined intervals.

CONCLUSIONS

All aspects of AV safety and reliability must be demonstrated before candidate AVs are deployed onto public roadways. This proposed methodology establishes a framework to quantify safety performance levels and includes the flexibility to incorporate new data describing driver performance or technological capabilities as AV technology evolves. To paraphrase Aristotle, the whole framework is greater than the sum of its parts. Blending established engineering concepts from motor vehicle safety, reliability, and systems engineering to form a new approach to specify benchmark test & evaluation scenarios places a reasonable burden on all stakeholders and is achievable well before driving billions of miles.

Table 4 lists one measure of vehicle incidence rates for different types of vehicles. Mindful that VMT and flight hours are not comparable metrics, the promise of AVs may be realized if it follows the trend of automated aviation safety.

Table 4.
Compilation of Vehicle Incidence Data

Vehicle	Incidence Data	
Police-reported Motor Vehicles Crashes	0.0205 crashes/ 10,000 VMT	[Ref. 2]
Estimate for All Motor Vehicles Crashes	0.0457 crashes / 10,000 VMT	[Ref. 2]
CA DMV AV Disengagements	38.6 disengagements / 10,000 AV VMT	[2017 data]
Automated Aviation	0.5 accidents / 1 million take-offs	[Ref. 6]
Commercial Aviation	0.149 accidents / 10,000 flight hours	[Ref. 6]
General Aviation	7.11 accidents / 10,000 flight hours	[Ref. 6]
Customs & Border Patrol Aviation	52.7 accidents / 10,000 flight hours	[Ref. 6]

REFERENCES

[1] Balci, O. and R.G. Sargent. 1981. "A Methodology for Cost-Risk Analysis in the Statistical Validation of Simulation Models." Communications of the ACM 24, No. 4, April 1981: 190-197.

[2] Blincoe, L. J., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2015, May). *The economic and societal impact of motor vehicle crashes, 2010. (Revised)* (Report No. DOT HS 812 013). Washington, DC: National Highway Traffic Safety Administration.

[3] Federal Highway Administration (FHWA) website, https://safety.fhwa.dot.gov/hsip/resources/fhwasa09029/app_c.cfm, retrieved March 8, 2019.

[4] National Center for Statistics and Analysis. (2018, September). *Summary of motor vehicle crashes: 2016 data*. (Traffic Safety Facts. Report No. DOT HS 812 580). Washington, DC: National Highway Traffic Safety Administration.

[5] National Highway Traffic Safety Administration. (2008, July). *National Motor Vehicle Crash Causation Survey Report to Congress*. (Report No. DOT HS 811 059). Washington, DC: National Highway Traffic Safety Administration.

[6] Ni, R. and Leung, J. *Safety and Liability of Autonomous Vehicle Technologies*. https://groups.csail.mit.edu/mac/classes/6.805/student-papers/fall14-papers/Autonomous_Vehicle_Technologies.pdf, retrieved March 8, 2019.

[7] O'Connor, A., Modarres, M. and A. Mosleh. (2016). *Probability Distributions Used in Reliability Engineering*. College Park, Maryland: Center for Risk and Reliability.

[8] Reliability Analytics Toolkit website, https://reliabilityanalyticstoolkit.appspot.com/sample_size, retrieved March 8, 2019.

[9] Singh, S. (2015, February). Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey. (Traffic Safety Facts Crash•Stats. Report No. DOT HS 812 115). Washington, DC: National Highway Traffic Safety Administration.

[10] U.S. Department of Transportation, Bureau of Transportation Statistics website, https://www.bts.gov/content/us-vehicle-miles, retrieved March 8, 2019.

[11] Webb, David W. March 2011. “A Comparison of Various Methods Used to Determine the Sample Size Requirements for Meeting a 90/90 Reliability Specification.” Army Research Laboratory paper number ARL-TR-5468, https://www.arl.army.mil/arlreports/2011/ARL-TR-5468.pdf, retrieved March 8, 2019.