EVALUATING AUTOMATED VEHICLES SAFETY PERFORMANCE: TOWARDS A SURVIVAL ANALYSIS APPROACH FOR ASSESSING AUTOMATED VEHICLE SAFETY

A Dissertation

by

SOHEIL SOHRABI

Submitted to the Graduate and Professional School of

Texas A&M University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Dominique Lord
Committee Members,	Sue Chrysler
	Yunlong Zhang
	Wei Li
	Bahar Dadashova
Head of Department,	Robin Autenrieth

December 2021

Major Subject: Civil Engineering

ABSTRACT

Automated vehicle (AV) safety needs to be evaluated for successful development and deployment. However, the AV safety evaluation literature is scarce, and AV safety has not been validated yet. This study (1) synthesizes the existing knowledge about AV safety evaluations, (2) proposes a new methodology to address the gaps in AV safety evaluations, and (3) designs an empirical study to assess the safety performance of existing AVs under the road tests.

A scoping review is designed and conducted to systematically synthesize the literature about AV safety evaluations. As a result of this review, six AV safety quantification methods were identified and compared. This review showed that existing methodologies for AV safety evaluation carry certain shortcomings and cannot be used for reliable evaluation of AV safety. In addition, major challenges in AV safety evaluations are highlighted, including uncertainties in AV implementations and their impacts on AV safety, potential riskier behavior of AV passengers as well as other road users, and emerging safety issues related to AV implementations.

A new methodology based on a survival analysis approach is proposed to evaluate AV safety with limited road test data. To this end, the time-to-event, in the form of the number of miles to a crash (MTC), is incorporated to add a new layer of information, time, into the analysis. The likelihood of failure for both AV and conventional vehicles is further estimated, and the difference of the failure functions is statistically examined using the Anderson-Darling and Kolmogorov–Smirnov tests. Moreover, a new metric for evaluating the safety performance of vehicles, "no-crash expectancy," is defined, which represents the average number of miles that a vehicle is expected to travel before a crash happens.

Elaborating on the hazard rate of AVs as a function of the number of miles driven by the vehicle, this study further formulates crash prediction models in the era of automation and indicates the necessity of revisiting existing road safety analysis methods.

An empirical study is designed to address the limitations in performing an apple-to-apple comparison between AVs and conventional vehicle safety and examine the proposed safety evaluation methodology. Conventional crashes, including non-police-reportable crashes, were sourced from the Second Strategic Highway Research Program's naturalistic driving study (NDS) data. NDS data comprise the driver's trip trajectory information, constantly collected from a sample of drivers, reflecting both major and minor crashes. AV crashes are sourced from the California Department of Motor Vehicles Autonomous Vehicle Tester program. The results of the empirical study on conventional vehicles and Level 3 AV crashes showed that, with 95% confidence, automated driving is safer in terms of MTC. The results indicated that the no-crash expectancy would be increased by 27% when switching from conventional vehicles to AVs in 150,000 miles of road operation. Despite the uncertainties in AV crash reports, this study can be considered the most accurate verdict regarding Level 3 of automation safety.

This study has certain limitations, mainly inherited in the availability of data. Future studies are required to address the limitations of this study and the identified gaps and challenges in AV safety evaluations. The proposed methodology can be further expanded to evaluate the vehicle-level crash contributing factors, such as vehicle technologies.

DEDICATION

To Arefeh, Farhaneh, Behrouz, and Setareh

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my committee chair, Dr. Dominique Lord, my role model in academia, for his unconditional support and understanding throughout my education. This research would not have been possible without their guidance and extremely useful advice.

I would like to thank my committee members, Dr. Sue Chrysler, Dr. Yulong Zhang, Dr. Wei Li, and Dr. Bahar Dadashova, for their guidance, insight, and support throughout the course of this research. I am grateful for Dr. Fred Mannering's insightful comments on this research.

I am indebted to all the professors, teachers, and mentors throughout my education and my professional appointments, who inspired me to continue my educations, namely, Dr. Haneen Khreis, Mr. Bryan Wilson, Dr. Fred Mannering, Dr. Mohammadali Shirazi, Dr. Alireza Ermagun, Dr. Rajesh Paleti, Dr. Mecit Cetin, Dr. Khosrow Ovaici, Dr. Mahmoud Mesbah, Dr. Amir Golroo, Dr. Masoud Tabibi, Dr. Meeghat Habibian, Mr. Ali Mirzaiee, Mr. Ali Arab, Dr. Mohammadreza Elyasi, and Dr. Ali Akbarieh, among others.

Thanks also to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Although thousands of miles away from my home, I felt content during my stay in the United States, thanks to my wife and my incredible friends, who are too numerous to be mentioned.

Finally, thanks to my parents for their encouragement and to my wife for her love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by the dissertation committee consisting of Professors Dominique Lord and Yunlong Zhang of the Department of Civil and Environmental Engineering, Professor Wei Li of the Department of Landscape Architecture and Urban Planning, and Drs. Sue Chrysler and Bahar Dadashova of the Texas A&M Transportation Institute. Professor Fred Mannering from the University of South Florida provided constructive feedback on the proposed methodology in this research.

The relevant study identification and review in Chapter Two were conducted in part by Ali Khodadi and Seyedeh Maryam Mousavi of the Zachry Department of Civil and Environmental Engineering. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

The project was funded by three sources:

1) the Safety through Disruption (Safe-D) National UTC, a grant from the U.S. Department of Transportation's University Transportation Centers Program (Federal Grant Number: 69A3551747115),
2) the Robert Wood Johnson Foundation, Grant ID 76489, Quantifying the Harms and Benefits of Connected and Automated Vehicle Technologies on Public Health and Equity, and
3) the A.P. and Florence Wiley Faculty Fellow provided by the Texas A&M University College of Engineering and the Zachry Department of Civil and Environmental Engineering to Dr. Dominique Lord.

NOMENCLATURE

AV	Automated Vehicles
ACC	Adaptive Cruise Control
A-D	Anderson-Darling
ADAS	Advanced Driver Assistance Systems
ADS	Automated Driving System
AEB	Automated Emergency Braking
AHR	Average Heart Rate
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
AVT	Autonomous Vehicle Tester
BIC	Bayesian Information Criteria
BSW	Blind-Spot Warning
CA DMV	California Department of Motor Vehicles
CAV	Connected and Automated Vehicles
CRI	Crash Risk Index
DDI	Diverse Diamond Interchange
DDT	Dynamic Driving Tasks
EPA	Environmental Protection Agency
ESC	Electronic Stability Control
FCW	Forward Collision Warning
GOF	Goodness-Of-Fit

IEEE	Institute of Electrical and Electronics Engineers
IRB	Institutional Review Board
K-S	Kolmogorov-Smirnov
LCW	Lane-Change Warning
LDW	Lane-Departure Warning
MLE	Maximum Likelihood Estimation
MPR	Market Penetration Rate
MTC	Mile-To-Collision
NCJ	Number of Critical Jerks
NDS	Naturalistic Driving Survey
NHTSA	National Highway Traffic Safety Administration
ODD	Operation Design Domain
PCAM	Pedestrian Collision and Mitigate
PET	Post-Encroachment Time
P-P Plot	Probability-Probability Plot
Q-Q Plot	Quantile-Quantile Plot
RMST	Restricted Mean Survival Time
SAE	Society Of Automobile Engineers
SCRT	Skin Conductance Response Time
SDLP	Standard Deviation of Lane Position
SHRP2	Second Strategic Highway Research Program
SSM	Surrogate Safety Measures
SWR	Steering Wheel Reversed

TET	Time-Exposed Time-To-Collision
TIT	Time-Integrated Time-To-Collision
TRID	Transport Research International Documentation
TTC	Time-To-Collision
US	United States
VIN	Vehicle Identification Number
VMT	Vehicle Miles Traveled

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION*

This chapter contains the problem statement, including some definitions and background information, the objectives of the study, and the structure of the research.

1.1 Statement of Problem

Automated vehicles (AVs) have the potential to improve traffic safety profoundly, mainly by eliminating driver error. According to the National Highway Traffic Safety Administration (NHTSA), human error contributes to 94% of crashes, and AVs are optimistically expected to prevent these crashes (NHTSA, 2018). AV safety, and the related safety implication complexity, vary in terms of driving automation levels, as defined by the Society of Automobile Engineers (SAE, 2018). In the lower levels of automation (Levels 1 and 2), the driver is responsible for dynamic driving tasks (DDTs), and advanced driver assistance systems (ADASs) on the vehicle can sometimes assist the human driver with either steering or braking/accelerating (SAE, 2018). ADASs have the potential to prevent or mitigate crashes by partially eliminating driver error. In higher levels of automation, the automated driving system (ADS) performs the entire DDT while engaged. In Level 3, the DDT fallback-ready user needs to intervene when requested

* Part of this chapter is reprinted with permission from Sohrabi, S., Khodadadi, A., Mousavi, S.M., Dadashova, B. and Lord, D., 2021. Quantifying the automated vehicle safety performance: A scoping review of the literature, evaluation of methods, and directions for future research. Accident Analysis & Prevention, 152, p.106003. Copyright [2021] by Elsevier.

(SAE, 2018). On the other hand, Levels 4 and 5 of automation do not require a DDT fallback-ready user, and Level 5 has an unlimited operation design domain (ODD). An ADS is expected to eliminate driver error entirely; however, disengagement from ADSs and DDT fallback can be challenging.

AV impacts on safety can be investigated at three levels: vehicle, transportation system, and society (Figure 1.1). At the vehicle level, AVs can be examined in terms of how they contribute to the critical driver-related reasons for crashes, such as inattention; internal and external distractions; inadequate surveillance; decision error caused by false assumptions and perceptions; performance (i.e., execution of improper driver response); and nonperformance mainly due to impairment, drowsiness, and fatigue (NHTSA, 2018). AVs implementation carries higher levels of uncertainty at the transportation system level. AV safety can be examined based on its potential to reduce traffic conflicts and, consequently, reduce crashes. At the society level, crashes pose a public health crisis, and the health impacts of AVs can be investigated based on the changes in motor vehicle crashes (Sohrabi et al., 2020). Previously, the role of motor vehicle crashes in public health has been measured in the form of premature mortalities from fatality crashes (Sohrabi and Khreis, 2020) and the disability-adjusted life year from injury crashes (Tainio, 2015).

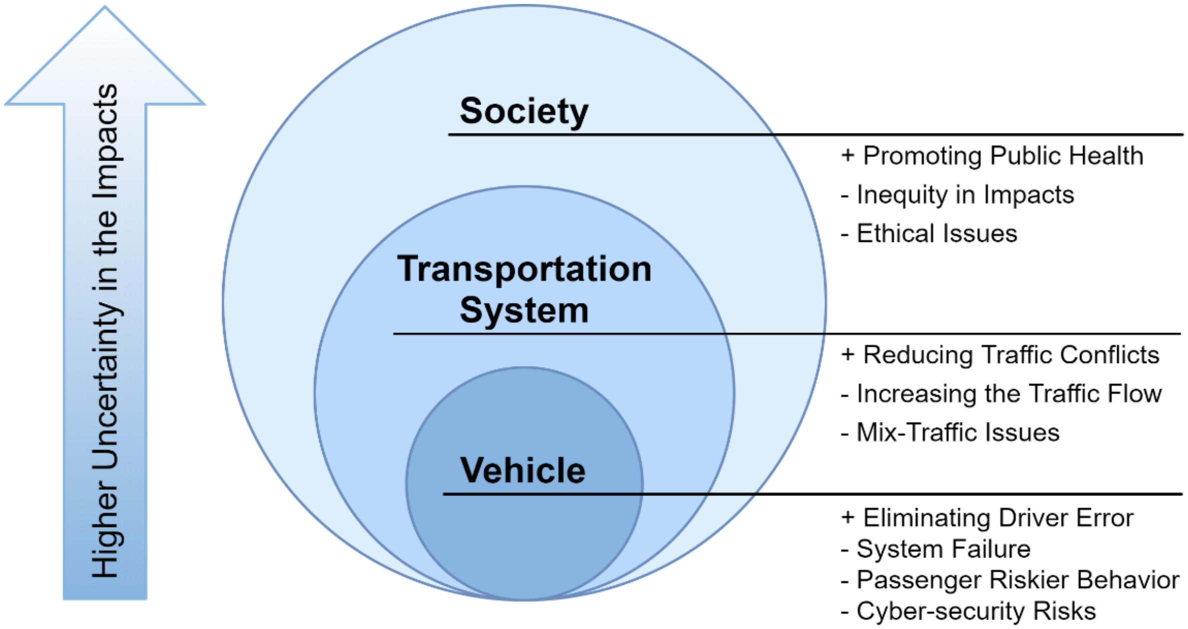


Figure 1.1. AV Levels of safety implications (Reprinted with permission from Sohrabi et al. 2021)

Although traffic crashes caused by driver error are expected to be eliminated after AVs' deployment, other safety issues may compromise the positive impacts (Kockelman et al., 2016, Litman, 2017, Yang et al., 2017). System operation failure (Koopman and Wagner, 2016), cybersecurity (Lee, 2017), and passengers' risky behaviors related to feeling overly safe while using AVs are some examples of potential safety concerns at the vehicle level. At the transportation system level, with AV market penetration rate (MPR) less than 100%, AVs may experience safety issues related to the interaction between human drivers and AVs in mixed traffic (Virdi et al., 2019, Taeihagh and Lim, 2018), as well as AVs' potential to increase traffic flow and, consequently, exposure to crashes as a result of induced demand, increased mobility, and changes in land use (Milakis et al., 2017). Moreover, due to the high cost of AVs, only wealthy consumers might be able to afford AVs as personal vehicles (Raj et al., 2019, Cohen and Shirazi, 2017) and, therefore, the disproportionate deployment of AVs may lead to health inequities that challenge AV safety impacts at the society level. The controversial discussion about how AVs should react during an unavoidable crash is another example of AV safety challenges at the society level.

Despite the complexities in AV safety evaluations and their impacts, accurate AV safety evaluations are required before deploying AVs. Particularly, the intent to use AVs and their market success are contingent upon the safety evaluation of AVs (Sener et al., 2019). In addition, not only can manufacturers and the automotive industry benefit from the accurate safety evaluations of AVs, but legislative and executive agencies require such information to advocate with industry stakeholders and society (Junietz et al., 2018). Evaluating the safety implications of AVs is necessary for formulating regulations and policies to alleviate the unintended consequences of AV implementations and increase their benefits, as outlined by the United States (US) Department of Transportation (US DOT, 2018) and the US Congressional Research Service (Canis, 2020).

The salient of the subject urged researchers to evaluate AV safety (reviewed by (Bagloee et al., 2016b, Sousa et al., 2017, Milakis et al., 2017, Martínez-Díaz and Soriguera, 2018, Montanaro et al., 2018)). Despite the previous effort, AV safety has not been validated yet (Milakis et al., 2017). The lack of AV safety validations could be associated with limitations in existing evaluation methods (Kalra, 2017). More specifically, Kalra (2017) pointed out the restrictions in AV road testing given the risk they impose on the other road users. As a result of these restrictions, there is not sufficient data for AV safety evaluations. The researcher resembles this situation with the “chicken and egg” paradox (further discussed in Chapter Three). Alternative evaluation methodologies were proposed to address the limitations in road test data by simulating AV operation and how they execute the DDT under different ODD (Wang et al., 2020) or reconstructing the crashes to explore measure AV contribution to the crashes (Kusano and Gabler, 2014), among others. This research targets AV safety evaluations to address its complexities and limitations in the existing methodologies.

1.2 Research Objectives

The main objective of this study is to contribute to the safety evaluation of AVs at the vehicle level, transportation system level, and society level. In this context, this study seeks the answer to three fundamental questions:

1- How can AV safety be validated, and what are the research gaps in the existing safety evaluation methods?
2- What methodologies are required to validate AV safety, evaluate their safety performance, and investigate the contributing factors to AV crashes?
3- How safe are the existing AVs in comparison with conventional vehicles?

First, this study is designed to conduct a comprehensive review of the AV safety evaluation literature, identify the existing AV safety evaluation approaches, and compare the identified approaches. Each approach is investigated in terms of its input, output, and application to estimate AV safety implications at the vehicle, transportation system, and society levels. The identified approaches are compared in terms of three criteria: availability of input data, suitability for evaluating different levels of automation, and reliability of estimations. Further, challenges in AV safety validation are identified. The results of this systematic review are expected to serve as a stop knowledge point and future research avenues to contribute to AV safety evaluation. Evaluating AV quantification methods can help researchers, policy makers, and practitioners choose an appropriate evaluation method based on their objectives.

This study proposes a methodology to statistically evaluate the safety of AVs in comparison with conventional vehicles with limited road test data. The proposed methodology adds a new layer of information, time, to address limitations in the availability of the road test data. The failure functions for AV and conventional vehicles are estimated, which represents the risk of being involved in a crash for each type of vehicle. Statistical tests are employed to compare the failure function of the vehicles and draw reliable conclusions regarding AV safety. In addition, a new metric is defined to compare the safety performance of vehicles effectively. The proposed framework for assessing AV safety, and its flexibility to enable further innovation, can address the decision maker's concerns (e.g., NHTSA's advance notice of proposed rulemaking for the development of a framework for the ADS safety [1]).

An empirical study is designed to (1) investigate the safety of existing AVs in comparison with conventional vehicles and (2) examine the proposed methodology. The results of implementing the proposed methodology on the designed empirical study are expected to offer the most accurate verdict regarding the safety of the existing AVs on a public road in terms of crash frequencies. This could contribute to the dialogue about AV safety among AV manufacturers, policymakers, and the public.

---

[1] Sourced from: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/ads_safety_principles_anprm_website_version.pdf (Accessed January 2021)

1.3 Structure of Research

This dissertation is divided into six chapters. Chapter Two reports the conducted scoping literature review on AV safety evaluation methodologies. This chapter contains the review methodology, a summary of review results, and an introduction to the identified AV safety evaluation methodologies. Two qualitative analyses on the identified methods and an extensive discussion on the challenges and gaps in AV safety evaluations are undertaken in this chapter.

Chapter Three introduces the proposed methodology for AV safety evaluation using road test data. This chapter describes how the proposed methodology addresses the identified gaps in Chapter Two. A brief introduction to the theories behind the proposed methodology and its literature in the context of traffic safety analysis is described in this chapter.

The designed three-step empirical study for AV safety evaluation is described in Chapter Four. This chapter also includes an introduction to AV crash datasets and the availability of the data. Finally, the data used for the empirical study and the sources of the data are introduced.

In Chapter Five, the results about the implementation of the proposed methodology on the designed empirical analysis are reported. This chapter is outlined based on the three steps of the empirical study, and the results of each step are reported subsequently.

Chapter Six contains a summary of the analysis and a discussion about the results of the research. This chapter reports the limitations in the literature review, the proposed methodology, and the conducted empirical study. Also, a set of recommendations for future research are included in this chapter.

CHAPTER 2

LITERATURE REVIEW*

In this chapter, a comprehensive literature review was conducted to (1) synthesize the AV safety evaluation literature, (2) summarize previous research findings, (3) compare the quantification methodologies, and (4) identify the gaps and limitations. In the subsequent section, the scoping review methodology is discussed. Then, the results of the literature review and elaborate on the identified quantification methodologies are reported. Next, two qualitative analyses on the identified methods are compared. This section is followed by a detailed discussion about AV safety evaluation challenges and limitations. Finally, a summary of the chapter is provided.

2.1 Review Methodology

A scoping review methodology framework proposed by Arksey and O’Malley (2005) was followed in this study. A scoping review methodology was selected rather than a systematic review since this study aims to identify previous studies answering a general question and then review the evidence from previous quantifications on AVs’ impact on traffic safety (Munn et al., 2018). In this context, the findings are not aggregated, nor is the quality of evidence assessed (Arksey and O'Malley, 2005).

---

* Part of this chapter is reprinted with permission from Sohrabi, S., Khodadadi, A., Mousavi, S.M., Dadashova, B. and Lord, D., 2021. Quantifying the automated vehicle safety performance: A scoping review of the literature, evaluation of methods, and directions for future research. Accident Analysis & Prevention, 152, p.106003. Copyright [2021] by Elsevier.

2.1.1 Review Question

The first step in a scoping review is to identify a research question to be answered (Arksey and O'Malley, 2005). The research question for this review was the following: “What are the methodologies and the gaps in the existing research on quantifying the potential impacts of AVs on traffic safety?” Specifically, this review identified the research that quantified the impacts of AVs rather than studies of a speculative nature.

2.1.2 Identifying Relevant Studies

A search strategy was developed to retrieve relevant research evidence from four electronic research databases—Scopus, Web of Science, Transport Research International Documentation (TRID), and Institute of Electrical and Electronics Engineers (IEEE) Xplore—as well as reference lists of the retrieved publications. IEEE Xplore is a research database that covers more than five million journal articles, conference proceedings, standards, and related materials on multiple disciplines, including but not limited to computer science, electrical engineering and electronics, and allied fields.[2] IEEE Xplore is sponsored by IEEE and other partner publishers. Scopus is Elsevier’s research database, which covers more than 75 million records from 50,000 publishers in four core areas: life sciences, social sciences, physical sciences, and health science.[3] The Web of Science, sponsored by the Institute of Scientific Information, is a publisher-independent research database that covers more than 79 million records from several areas, such as life sciences, biomedical sciences, engineering, social sciences, arts and humanities, natural

[2] Sourced from: https://innovate.ieee.org/about-the-ieee-xplore-digital-library/
[3] Sourced from: https://www.elsevier.com/solutions/scopus/why-choose-scopus

sciences, health sciences, engineering, computer science, and materials sciences.[4] TRID database, a research database that combines the records from the Transportation Research Board's Transportation Research Information Services, is also explored, which is solely focused on transportation research and provides access to more than 1.25 million records.[5]

The databases were searched to identify published articles, letters, reports, book chapters, and books using any combination of two sets of keywords in their title, abstract, and keywords: ["autonomous vehicle" or "autonomous car" or "self-driving car" or "driverless car" or "automated driving"] and ["crashes" or "accidents" or "collision" or "safety"]. Due to the burdensome translating process, only the published material written in English is included in this review. All material considered in the review was published as of October 2020.

2.1.3 Study Selection

To ensure consistency in selecting studies that answered the review's question and excluded irrelevant studies, a set of inclusion and exclusion criteria are defined. The included studies had to meet the following established criteria:

1. Must explicitly quantify AVs' impacts on traffic safety rather than merely offer speculations and qualitative assessments.
2. Must evaluate AV as a vehicle for ground transportation, such as automated cars, buses, shuttles, trucks, and the like.
3. Must investigate the safety of different levels of vehicle automation rather than individual AV technologies (e.g., ADASs, sensors, and algorithms).

[4] Sourced from: https://clarivate.libguides.com/webofscienceplatform/coverage
[5] Sourced from: http://www.trb.org/InformationServices/AboutTRID.aspx

Based on the inclusion criteria, connected vehicles' safety evaluations did not fall within the scope of this study. However, the literature on connected and automated vehicles (CAVs) is included in the review, with a focus on the safety evaluation of automation components of CAVs. The selection process was divided into two stages. First, the titles and abstracts of the identified publications were reviewed, and potentially relevant publications were selected. Second, the full text of the potentially relevant publications was retrieved and reviewed against the inclusion criteria, and studies that did not meet all inclusion criteria were excluded. The reference lists of included publications were also reviewed to find any relevant articles that were not identified through the developed search strategy.

2.2 Search Results and Characteristics of Included Studies

The implemented scoping review process is shown in Figure 2.1. As of October 2020, a total of 1,859 publications were identified using the developed search strategy. After checking for duplicates, screening the identified articles, and reviewing articles' full text, 1,809 articles were excluded: 324 duplicates, 1,396 after screening, and 89 after full-text review. Ultimately, 50 articles met the inclusion criteria and were included in this review.

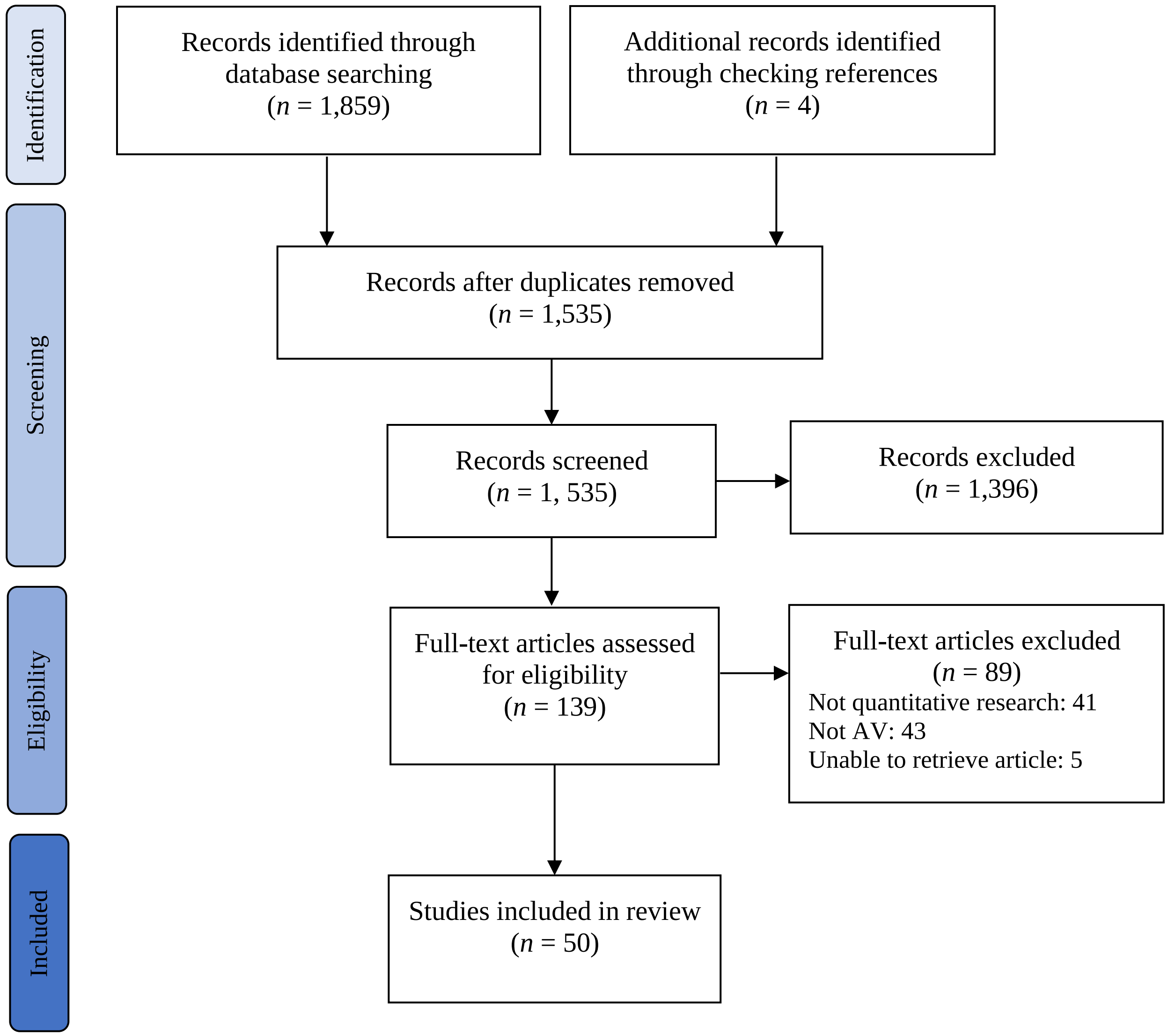


Figure 2.1. Study identification and selection mechanism of the implemented scoping review

The number of publications increased significantly beginning in 2012, although in 2019, only 14 articles were published on quantifying AV safety implications (Figure 2.2a). The AV safety quantification approaches can be classified into six groups: target crash population, traffic simulation, driving simulator, road test data analysis, system failure risk assessment, and safety effectiveness estimation. Figure 2.2b shows the distribution of quantification approaches. Road test data analysis and simulation studies were more commonly used in the literature, followed by the driving simulator and target crash population approaches. Failure risk assessment and safety effectiveness quantification received the least attention. A time-series analysis of publications indicated that traffic simulation and road test data analysis methods began receiving more attention over time. Increases in road test data may be one of the reasons behind this change.

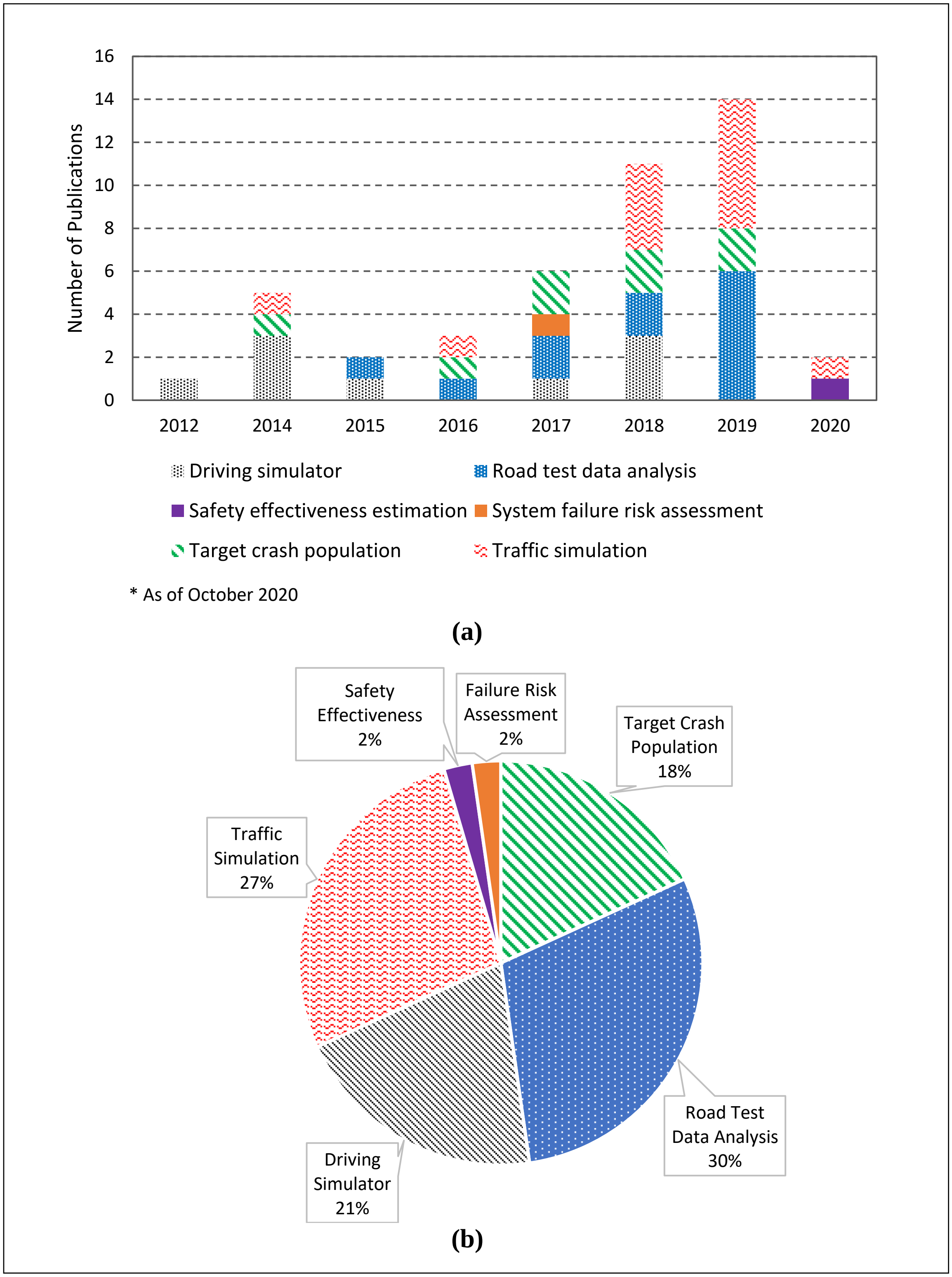


Figure 2.2. (a) Publication date of the studies included in this review, and (b) distribution of the identified AV safety quantification approaches (Reprinted with permission from Sohrabi et al. 2021)

2.3 Identified AV Safety Evaluation Approaches

In this section, the included studies are reviewed. A narrative review of the literature under the six identified AV safety quantification approaches is reported.

2.3.1 Target Crash Population

The target crash population approach quantifies the number of preventable crashes after AV implementation. The quantification process in the examined studies followed three steps (Rau et al., 2015, Yanagisawa et al., 2017):

1. Identify AVs' ADS and ADAS functionality.
2. Match AV functionality with the target crash type.
3. Explore the crash datasets and identify preventable crashes.

In the first step, AV functions were investigated on the basis of (a) levels of automation (Lubbe et al., 2018, Agriesti et al., 2019) and (b) individual or combined ADS and ADAS functions (Combs et al., 2019, Detwiller and Gabler, 2017, Hendrickson and Harper, 2018, Li and Kockelman, 2016, Kusano and Gabler, 2014).

In the second step, AV functionality was matched with corresponding crash characteristics. Previous studies assessed AV technology to mitigate either specific crash types (e.g., rear-end collision, pedestrian crashes) (Combs et al., 2019, Detwiller and Gabler, 2017, Hendrickson and Harper, 2018), specific crash-contributing factors (e.g., distracted driving, speeding, etc.), or critical pre-crash events (e.g., running a red light, vehicle failure) (Yanagisawa et al., 2017, Lubbe et al., 2018, Li and Kockelman, 2016, Kusano and Gabler, 2014). In addition, some AV functions are programmed to operate under a certain ODD to activate and achieve the maximum desired effectiveness; therefore, the crash dataset had to be filtered out to mirror those conditions properly. Lighting condition (day/night) (Yanagisawa et al., 2017, Agriesti et al., 2019), weather condition (clear/adverse) (Yanagisawa et al., 2017, Agriesti et al., 2019), road surface condition (wet/dry) (Yanagisawa et al., 2017, Agriesti et al., 2019), travel speed range (Yanagisawa et al., 2017, Agriesti et al., 2019, Hendrickson and Harper, 2018), visual obstruction (Lubbe et al., 2018, Combs et al., 2019), pedestrian crossing condition (Lubbe et al., 2018, Detwiller and Gabler, 2017), lane marking condition (Lubbe et al., 2018, Agriesti et al., 2019), and stable vehicle condition (Lubbe et al., 2018) are conditions under which AV safety was examined in the literature. AV safety implications were explored for various road facilities and areas (Detwiller and Gabler, 2017, Hendrickson and Harper, 2018) as well. However, in some studies, facility type was automatically filtered out by selecting possible crash scenarios (e.g., running a red light, which is specific to intersections only) and beneficial safety equipment specific to that facility (e.g., cooperative intersection collision avoidance systems, which are applicable in intersections only) (Li and Kockelman, 2016, Kusano and Gabler, 2014). The safety effectiveness of AV technology was widely presumed to be 100% in the literature (Yanagisawa et al., 2017, Agriesti et al., 2019, Detwiller and Gabler, 2017, Hendrickson and Harper, 2018, Kusano and Gabler, 2014); however, some studies accounted for the shortcomings in the safety implications of AVs by considering the effectiveness of AV technology (Lubbe et al., 2018, Combs et al., 2019, Li and Kockelman, 2016). AV safety effectiveness was either extracted from simulation studies (Combs et al., 2019) or indirectly through defining different sets of rules (Lubbe et al., 2018, Li and Kockelman, 2016). Each set consisted of assumptions regarding weather, road condition, vehicle condition, speed range, and so forth, through which both maximum effectiveness and lower effectiveness due to adverse conditions could be taken into account. Moreover, different rule sets provided a lower and upper bound for the expected number of preventable crashes instead of a constant value for effectiveness. Most of the literature assumed a 100% MPR; indeed, only two studies considered the MPR in their analysis (Agriesti et al., 2019, Li and Kockelman, 2016).

In the third step, the crash datasets were explored, and the crash characteristics were extracted. Next, the safety benefits of AVs were quantified in terms of the number of preventable crashes (Yanagisawa et al., 2017, Lubbe et al., 2018, Agriesti et al., 2019, Combs et al., 2019, Detwiller and Gabler, 2017, Hendrickson and Harper, 2018, Kusano and Gabler, 2014) and/or reduced cost of crashes (Yanagisawa et al., 2017, Hendrickson and Harper, 2018, Li and Kockelman, 2016). As a result, AV safety was attributed to ADSs (Yanagisawa et al., 2017, Lubbe et al., 2018, Agriesti et al., 2019, Combs et al., 2019, Detwiller and Gabler, 2017, Hendrickson and Harper, 2018, Kusano and Gabler, 2014, Li and Kockelman, 2016) and ADASs (Combs et al., 2019, Hendrickson and Harper, 2018, Li and Kockelman, 2016, Kusano and Gabler, 2014). The total number of preventable crashes was estimated in the target crash population methodology, and some studies stratified crashes based on severity level (Detwiller and Gabler, 2017, Hendrickson and Harper, 2018, Li and Kockelman, 2016, Kusano and Gabler, 2014). Table A1 in the appendix summarizes the target population studies.

2.3.2 **Road Test Data Analysis**

Analyzing AV road tests is one of the approaches used in the literature to evaluate AV safety. AV incident data were sourced from the California Department of Motor Vehicles (CA DMV) (Schoettle and Sivak, 2015, Teoh and Kidd, 2017, Favarò et al., 2017, Matysiak and Razin, 2018, Banerjee et al., 2018, Xu et al., 2019, Wang and Li, 2019, Petrović et al., 2020, Boggs et al., 2020, Das et al., 2020), US National Transportation Safety Board (NTSB) (Wang and Li, 2019), or AV manufacturers' self-reports (Schoettle and Sivak, 2015). CA DMV mandates that all manufacturers testing AVs on public roads file two different types of reports: (a) a report of a collision involving an AV within ten days after the collision; and (b) an annual report summarizing the disengagements.

Three types of analyses were found in the literature. First, the rate of AV incidents was compared to conventional car crashes as a benchmark (Schoettle and Sivak, 2015, Teoh and Kidd, 2017, Matysiak and Razin, 2018, Banerjee et al., 2018, Favarò et al., 2017). The AV incident rate was estimated as either number of crashes per number of AV vehicle miles traveled (VMT) (Schoettle and Sivak, 2015, Teoh and Kidd, 2017, Favarò et al., 2017) or the number of disengagements per VMT (Matysiak and Razin, 2018, Banerjee et al., 2018). AV incident rates were then compared to either conventional vehicle crash rates (Schoettle and Sivak, 2015, Teoh and Kidd, 2017, Favarò et al., 2017, Banerjee et al., 2018) or injury and fatality crash rates (Matysiak and Razin, 2018). Unlike AV crashes, where the auto manufacturers report every single incident involving AVs, conventional vehicle crashes are reported by police based on the dollar amount of the property damage and therefore are significantly underreported. To have a fair comparison between AVs and conventional vehicle crash rates, Toeh and Kidd (2017) used AV police-reportable crashes, and Schoettle and Sivak (2015) adjusted the conventional vehicle crash rates for underreporting. Given the disparities in the equivalence between AV and conventional vehicle crash rates, mixed conclusions were drawn in the literature regarding AV safety in terms of crash rates.

Second, some studies investigated the characteristics of AV crashes in terms of collision type, crash location, speed, and causes of the crash. The majority of the literature ran a descriptive analysis of AV characteristics (Schoettle and Sivak, 2015, Favarò et al., 2017, Xu et al., 2019, Petrović et al., 2020), whereas some compared AV crash characteristics to conventional vehicle crashes (Schoettle and Sivak, 2015, Favarò et al., 2017, Petrović et al., 2020). Researchers found that the rate of rear-end crashes is higher in AV crashes (Schoettle and Sivak, 2015, Favarò et al., 2017, Petrović et al., 2020), while the severity of crashes is lower (Schoettle and Sivak, 2015). More rigorous statistical analyses, in the form of logistic regression (Wang and Li, 2019, Xu et al., 2019), a decision tree (Wang and Li, 2019), a Bayesian latent class model (Das et al., 2020), and logit discrete choice models (Boggs et al., 2020) were used to uncover the factors contributing to AV crash risk (Boggs et al., 2020), collision type (Xu et al., 2019, Wang and Li, 2019), and severity (Xu et al., 2019, Wang and Li, 2019). Driving speed, on-street parking, speed limit, and collision location—highway, arterial and collector, streetlights, and intersections—were shown to be associated with AV crash risk. The number of lanes marked with a centerline and clear weather conditions were shown to reduce the likelihood of AV crashes. AV driving mode (AV mode or conventional driver), collision location, roadside parking, rear-end collision, and one-way road were the main factors found to contribute to the severity level of AV-involved crashes. AV driving mode, AV stopped or not, vehicle turning movement, and whether crashes were associated with yielding to pedestrians/cyclists were the factors found to affect the collision type of AV crashes. The cause of AV disengagement was investigated by Banerjee et al. (2018), who found that 64% of disengagements were the result of problems in, or untimely decisions made by, the machine learning system.

Third, the safety reliability of AVs was examined by comparing (a) the AV failure rate to other safety-critical autonomous systems (Banerjee et al., 2018); (b) the number of miles driven by AVs until a crash to the number of miles driven by conventional cars until a crash (Favarò et al., 2017); (c) the number of failure-free miles AVs should drive to reach conventional cars' failure rates (Kalra and Paddock, 2016, Li and Zhai, 2019); (d) the total number of miles driven to evaluate AV failure rate (Kalra and Paddock, 2016, Li and Zhai, 2019); and (e) the total number of miles AVs need to drive to demonstrate their failure rate is statistically lower than that of conventional cars (Kalra and Paddock, 2016). Banerjee et al. (2018) compared AV reliability with other safety-critical autonomous systems in terms of reliability per mission and demonstrated that AVs are 4.22 times worse than airplanes and 2.5 times better than surgical robots. Favarò et al. (2017) estimated that AVs drive 500,000 miles before a crash, which shows AVs' reliability versus conventional vehicles. However, estimations regarding the number of failure-free miles AVs should drive to reach conventional vehicles' failure rate resulted in higher thresholds of 1.6 million miles (Kalra and Paddock, 2016) and 140 million miles (Li and Zhai, 2019). Kalra and Puddok (2016) showed that AVs need to be driven 51 and 61 million miles to be able to test their failure rate and statistically examine their failure rate, respectively. However, much higher numbers (71 billion miles) have been estimated for AV testing requirements to be able to properly investigate AV safety (Kalra and Paddock, 2016). Table A2 summarizes the studies that used AV road test data to evaluate their safety.

2.3.3 Traffic Simulations

During the last decade, traffic simulation models have been frequently implemented to replicate conventional vehicles' driving characteristics in a fleet (Young et al., 2014). Research studies have employed traffic simulation models to assess AVs' safety effects and the assumption, methodologies, and limitations behind them (see Table A3 for a summary of related literature).

In the identified traffic simulation studies, various traffic microsimulation computer software was used, such as VISSIM (Kockelman et al., 2016, Katrakazas et al., 2019, Morando et al., 2018, Deluka Tibljaš et al., 2018, Rahman et al., 2019, Arvin et al., 2020, Mousavi et al., 2020), MATLAB, SUMO, VENTOS, and PELOPS (Bahram et al., 2014, Arvin et al., 2018, Arvin et al., 2019, Qin and Wang, 2019). Depending on the study purpose, safety was evaluated at roadway segments (Katrakazas et al., 2019, Bahram et al., 2014, Ye and Yamamoto, 2019, Virdi et al., 2019, Qin and Wang, 2019, Zhang et al., 2015, Sinha et al., 2020), intersections (Kockelman et al., 2016, Arvin et al., 2018, Arvin et al., 2019, Morando et al., 2018, Virdi et al., 2019, Rahman et al., 2019, Arvin et al., 2020, Mousavi et al., 2020), roundabouts (Morando et al., 2018, Deluka Tibljaš et al., 2018), or on/off-ramps (Kockelman et al., 2016).

For developing the simulation scenarios, different car-following models were utilized for conventional vehicles and AVs. Various car-following models were implemented to replicate conventional vehicles' driving behavior, such as Wiedemann 74 (Arvin et al., 2018, Deluka Tibljaš et al., 2018, Virdi et al., 2019, Arvin et al., 2020, Mousavi et al., 2020), Wiedemann 99 (Katrakazas et al., 2019, Morando et al., 2018, Zhang et al., 2015, Sinha et al., 2020), and user-defined models (Ye and Yamamoto, 2019). For AVs, car following was in the form of modified built-in models, including modified Wiedemann models (Kockelman et al., 2016, Arvin et al., 2018, Morando et al., 2018, Deluka Tibljaš et al., 2018, Arvin et al., 2020, Mousavi et al., 2020) or AV-specific models using external coding interfaces to either adjust a variable, introduce a new following strategy, or test various models (Bahram et al., 2014, Arvin et al., 2018, Ye and Yamamoto, 2019, Papadoulis et al., 2019, Virdi et al., 2019, Sinha et al., 2020). In general, Wiedemann characterizes the car-following behavior by look-ahead distance, look-back distance, and average standstill distance, while modified Wiedemann 99 also considers headway time (PTV, 2018).

Based on driving behaviors, various scenarios were developed to evaluate the impact of AVs on safety. The majority of the studies explored different AV MPRs as the main variable (Katrakazas et al., 2019, Bahram et al., 2014, Rahman et al., 2019, Arvin et al., 2018, Arvin et al., 2019, Morando et al., 2018, Deluka Tibljaš et al., 2018, Ye and Yamamoto, 2019, Papadoulis et al., 2019, Qin and Wang, 2019, Arvin et al., 2020, Sinha et al., 2020). Depending on the study, each simulation scenario was run multiple times to obtain reliable outputs for evaluating traffic safety. Since simulations do not lead to any crash, near-miss events were used instead to assess safety, which is an important limitation for using traffic simulation programs (Lord et al., 2021).

Surrogate safety measures (SSMs) were used to determine the number of near-miss events and, consequently, the associated level of traffic safety. The most commonly used SSMs in the studies were time-to-collision (TTC) and post-encroachment time (PET) (Kockelman et al., 2016, Katrakazas et al., 2019, Bahram et al., 2014, Arvin et al., 2018, Arvin et al., 2019, Morando et al., 2018, Deluka Tibljaš et al., 2018, Ye and Yamamoto, 2019, Papadoulis et al., 2019, Mousavi et al., 2020, Sinha et al., 2020). Acceleration rate and velocity difference (Ye and Yamamoto, 2019, Sinha et al., 2020), time-exposed time-to-collision (TET) (Bahram et al.), time-integrated time-to-collision (TIT) (Bahram et al., 2014, Qin and Wang, 2019, Zhang et al., 2015, Rahman et al., 2019), time-exposed rear-end crash risk index (TERCRI) (Zhang et al., 2015, Rahman et al., 2019), the number of critical jerks (NCJ) (Rahman et al., 2019), and lane-change conflicts (Zhang et al., 2015) were the other types of SSMs used in these studies.

Most of the studies concluded that by increasing the AV MPR, the number of near-miss events decreased on-road segments (Bahram et al., 2014, Morando et al., 2018, Ye and Yamamoto, 2019, Qin and Wang, 2019, Sinha et al., 2020), at intersections (Kockelman et al., 2016, Arvin et al., 2018, Arvin et al., 2019, Morando et al., 2018, Rahman et al., 2019, Arvin et al., 2020, Mousavi et al., 2020), at priority intersections (Virdi et al., 2019), in bottlenecks, at on/off-ramps (Kockelman et al., 2016), and in roundabouts (Morando et al., 2018, Virdi et al., 2019). However, Deluka et al. (2018) indicated that an increase in the AV MPR in roundabouts led to an increase in the number of conflicts. Moreover, Kockelman et al. (2016) showed an increase in conflicts by increasing the AV MPR at intersections. On the other hand, other studies showed that low AV MPRs were associated with a higher number of conflicts compared to zero MPR, yet, the number of conflicts decreased at intersections (Arvin et al., 2018, Virdi et al., 2019) and diverse diamond interchange (DDI) intersections (Virdi et al., 2019) by increasing the MPR in the simulation environment. Katrakazas (2019) also proposed a method to enable AVs to determine their trajectories to enhance safety in emergency situations. Study results indicated that the proposed method is capable of improving safety.

2.3.4 Driving Simulators

Probable challenges in human-vehicle interaction in the AV domain can take place in either the AV driver and AV interface stage (e.g., taking-over process) or the interaction between conventional vehicles and AVs (e.g., conventional vehicles entering the platoon of AVs). At different levels of automation, the AV driver needs to monitor or even intervene in the automation system to some extent in order to compensate for automation biases. On the other hand, AVs, at any MPR, will interact with conventional vehicles before they entirely dominate the future transportation system. In both cases, detailed knowledge of human driving behavior and reactions is necessary to evaluate AV safety. All the safety-related scenarios in reviewed studies could be categorized as (a) vehicle-human interaction (take-over situations in different driving states, such as drunk driving, drowsy driving, distracted driving, unplanned disengagement from the ADS, planned disengagement, etc.) (Strand et al., 2014, Kundinger et al., 2018, Berthelon and Gineyt, 2014, Gold et al., 2018, Happee et al., 2017, Blommer et al., 2015, Yun and Yang, 2020, Lee et al., 2020), or (b) vehicle-vehicle interaction (joining a conventional vehicle to a platoon of AVs) (Gouy et al., 2012, Lee et al., 2018). In both categories, a hazard scenario must be designed to determine the driver's performance in the evasive situation of interest. A hazard scenario is a situation that triggers the driver to make a maneuver and might be (a) a suddenly blocked lane by another vehicle(s) or an obstacle (Gold et al., 2018, Happee et al., 2017, Blommer et al., 2015, Yun and Yang, 2020, Lee et al., 2020), a sudden drift toward the edge of the road (Desmond et al., 1998), or a deceleration failure (Strand et al., 2014); or (b) safety challenges faced during driving, such as entering a platoon environment (Gouy et al., 2012, Lee et al., 2018) or controlling the vehicle while drowsy or drunk (Kundinger et al., 2018, Berthelon and Gineyt, 2014). The simulator experiments included three aspects—participants, experimental variables, and safety measurements—that had to be designed before the main experiment.

Different characteristics of participants used in designing simulator experiments included the following: age (Berthelon and Gineyt, 2014, Gold et al., 2018, Happee et al., 2017, Blommer et al., 2015, Gouy et al., 2012, Strand et al., 2014, Lee et al., 2018, Kundinger et al., 2018, Desmond et al., 1998, Yun and Yang, 2020, Lee et al., 2020), gender (Happee et al., 2017, Gold et al., 2018, Blommer et al., 2015, Gouy et al., 2012, Strand et al., 2014, Lee et al., 2018, Kundinger et al., 2018, Berthelon and Gineyt, 2014, Desmond et al., 1998, Yun and Yang, 2020, Lee et al., 2020), annual mileage driven (Strand et al., 2014), driving experience (Strand et al., 2014, Gouy et al., 2012, Berthelon and Gineyt, 2014, Yun and Yang, 2020, Lee et al., 2020), previous experience with automated driving (Strand et al., 2014, Blommer et al., 2015), prior experience with a driving simulator (Gouy et al., 2012, Gold et al., 2018, Happee et al., 2017), and mental/physical health condition (Kundinger et al., 2018, Berthelon and Gineyt, 2014, Lee et al., 2020). Each experiment took place in a controlled ODD and was based on a predefined procedure. Predesigned factors, such as (a) traffic density (Gold et al., 2018, Happee et al., 2017, Blommer et al., 2015, Strand et al., 2014, Gouy et al., 2012, Lee et al., 2018, Kundinger et al., 2018, Berthelon and Gineyt, 2014), (b) MPR (Lee et al., 2018), (c) facility type (Gold et al., 2018, Happee et al., 2017, Blommer et al., 2015, Strand et al., 2014, Gouy et al., 2012, Lee et al., 2018, Kundinger et al., 2018, Berthelon and Gineyt, 2014, Yun and Yang, 2020), and (d) repetition of experiment (Happee et al., 2017, Gold et al., 2018, Strand et al., 2014, Gouy et al., 2012, Desmond et al., 1998, Yun and Yang,

2020) and controlled factors—including the facility geometry design characteristics (Gold et al., 2018, Happee et al., 2017, Blommer et al., 2015, Gouy et al., 2012, Lee et al., 2018, Berthelon and Gineyt, 2014) and speed (Gold et al., 2018, Happee et al., 2017, Blommer et al., 2015, Strand et al., 2014, Gouy et al., 2012, Lee et al., 2018, Kundinger et al., 2018, Berthelon and Gineyt, 2014, Desmond et al., 1998, Yun and Yang, 2020)—were common experimental characteristics found in simulator studies. Some studies conducted only one experiment per participant to avoid learning effect bias (Blommer et al., 2015, Kundinger et al., 2018, Lee et al., 2018); others repeated the experiment to extract the maximum information from the available resources and tried to mitigate the learning effect bias by incorporating it as a variable in the model. However, almost all studies conducted a trial run before the main experiment to familiarize the participants with the simulator environment.

A metric is required to measure AVs' performance and quantify the risks and benefits of AVs using simulator studies. To this end, SSMs were widely used as the response variable to quantify safety risks and benefits of AVs, namely average/maximum/minimum speed (Berthelon and Gineyt, 2014, Lee et al., 2020), time headway (Strand et al., 2014, Gouy et al., 2012), take-over time (TOT) (Gold et al., 2018), TTC (Gold et al., 2018, Happee et al., 2017, Strand et al., 2014, Lee et al., 2020), distance to collision (DTC) (Lee et al., 2020), time to lane change (TTL) (Yun and Yang, 2020), brake application (Gold et al., 2018), crash/crash probability (Gold et al., 2018, Berthelon and Gineyt, 2014), steering response time (Happee et al., 2017, Lee et al., 2018), response time (Blommer et al., 2015, Strand et al., 2014, Yun and Yang, 2020), percent of the time with eyes on the road (Blommer et al., 2015), clearance toward the obstacle (Happee et al., 2017), road clearance metric (Happee et al., 2017), steering magnitude (Lee et al., 2018), lateral/longitudinal control (e.g., longitudinal/lateral deceleration) (Desmond et al., 1998, Lee et al., 2020), standard deviation of lane position (SDLP) (Yun and Yang, 2020, Lee et al., 2020), steering wheel reversed (SWR) (Yun and Yang, 2020), Karolinska Sleepiness Scale (Kundinger et al., 2018), physical and perceptual fatigue (Desmond et al., 1998), skin conductance response time (SCR) (Yun and Yang, 2020), and average heart rate (AHR) (Yun and Yang, 2020). The point of modeling different SSMs relates to the difference in their ability to capture near-crash events and critical maneuvers.

Finally, the SSMs were used to (a) find contributing factors to safety risk and benefits of AVs in different settings (Gold et al., 2018, Happee et al., 2017, Blommer et al., 2015, Strand et al., 2014, Gouy et al., 2012, Lee et al., 2018, Berthelon and Gineyt, 2014, Yun and Yang, 2020, Lee et al., 2020), and (b) compare AV safety with conventional vehicle safety (Happee et al., 2017, Kundinger et al., 2018, Desmond et al., 1998). Linear regression (Gold et al., 2018), logistic regression (Lee et al., 2018), univariate/multivariate analysis of variance (ANOVA) (Blommer et al., 2015, Strand et al., 2014, Gouy et al., 2012, Lee et al., 2018, Berthelon and Gineyt, 2014, Yun and Yang, 2020, Lee et al., 2020), Fisher's exact test (Strand et al., 2014), analysis of covariance (ANCOVA) (Strand et al., 2014), and Cochran's Q test (Strand et al., 2014) were used to identify significant variables that influenced AV safety. Besides the participant characteristics and experiment characteristics (or elements) mentioned before, other variables—such as time budget (Gold et al., 2018, Happee et al., 2017), lanes driven (Gold et al., 2018, Happee et al., 2017), type of secondary tasks (Gold et al., 2018, Happee et al., 2017, Blommer et al., 2015, Lee et al., 2020), automation level (Strand et al., 2014), disengagement scenarios (planned/unplanned) (Yun and Yang, 2020), types of take-over warnings (Yun and Yang, 2020), the extent of hazard scenario and challenges (e.g., moderate/severe/complete deceleration failure, or different time headway within the platoon) (Strand et al., 2014, Gouy et al., 2012), platoon size (Lee et al., 2018), and alcohol concentration (Berthelon and Gineyt, 2014)—were considered. Results showed that take-over scenarios, traffic density, experiment repetition, and defined time budget were highly influential factors affecting SSMs (Gold et al., 2018). In addition, scheduled disengagement (Blommer et al., 2015), lower automation levels, lower extent of hazard scenarios (Strand et al., 2014), engaging in non-driving-related tasks with less cognitive load (Lee et al., 2020), and use of multimodal take-over warning systems (Yun and Yang, 2020) led to better performance of drivers during the take-over situation. Drunk driving affected the longitudinal and lateral control of the vehicle and driver reaction to evasive maneuver, especially in lower automation levels (Berthelon and Gineyt, 2014). Moreover, in the platoon environment, the higher MPR (Lee et al., 2018) and lower time headway of AVs resulted in more aggressive driving behavior from conventional vehicles joining the platoon. To compare conventional vehicles and AVs in terms of safety risks and benefits, researchers mostly used ANOVA (Kundinger et al., 2018, Desmond et al., 1998) and Fisher's exact test (Happee et al., 2017). Results showed that automated driving would negatively affect a take-over scenario in response to a risk while the vehicle is disengaged from the ADS (Happee et al., 2017, Desmond et al., 1998) and increase driver drowsiness (Kundinger et al., 2018) compared to manual driving.

More details on the reviewed driving simulator studies can be found in Table A4.

2.3.5 System Failure Risk Assessment

System operation failure is one probable risk that AVs encounter (Koopman and Wagner, 2016). Malfunctioning sensors in detecting objects (pedestrians, bikes and cyclists, vehicles, obstacles, etc.), misinterpretation of data, and poorly executed responses can jeopardize AVs' reliability and have serious safety consequences in an automated environment (Bila et al., 2017). The failure rate of each component of AVs was synthesized by Bhavsar et al. (2017). To this end, each component of the ADS and ADAS was examined individually, and the failure rate was determined for each component based on the evidence from the existing literature. The researchers developed a hierarchical model to synthesize AV failure risks associated with the vehicle and infrastructure. The communication system's failure risks, hardware system (sensor and integration platform failure), and software system were ranked the highest, with 9.5%, 4.2%, and 1.0% failure probability, respectively. The failure probability of an AV involved in a crash with a non-AV was also calculated by multiplying the risk of failure of AVs and the crash probability of conventional vehicles.

2.3.6 AV Safety Effectiveness

AV safety effectiveness can be defined using AV SSMs and crash rates. For example, the safety effectiveness of AVs can be estimated as (Equation 2.1:

$$\text{Safety Effectiveness} = 1 - \frac{\text{AVs}'\text{crash rate}}{\text{Conventional vehicles}'\text{ crash rate}} \qquad \text{(Equation 2.1)}$$

However, decisions about AV safety effectiveness or AV safety validity cannot be based on the results of a single study because results typically vary from one study to the next (see Sections 4.3.3 and 4.3.4 for more details). Rather, a mechanism is needed to synthesize data across studies. Wang et al. (2020) synthesized the results of previous simulation and field experiments that estimated safety effectiveness by performing a meta-analysis of 89 studies. They estimated the safety effectiveness of nine ADASs, in descending order: intersection movement assists, pedestrian collision and mitigate (PCAM), lane-departure warning (LDW), lane-change warning (LCW), forward collision warning (FCW), electronic stability control (ESC), blind-spot warning, automated emergency braking (AEB), and adaptive cruise control (ACC).

Wang et al. (2020) further designed a target crash population study to implement the estimated ADASs' safety effectiveness rates and quantify the potential impacts of CVs and AVs on different crash types. The results of their analyses showed that 3.4 million crashes could be prevented between 2012 to 2016; this figure represented a significant reduction in crashes in India (54.24%), Australia (51.55%), the United States (48.07%), New Zealand (45.36%), Canada (44.71%), and the UK (40.95%).

2.4 Comparing AV Safety Evaluation Approaches

Six approaches for quantifying AV safety were identified in this review. The identified approaches were investigated in terms of their input, output, and level of safety implications they address. The identified approaches' inputs included predefined information on AVs' functionality, conventional vehicle crashes, AV road test crashes and errors, study-specific observations, and assumptions and speculations regarding AV implementation. This review showed that the target crash population approach could be used to estimate the number of preventable crashes for evaluating AV safety at the transportation system and society level. Road test data analysis, which mainly focuses on AV crashes' characteristics, compares system failure and crash frequencies of AVs with conventional vehicles. The road test data analysis approach can be used for evaluating AV safety at the transportation system and society levels. Driving simulators and traffic simulation studies can be used for evaluating AV safety in terms of SSMs under different implementation scenarios. While driving simulators investigate AV safety and its potential operational challenges (e.g., disengagement from ADS) at the vehicle level, traffic simulation studies consider AVs' performance and their interactions with other vehicles in a fleet at the transportation system level. The driving simulator studies also unveil some information regarding the user's behavior, such as car-following behavior, that is later used as an input in the traffic simulation studies. AVs' safety effectiveness is estimated as a result of synthesizing the simulator and simulation studies and statistically analyzing their outputs. Although safety effectiveness was defined for ADASs in the literature, this method can be used to evaluate the safety of ADS as well. The estimated safety effectiveness (from traffic simulations or driving simulators) is then used to provide insights into AV safety at both the transportation system and society levels. The system failure assessment approach can evaluate AVs' safety at the vehicle level in terms of the system components' failure rate. Figure 2.3 summarizes the inputs, outputs, and potential applications of the identified approaches.

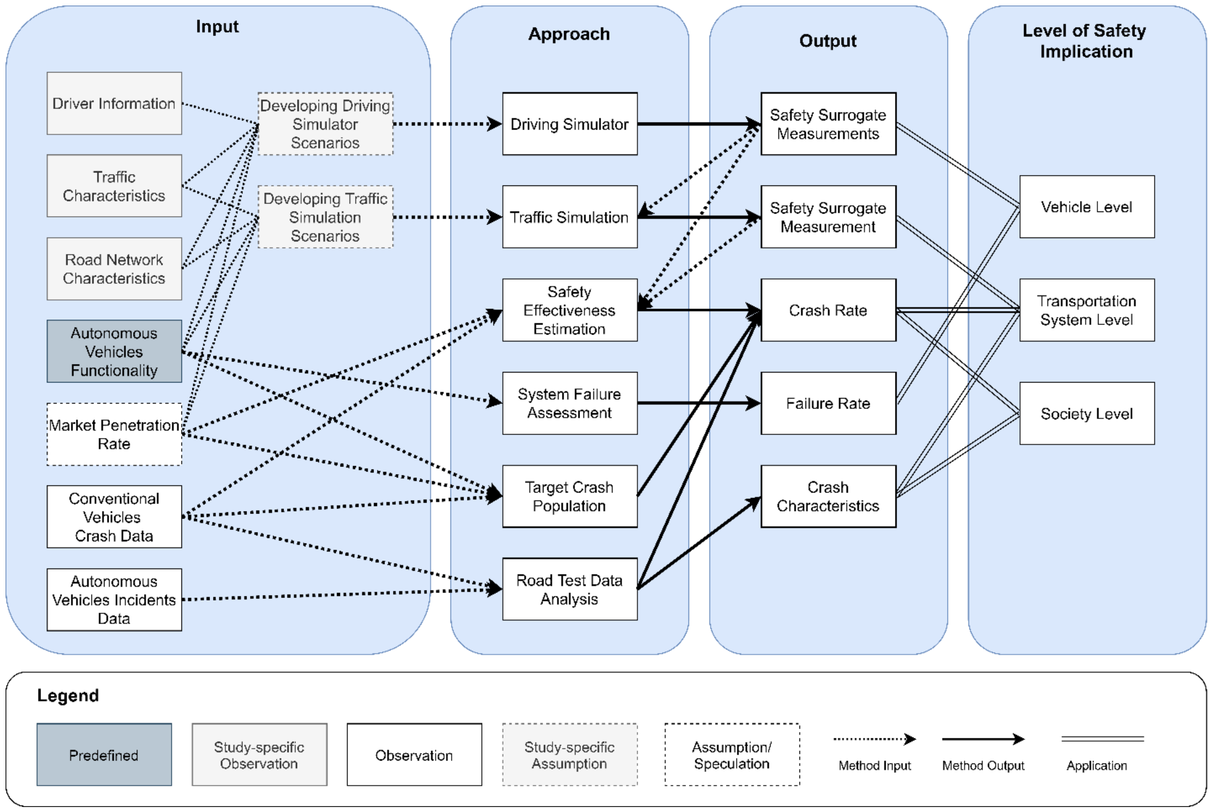


Figure 2.3. Summary of the inputs, outputs, and potential application of AV safety quantification methodologies (Reprinted with permission from Sohrabi et al. 2021)

The AV safety quantification approaches vary in terms of (a) availability of input data, (b) suitability for evaluating different levels of automation, and (c) reliability of estimations. Figure 2.4 shows the trade-off between AV safety quantification methods based on their relative capabilities in terms of these three criteria. This qualitative analysis is based on a comprehensive review of the literature and a detailed evaluation of each approach's capabilities rather than quantitative analyses.

The road test data analysis method is able to evaluate the safety of higher levels of automation with minimal uncertainty; however, it requires extensive and reliable AV crash data. The target crash population method needs relatively fewer input data and can estimate the safety benefits of lower automation levels; nevertheless, considerable uncertainty exists in the estimates. The traffic simulation, driving simulator, safety effectiveness, and system failure assessment approaches can be used to evaluate all levels of automation.

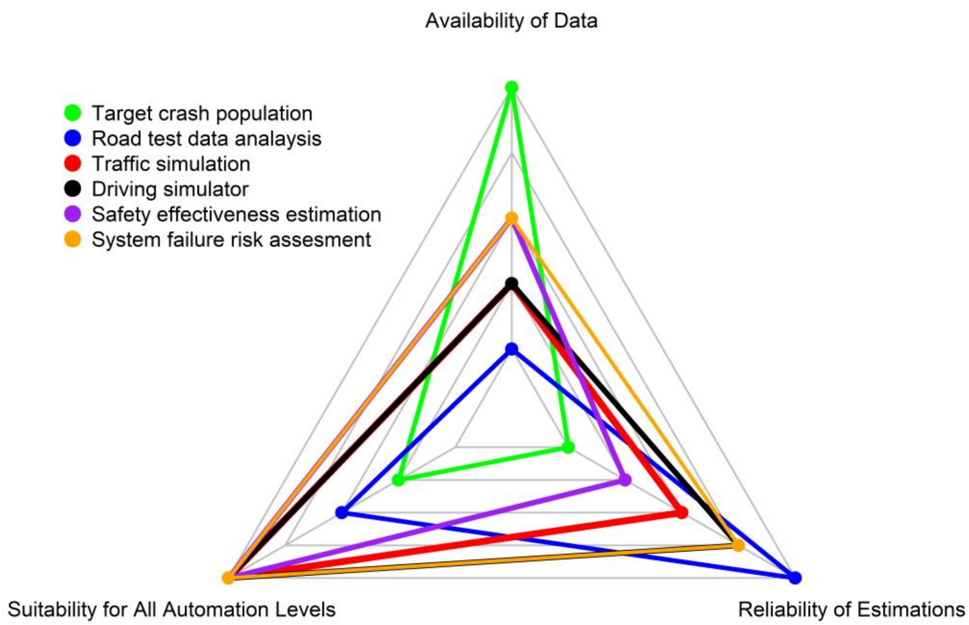


Figure 2.4. Trade-offs between relative availability of data, suitability for evaluating levels of automation, and reliability of estimations (Reprinted with permission from Sohrabi et al. 2021)

2.5 AV Safety Evaluation Challenges

We identified four challenges to AV safety evaluation:

1. Limitations in the existing quantification methodologies
2. Uncertainties in AV implementations and their impacts on AV safety
3. Potential riskier behaviors of AV passengers as well as other road users
4. New safety issues related to AV implementations

2.5.1 Limitation in the Existing Quantification Methodologies

Certain limitations in existing AV safety quantification methodologies can jeopardize the safety evaluation of this new technology. The target crash population studies did not account for the risky scenarios that AVs might cause (e.g., disengagement or system failure) and totally disregarded probable new crashes. The mixed traffic safety issues (interaction of AVs and conventional vehicles) and the way an AV driver reacts to hazards were not considered in the target crash population methodology as well. Thus, this method is expected to represent a theoretical upper bound (or optimistic estimations) of AVs' potential safety benefits, as opposed to their expected actual benefits.

Driving simulator studies were designed to evaluate AVs' potential safety challenges. Traffic simulation studies can also be used to account for both AV and conventional vehicles' driving behaviors and mixed traffic safety issues. Nevertheless, driving simulators and traffic simulation studies have certain limitations. They are subject to biases from a variety of sources, such as participants (e.g., driving behavior and fatigue), simulator and simulation environment (e.g., physical fidelity and functional fidelity), and SSM selection. Employing different SSMs to evaluate AV safety in simulators and simulations makes it almost impossible to directly compare the literature, although a general comparison in terms of the overall safety trend of AVs could be conducted using SSMs. Another challenge in simulator and simulation studies is the limitations in calibration and validation of experiment results since AV road test data—which is the ground truth data—are limited. Because safety effectiveness estimations are based on the results of simulation and simulator studies, they carry remarkable uncertainty as well.

The system failure assessment methodology was used to quantify the crash risks associate with the failure probability of ADASs/ADSs technologies. However, looking at the system failure rates individually can result in overestimating AV failures, given that other components can compensate for the failure of the deficient components. For example, in the event of an AV radar malfunction, the camera vision can help to activate the collision prevention system and avoid a collision. Moreover, system failure assessment relies on system failure rates from private companies. Collecting accurate system failure rates is challenging since this information should be collected from the manufacturer and might be underreported.

The road test data analysis was purported to be the most reliable method for evaluating AV safety. However, existing road tests are limited, and more data are required to draw reliable conclusions on AV safety. Accounting for AVs' safety implications at different MPR levels is another limitation of road test data, given that higher MPR cannot be expected in the near future. Also, a decisive comparison between AV and conventional vehicle crashes is subject to accurate and reliable information about the AV testing environment (ODD and fallback-ready user) as well as conventional vehicle crashes (non-reportable crashes). Increases in AV road test analysis studies in recent years (Figure 3) can be associated with larger and more reliable road test datasets. However, quantifying AV safety with road test data has been criticized because they expose road users to road hazards (Kalra, 2017).

2.5.2 Uncertainties in AV Implementations and Their Impacts on AV Safety

AV impacts on transportation go beyond safety impacts. By offering a safer, cheaper, and more comfortable travel option to individuals with disabilities, AVs may induce additional transportation demand and encourage longer trips. AVs can also encourage shifting from public transit and active transportation (walking and cycling) to private cars (Fagnant and Kockelman, 2015). Transportation and land use are tightly linked in urban areas (Rodrigue et al., 2016); consequently, changes in transportation can ultimately result in urban sprawl (i.e., migrating to areas with lower density and consequently spreading a city's boundaries). Urban sprawl increases total VMT (Childress et al., 2015) and negatively influences accessibility in an urban area (Milakis et al., 2017). In addition, the uncertainties in AVs' intention of use and disproportionate ownerships will affect transportation systems, travel patterns, and urban design.

Changes in VMT and modal shifts, along with the level of MPR, are factors that can impact traffic safety at the transportation system and society levels. Therefore, these changes need to be considered in AV safety evaluations to attain accurate insights into AV safety implications. Full-chain assessment of AV safety—including AV adoption modeling, urban growth modeling, travel demand modeling, and safety analysis—can be a potential avenue to address the uncertainties associated with AV implementations in the transportation system, travel patterns, and urban design.

2.5.3 The Potential Risky Behaviors of AV Passengers and Other Road Users

Changes in AV and conventional vehicle users' behavior need to be considered in AV safety evaluation. Based on research conducted by AAA Foundation, a substantial minority of early adopters of braking assistance systems reported having had a crash or near-crash while driving a vehicle without this technology, supposedly because of incorrect expectations from the unequipped vehicle to provide warnings (Jenness et al., 2007). Gouy et al. (2012) ran a driving simulator experiment and showed that the conventional vehicles would be driven more aggressively if joining a platoon of AVs. The riskier behavior of drivers during interaction with AVs can be explained by the *risk homeostasis hypothesis* (Wilde, 1998). Based on this hypothesis, every person has an acceptable amount of risk that they find tolerable. According to Wilde (1998), “If the perceived level of risk in one part of a person’s life changes, they will compensate by either reducing or increasing the risks they take—all in order to maintain an equilibrium of perceived risk.”

2.5.4 New Safety Issues Related to AV Implementations

Cybersecurity is another potential concern related to AV operation because hacking and vehicle misuse can result in catastrophic crashes (Lee, 2017, Taeihagh and Lim, 2018, Cui et al., 2019). A car hacking experiment conducted by (Jafarnejad et al., 2015) demonstrated that electric vehicles could be easily controlled remotely by mobile applications that forced the vehicles to go forward or backward, limited their speed, and so on. In addition, the ethical dilemma associated with AV reactions during unavoidable situations introduces another challenge in AV operation (Goodall, 2014, Awad et al., 2018) that requires further attention. Although AVs’ ethical issues cannot directly impact AV safety evaluation, they concern about the liability of AVs in crashes, which requires judiciary attention.

2.6 Chapter Summary

This chapter has documented the scoping review methodology used for synthesizing the AV safety quantification methods. The chapter first identified and provided an evaluation of the quantification methods and uncovered the gaps and challenges in AV safety evaluation. The AV safety quantification methods were categorized into six groups: target crash population, road test analysis, traffic simulation, driving simulator, safety effectiveness estimation, and system failure assessment. This review showed that existing methodologies for AV safety evaluation carry certain shortcomings and cannot be used for reliable evaluation of AV safety. In addition, the major challenges in AV safety evaluations are discussed, including uncertainties in AV implementations and their impacts on AV safety, potential riskier behavior of AV passengers as well as other road users, and emerging safety issues related to AV implementations. The next chapter describes the proposed methodology for quantifying the safety performance of AVs and its contribution to the literature.

CHAPTER 3

AUTOMATED VEHICLE SAFETY EVALUATION METHODOLOGY

This chapter introduces the proposed AV safety evaluation methodology. In the following sections, first, the motivations behind proposing a new methodology and how it addressed the limitations of AV safety evaluation methodologies (extensively discussed in the comprehensive literature review conducted in Chapter Two) are highlighted. Then, the researcher skims over the survival analysis and its history in crash prediction models. Then, the survival function for the AV safety evaluation problem is formulated, and the proposed AV safety evaluation methodology, and the theory behind it, are then explained.

3.1 Motivations and Contributions to the Literature

In Chapter Two, the comprehensive review of the existing AV safety evaluation methodologies identified six approaches through which AV safety can be quantified. Each approach was further analyzed and evaluated in terms of:

1- required inputs and availability of data,

2- output and metrics through which they measure AV safety,

3- their application for vehicle, system, and society level safety evaluations,

4- their suitability for diffident levels of automation safety evaluation, and

5- the reliability of the estimation.

As a result, the gaps and limitations of AV safety evaluation methodology and challenges in AV safety evaluations were identified. Among the identified evaluation approaches, road test analysis was discussed to result in the most reliable estimations of

AVs safety since it automatically considers AV safety challenges such as riskier behavior of AV users, risks of AV interaction with human-driven vehicles, and system failure risk. To achieve reliable estimations using road test data analysis, AVs are required to be tested extensively on roads. Although AV road tests are permitted in many states, the potential safety concerns associated with AV operation persuades decision-maker to hinder immense road tests of AVs. This “chicken or egg” paradox urges the need for safety evaluation of AVs in the meantime, with limited road test data. The existing road test analysis literature assumes conventional vehicle safety as the benchmark and evaluates AV safety using this benchmark. Conventional vehicles and AV safety were mainly measured using the rate of crashes per VMT. This analysis can be biased given that (1) conventional vehicle crashes are underreported, and comparing AV and conventional vehicle crashes will be unfair, and (2) the rate of crashes provides limited information about AV safety, especially when dealing with small datasets. This resulted in biased, unreliable analysis of AV crashes in the literature.

Another application of road test data is identifying and assessing the factors that contribute to AV crashes. Although the existing econometrics methodologies for crash predictions can examine the impacts of road characteristics and environmental conditions on AV crashes frequency at the road segment level, the impacts of vehicle-level factors remain unclear. For instance, the safety impacts of AV technology improvements can be of interest to manufactures which cannot be considered in traditional econometrics methodologies used in road test analyses. The contribution of AV driver’s characteristics (in levels 1, 2, and 3) to crash frequency is another example in which the existing road segment-level methodologies cannot be used for investigating vehicle-level contributing factors.

Analysis of road test data can be used to support efficient safety effectiveness estimation of AVs, given that the safety challenges of AV operations are reflected in road test data. In this case, the road test data analysis can be used for AV safety evaluations not only at the vehicle level but also transportation system and society level.

In this research, the limitations in the existing AV safety evaluation and road test analysis method are addressed by rethinking the AV safety evaluation problem and proposing a new safety performance evaluation methodology based on survival analysis. The proposed methodology could be used for evaluating AV safety with limited data. On the basis of the proposed methodology, new metrics can be defined to estimate the safety effectiveness of AVs and support their safety validations. The new methodology can further expand to study the impacts of vehicle-level contributing factors to AV crashes, such as safety technologies.

3.2 Background

This section contains an introduction to survival analysis and its application in traffic safety and road crash analysis.

3.2.1 An Introduction to Survival Analysis

The first survival analysis can be traced back to the 17th century, where it was organically used in biomedical science, and then, it has been used in many disciplines, including engineering (Liu, 2012). Survival analysis refers to methods to investigates the length of time to occurrence of an event in a specific observation period. *Survival* is the process or the life span from a specific starting time to an *event*. While events in survival analysis evoke morbidity or mortality, they were also defined as divorce or collapse of the political system in social science and system failure or product deficiency in engineering. An example of survival analysis can be exploring the association between breast and coping strategies among Black and white women (Reynolds et al., 2000). In this example, patients' survival is studied in an observation period interval, and the event can be defined as deaths from breast cancer.

More formally, survival analysis enables us to make predictions not only based on events but also "time-to-event." Unlike comparing the rate of an event between two datasets, survival analysis explores another layer of information, the time of a particular event. Suppose that in the breast cancer survival example, Black and white patients have similar rates of death in an observation period, but the events are observed sooner among Black patients. While the two datasets are similar in terms of the rate of event, considering the time-to-event would distinguish these two datasets. This implies that survival analysis considers information about the event frequency as well as the time that event occurred. If we consider an event as a change in status, survival prediction models resemble the qualitative choice analysis models, such as logistic, logic, and probit models (Train, 2009). However, in survival analysis, another layer of information, time-to-event, is incorporated in predictions.

Back to the women breast cancer example, the survival of patients was examined in a specific observation period. Therefore, the survival analysis includes the events that occur within the observation period, although some patients may die after. In this case, the patient information cannot be included in the survival analysis. Another example can be losing contact with the patient throughout the longitudinal survival analysis. The loss of observation in the survival analysis is referred to as *censoring*. Censoring would result in incomplete survival data, which adds complexities to survival analysis.

To this point, the discussion about survival function was limited to a homogeneous population. But the survival function needs to be adjusted in a heterogeneous population using explanatory variables. In the women's breast cancer example, the patients have different characteristics (age, race, tumor stage, study location, etc.) that can be associate with the event. The time to event, in this case, can be characterized using such factors. In this case, age, race, tumor stage, and study locations will be the explanatory variables that can address the heterogeneity in the population.

3.2.2 Survival Analysis in Traffic Safety

Survival analysis in the context of traffic safety and crash analysis can be traced back to the 80s, where Jovanis and Chang (1989) studied the probability of accident occurrence on individual trips. They further defined a general structure for studying accident occurrence using survival analysis (Chang and Jovanis, 1990). In the 90s, Lin et al. (1993) explored the safety impacts of driving-hour regulations on less-than-truckload carriers, and Mannering (1993) examined the role of gender in crash risk using survival analysis. Since then, survival analysis was revisited for different types of traffic safety analyses—including investigating drink and drive events (Ferrante et al., 2001), assessing pedestrian risk exposure at signalized intersections (Tiwari et al., 2007), exploring contributing factors to driving-under-influence crashes (Fu, 2008), examining the risk of driving for older drivers (Caragata Nasvadi and Wister, 2009), comparing intersections and local road crashes (Bagloee et al., 2016), contributing factors to motorcycle crashes (Chen et al., 2018, Balusu et al., 2020) and contributing factors to sever crash events (Xu et al., 2018). More recently, Xie et al. (2019) have used survival analysis for evaluating the impacts of safety treatments in before and after studies. Table 1 represents a summary of survival analysis in the context of traffic safety.

Survival analysis in the context of traffic safety was limited to road user (driver or pedestrian) (Mannering, 1993, Ferrante et al., 2001, Tiwari et al., 2007, Caragata Nasvadi and Wister, 2009, Balusu et al., 2020), crash type (Fu, 2008, Xu et al., 2018), and road infrastructure (Bagloee et al., 2016, Xie et al., 2019) perspective. Also, in the previous time-to-event analyses, “time” and “event” were defined based on the research questions. Time (mainly in the form of the number of days to an event) was mainly used (Jovanis and Chang, 1989, Lin et al. 1993, Mannering, 1993, Ferrante et al., 2001, Tiwari et al., 2007, Fu, 2008, Caragata Nasvadi and Wister, 2009, Xu et al., 2018, Xie et al., 2019, Balusu et al., 2020) to represent the longitudinal data while Bagloee et al., (2016) considered the distance of crash to the intersection as the longitudinal parameter in survival analysis. The events were considered as the occurrence of crash (Jovanis and Chang, 1989, Lin et al. 1993, Mannering, 1993, Ferrante et al., 2001, Fu, 2008, Caragata Nasvadi and Wister, 2009, Bagloee et al., 2016, Xu et al., 2018, Xie et al., 2019, Balusu et al., 2020), unsafe intersection passing events (Tiwari et al., 2007), Driving Under Influence (DUI) arrest (Ferrante et al., 2001).

3.3 Survival Process of a Vehicles

While the literature about survival analysis in the context of traffic safety is limited to the road user-level, crash type-level, and road infrastructure-level analyses, survival analysis can be used for vehicle-level analysis, considering patients as vehicles. In this case, the survival of a vehicle can be studied, and those vehicles that cannot be tracked will be censored. In the context of AV safety evaluation, vehicle-level survival analysis would be challenging given the limitation in the available data. To overcome such issues and the need for censored survival analysis and its complexities, we can aggregate individual vehicles to the type of vehicle and assume the lifetime is shorter than the duration of the experiment. If we consider crashes as the events and VMT as the time, time-to-event can be defined as miles-to-crash. Suppose X is the number of miles to crash with a distribution function $f(x)$. The cumulative distribution of $f(x)$, $F(x) = \Pr(X < x)$, represents the probability that a crash occurred before x miles. The probability of survival beyond x miles can then be defined Equation 3.1 (Liu, 2012):

$$S(x) = \Pr(X > x) = \int_x^\infty f(t)\mathrm{d}t \qquad \text{(Equation 3.1)}$$

where $S(x)$ is the survival function. Since X is a continuous variable, the survival function is a strictly decreasing function. The cumulative distribution function of MTC is the complement of the survival function (Equation 3.):

$$F(x) = 1 - S(x) \qquad \text{(Equation 3.2)}$$

If we consider that survival function represents the safety reliability of vehicles, then $F(x)$ can be referred to as the *failure function*. Unlike the survival function, the failure function is strictly increasing. Failure function represents the likelihood of a vehicle to be involved in a crash after being driven for a given number of miles. Hereafter, the main discussion is around failure functions rather than survival functions for the sake of more tangible interpretations.

3.3.1 Hazard function

The instantaneous rate of failure (crashes) can be estimated using the hazard function. The hazard function, $h(x)$, can be estimated as the instantaneous rate of failure relative to the survival rate at time x (Washington et al., 2020):

$$h(x) = -\frac{1}{S(x)}\frac{\mathrm{d}S(x)}{\mathrm{d}x} = \frac{f(x)}{S(x)} \qquad \text{(Equation 3.3)}$$

Interpreting the hazard rate can be challenging since the vehicles are expected to maintain a constant hazard rate λ throughout their operation. This implies that the exponential survival function (with constant hazard rate) can accurately and sufficiently characterize the survival function. In the context of AVs, however, there are certain scenarios under which the hazard rate can be a function of mile driver $\lambda(x)$, i.e., it can change over the number of miles driven by the vehicle. For example, one can resemble the learning process of AVs with a human driver. Over time as the number of miles driven increases, the human driver is expected to gain more experience and drive safer, as do AVs; hence the hazard rate decreases. Therefore, we can characterize the learning curve for AVs by hazard function (a function of the number of miles driven). Also, from the manufacturing life cycle point of view, defective vehicles will be failed early on in the life cycle. Once they are removed from the testing sample, the hazard rate decreases over time.

It can be shown that for hazard rate $\lambda(x)$, the survival function will be $e^{-\int_0^x \lambda(y)dy}$ (see (Klugman et al., 2012) for derivation). The failure function and the distribution of the number of miles to crash can also be derived as (Klugman et al., 2012):

$$F(x) = 1 - e^{-\int_0^x \lambda(y)dy} \qquad \text{(Equation 3.4)}$$

$$f(x) = \lambda(x)e^{-\int_0^x \lambda(y)dy} \quad \text{(Equation 3.5)}$$

Respectively, the hazard function can be estimated using Equation 3.3.

3.3.2 Parametrized survival function

While empirical (non-parametric) distributions are informative and flexible, the parametric distribution can help to describe a theoretical problem. It would be specifically helpful when dealing with a smaller dataset, which its parametrization will result in a smooth function.

The maximum likelihood estimation (MLE) method can be used for estimating the parameters of the parametric failure function. Suppose that random variables $x_1, x_2, \dots, x_n$ have joint probability function $f(x_1, x_2, \dots, x_n|\theta)$, where θ represents the parameters of the density function. Then, the likelihood of θ as a function of $x_1, x_2, \dots, x_n$:

$$L(\theta) = f(x_1, x_2, \dots, x_n|\theta) \quad \text{(Equation 3.6)}$$

The MLE of θ would be the value that maximizes $L(\theta)$. For observed values $X_i = x_i$, where $i = 1, 2, \dots, n$, if X_i is identically independently distributed ($i.i.d.$), then the joint probability function is the product of marginal probabilities:

$$L(\theta) = \prod_{i=1}^{n} f(X_i|\theta) \quad \text{(Equation 3.7)}$$

To solve the maximization problem and find θ, the log-likelihood function will be maximized:

$$l(\theta) = \sum_{i=1}^{n} \log\left[f(X_i|\theta)\right] \quad \text{(Equation 3.8)}$$

The goodness-of-fit (GOF) of the fitted parametric distribution to the empirical failure functions can be examined in terms of Akaike information criterion (AIC) and Bayesian information criteria (BIC), and visuality using quantile-quantile plot (Q-Q plot) and probability-probability plot (P-P plot). Moreover, the fitted distribution functions should be statistically significant, which is examines using hypothesis tests. Section 3.3.3 discusses the hypothesis testing in more detail.

The choices of parametric distribution functions should be in line with the theories behind the changes in hazard rate over time. As discussed earlier, for conventional vehicles, the hazard rate cannot change over time and therefore is assumed to be constant. In this case, the distribution of the number of miles to crash can be shown to follow an exponential distribution. On the other hand, for AV, the hazard rate can vary over time, and therefore, the survival function can be characterized by Weibull, log-normal, log-logistics, and gamma distribution functions.

3.3.3 Hypothesis test

To validate AV safety, we can test the hypothesis that automated and conventional vehicles' failure functions are statistically consistent. Two powerful statistical tests, Kolmogorov-Smirnov (K-S) test and Anderson-Darling (A-D) test, are suggested to test this hypothesis (Razali and Wah, 2011). K-S and A-D tests are two sample tests that can be used for comparing both parametric and non-parametric failure functions.

The hypothesis that whether the AVs parametric failure function $F_{AV}(x)$ is the same as the conventional vehicles failure function $F_{CV}(x)$ or not need to be examined, i.e., the null hypothesis being that the two parametric distributions are identical:

$H_0: F_{AV}(x) = F_{CV}(x)\ for\ all\ x$
$H_A: F_{AV}(x) \neq F_{CV}(x)\ for\ some\ x$

According to the K-S test, the test statistic would be the maximum distance of two distributions:

$$D_{1,2} = \operatorname*{Sup}_x \left|F_{AV}(x) - F_{CV}(x)\right| \qquad \text{(Equation 3.9)}$$

where "Sup" stands for Supremum function. The K-S test statistic $D_{1,2}$ is then compared to critical values $D_{1,2,\alpha}$ for desired significance level. The critical values are estimated from the Kolmogorov distribution. If $D_{1,2}$ exceeds the $D_{1,2,\alpha}$, then the null hypothesis (H_0) can be rejected and we conclude the $F_{AV}(x)$ and $F_{CV}(x)$ are different distributions.

Unlike the K-S test, the A-D test finds the difference between two distributions giving more weight to the differences between the tails of distributions $F_{AV}(x)$ and $F_{CV}(x)$. While the test hypothesis is defined similar to the K-S test, the difference between the distributions can be defined as (Anderson, 2011):

$$W_n^2 = n \int_{-\infty}^{\infty} \left[F_{AV}(x) - F_{CV}(x)\right]^2 \psi\left(F_{CV}(x)\right) dF_{CV}(x) \qquad \text{(Equation 3.10)}$$

where $\psi(z)$ is the weight function such that $\psi(z) > 0$ and $\psi = \left[F_{CV}(z)\left(1 - F_{CV}(z)\right)\right]^{-1}$. When $U = F(x)$ is a random variable with distribution function $u = \Pr\,(U < u = \Pr(F(x) < u\}\},\ 0 \leq u < 1$, Anderson and Darling (1954) showed that the Equation 10 can be written as:

$$A_n^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)[\log u_{(j)} + \log\left(1 - u_{(n-j+1)}\right)] \qquad \text{(Equation 3.11)}$$

where $u_{(j)} = F_{CV}(x_{(j)})$ and $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ is the ordered sample. The A-D test critical values are estimated for different distribution functions (Jäntschi and Bolboacă, 2018). Similar to K-S test, null hypothesis is rejected if A-D test statistics (A_n^2) exceed the critical values.

In addition, the suggested hypothesis test can be used to examine the GOF of the parametric distribution functions to the empirical distribution functions. The fitted functions should be statistically significant.

3.3.4 Automated vehicle safety effectiveness

In addition to graphical comparison and interpretation of the survival curves, Restricted Mean Survival Time (RMST) is another informative metric used in the survival analysis literature to compare survival functions (Royston and Parmar, 2013, Harhay et al., 2018). The RMST of a random variable T, $\mu(x^*)$, is the expected value of $\min(X, x^*)$—i.e., the area under the survival curve $S(x)$ up to x^*:

$$\mu(x^*) = E(\min(X, x^*)) = \int_0^{x^*} S(x)\mathrm{d}x = \int_0^{x^*} (1 - F(x))\mathrm{d}x \qquad \text{(Equation 3.12)}$$

Since X is the number of miles to a crash, RMST can be interpreted as the *no-crash expectancy* until x^* miles. For example, AV no-crash expectancy in the next 1 million miles can be 0.5 million miles, which means no-crashes are expected after 0.5 million miles of driving in the next 1 million miles of AV operation. The no-crash expectancy can be defined similarly for conventional vehicles. To compare the conventional vehicle and AV safety at x^*, RMST is estimated for, and their ratio is calculated. For entire AV and conventional vehicle operations, when $X \to \infty$, the calculated ratio resembles the safety effectiveness (SE) of AVs in comparison to conventional vehicles:

$$SE = \frac{RMST_{AV}}{RMST_{CV}} = \frac{\mu_{AV}(X)}{\mu_{CV}(X)} = \frac{\mathrm{E}_{AV}(X)}{\mathrm{E}_{CV}(X)} = \qquad \text{(Equation 3.13)}$$

$$= \frac{AV\ no\text{–}\ crash\ expentency}{CV\ no\text{–}\ crash\ expentency}$$

where X is the miles to crash (MTC), $\mathrm{E}_{AV}(X)$ is expected value of AV's MTC and $\mathrm{E}_{CV}(X)$ is the expected value of conventional vehicles MTC. If SE is larger than 1, then it is assumed that AVs are safer; otherwise, conventional vehicles are safer.

3.3.5 Automated vehicle crash contributing factor

The contributing factors to AV crashes can be investigated at the vehicle level road segment level. The survival analysis allows us to conduct vehicle-level safety analysis. However, at the road segment level, considering hazard rate as a function of time would violate the assumptions behind existing crash prediction models in the road safety literature (Lord et al., 2005, Lord and Mannering, 2010). We further discuss each level of AV safety analysis in the subsequent sections.

3.3.5.1 Vehicle-level safety analysis

The hazard function can also be estimated as the ratio of the condition probability at x given the condition $X \geq x$ over an infinitesimal time change (Liu, 2012):

$$h(x) = \lim_{\Delta x \to 0} \frac{\Pr\{X \in (x, x + \Delta x] | X \geq x\}}{\Delta x} \qquad \text{(Equation 3.14)}$$

Therefore, the hazard rate is the conditional probability of failure with respect to the limit of a time interval.

Now, let us associate the failure at mile x with a vector of explanatory variables $\boldsymbol{Z} = (Z_1, \dots, Z_p)$, where $\boldsymbol{Z}$ includes vehicle characteristics such as vehicle make, vehicle age, driver skill, vehicle safety technology, etc. The effect of explanatory variables can be captured by classical linear regression, modeling the natural logarithm of the survival time $Y = \ln(X)$. However, the linear regression approach requires assumptions regarding the distribution of survival time. An alternative approach can be modeling the condition hazard rate as a function of the explanatory variables. In this case, the condition hazard rate with covariate vector $\boldsymbol{z}$ is a product of baseline hazard rate $h_0(x)$ and non-negative function of covariates (Liu, 2012):

$$h(x|\boldsymbol{z}) = h_0(x)c(\beta^t \boldsymbol{z}) \qquad \text{(Equation 3.1)}$$

where β^t is the coefficient of covariates (explanatory variables). In these models, the hazard rate of two individuals with a distinct value of $\boldsymbol{z}$ is proportional at mile x. For instance, for covariate values $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ we have the constant and independent of time ratio of:

$$\frac{h(x|\boldsymbol{z}_1)}{h(x|\boldsymbol{z}_2)} = \frac{h_0(x)c(\beta^t \boldsymbol{z_1})}{h_0(x)c(\beta^t \boldsymbol{z_2})} = \frac{c(\beta^t \boldsymbol{z_1})}{c(\beta^t \boldsymbol{z_2})} \qquad \text{(Equation 3.16)}$$

In the Cox proportional regression, the link function $c(\quad)$ is considered as an exponential function (a monotonic increasing function) (Liu, 2012, Washington et al., 2020). Unlike classical regression models, the Cox proportional hazards model is a semi-parametric model, with no assumptions about the shape of the baseline hazard function. However, other assumptions such as independence and linear association between the natural logarithm of the hazard rate and covariates exist.

3.3.5.2 Road-segment level safety analysis

It is shown that the fundamental crash process follows a Bernoulli trial with an unequal probability of independent events, also known as the Poisson process (Lord et al., 2005). Three postulates of Poisson processes are listed below:

1- The changes occurring in non-overlapping intervals are independent.
2- The probability of two or more changes taking place in sufficiently small intervals is essentially zero.
3- The probability of exactly one change in a short interval $(x, x+\delta)$ is approximately $\lambda\delta$ where δ is sufficiently small and λ is positive constant.

If we relax the third assumption and assume the hazard rate λ as a function of the number of miles $\lambda(x)$, then the AV crash occurrence can be considered as a non-homogeneous Poisson process. In this case, the ubiquitous Poisson and Negative binomial regression models can be effectively used for AV crash prediction at the road-segment level. For instance, if the mile-to-crash distribution follows the Weibull distribution, the hazard rate function will become $\lambda(x) = \frac{\alpha}{\beta}(\frac{x}{\beta})^{\alpha-1}$. In case α = 1, the hazard rate becomes constant and AV crash occurrence follows the Poisson process. When the parameter α < 1, the failure rate decreases over time, which can represent the learning curve of AV over the number of miles traveled or be used to model life cycle of AV from the manufacturing perspective.

3.4 Chapter Summary

This chapter introduced the proposed AV safety methodology on the basis of survival analysis. The proposed methodology addresses some of the limitations and gaps in AV safety evaluations. First, the proposed model adds a new layer of information (time of the crash) to the analysis that addresses the limited availability of road test data and uncertainties in their analyses. Second, in light of rethinking AV safety evaluation using survival analysis, a new metric is defined, no-crash expectancy, which can support AV safety effectiveness estimations for system-level and society-level safety evaluations. Third, the posed methodology can be used for vehicle-level crash contributing factor analysis. Fourth, we showed that the existing road safety analysis methods must be revisited in the era of automation to account for the decreasing AV hazard rate over time (number of miles the vehicle is operated). An empirical study is designed to examine the proposed methodology and evaluate the safety of existing AVs under road tests while conducting a fair comparison between AV and conventional vehicle crash frequencies. The empirical study is discussed in the next chapter.

CHAPTER 4

EMPIRICAL STUDY DESIGN

In this chapter, the researcher describes the rational motivation and rationale behind designing an empirical study for AV safety evaluations. This chapter contains an introduction about the data used in the empirical study, including an overview of the available AV crash data is presented along with the source of conventional vehicles crashes data. The researcher further elaborates on the designed empirical study architecture, including the process of creating the datasets, examining the proposed methodology, and comparing AV and conventional vehicle safety using the proposed methodology.

4.1 Motivation

An empirical study is designed to evaluate AV safety with two objectives. First, the empirical study targets the false equivalency between automated and conventional vehicles crashes stemmed from the limitations in the availability of conventional vehicles' non-police-reportable crashes (discussed in chapters 2 and 3). To this end, the fallacy in comparison between automated and conventional vehicles crash rates is addressed by sourcing the conventional vehicle crashes from the NDS database, which included both minor and major crashes―hence making it comparable to AV crashes reported by the automakers. Second, the proposed methodology is examined using comparable automated and conventional vehicle crash datasets. The limitations in data availability do not allow applying the proposed method thoroughly, e.g., the applicability of Cox proportional regression for exploring the contributing factors to the crashes and AVs hazard rate estimations. Nevertheless, the applicability of survival analysis for AV safety evaluations is investigated.

Given the scarcity of AV test drives and lack of transparency in AV road tests, the empirical study faced multiple challenges. The NDS data are available for purchase upon the Institutional Review Board (IRB) approval. The data availability and data sources are explained in the subsequent sections.

4.2 AV Crash Data

Automated driving road tests have been growing in recent years. According to NHTSA's AV test initiative, 25 AV manufacturers and developers are testing their cars on United States public roads[6]. As shown in Figure 4.1, AVs are tested in 21 states (have submitted AV test information to NHTSA). A total number of 93 AV test sites are recognized by NHTSA, which the majority of sites are public streets (Figure 4.2a). Shuttles and cars are mainly testes on these sites (Figure 4.2b).

[6] Sourced from: https://www.nhtsa.gov/automated-vehicle-test-tracking-tool (May 2021)

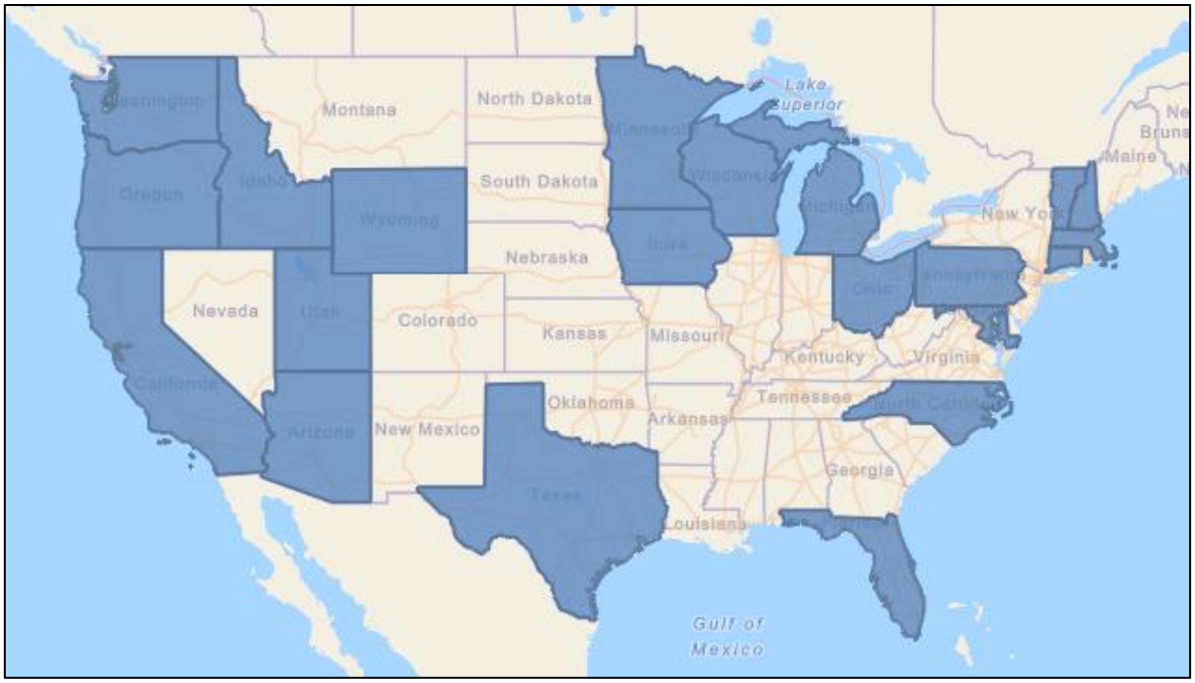
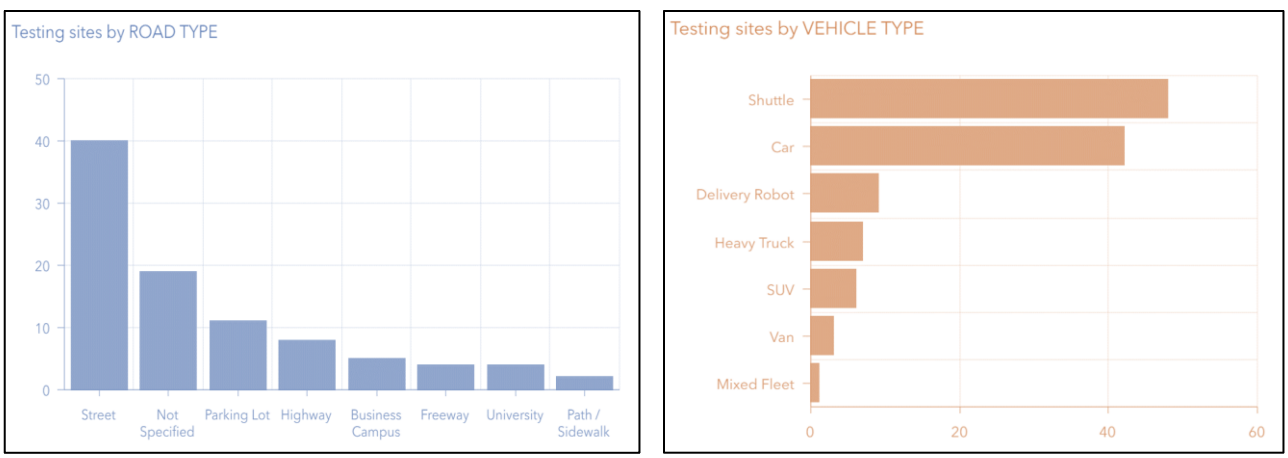


Figure 4.1. States where AVs were tested on public roads (source: NHTSA AV test initiative)



(a) **(b)**

Figure 4.2. AV test sites (a) road types and (b) vehicle types

Although AV companies are not required to report their vehicle information based on federal rules, state regulations can help to keep road tests transparent or make data available for the public. For example, the California Department of Motor Vehicles (CA

DMV) Autonomous Vehicle Tester (AVT) program in 2014 with the aim of testing autonomous vehicles with fallback users (test vehicles require a human in the driver seat who can take control of the vehicle at any time). According to this program, all manufacturers testing AVs on public roads are mandated to report crashes involving an AV within ten days after the collision.

AV crash data are sourced from the CA AVT program. Crashes that occurred in one year of AVs operation on CA public roads are investigated. As of November 2020, 59 permit holders are testing their AV under this program. CA AVT program defined AVs as "a vehicle that has been equipped with technology that is a combination of both hardware and software that, when engaged, performs the dynamic driving task, but requires a human test driver or a remote operator to continuously supervise the vehicle's performance of the dynamic driving task."[7] According to this definition, vehicles equipped with one or more ADAS (Levels 1 and 2 of automation) are not tested in this program, and the AVT program is limited to testing level 3 of automation.

Based on CA AVT regulations, AV companies are mandated to report AV-involved crashes in fewer than ten days after the time of the crash, and so it can be assumed that the crash reports contain all AV-involved crashes. The crash reports consist of information regarding the crash time, cause of the crash, crash type, crash severity, and whether the crashes occurred under ADS operation or manual driving. Also, the annual mileage of each vehicle's operation on public roads must be reported by the end of the year. The mileage dataset includes the vehicle identification number (VIN) and the number

7 Sourced from https://www.dmv.ca.gov/portal/uploads/2020/06/Adopted-Regulatory-Text-2019.pdf (May 2021)

of miles it was operated during each month. No information regarding the environment under which AVs were test is publicly available.

As of the time of developing this study, AV road test data were available until November 2020. Figure 4.3 represents the distribution of crashes and VMT from January 2019 to November 2020. The data for the year 2019 is used, given the restriction in AV testing, and road traffic in general, because of the global pandemic in the year 2020.

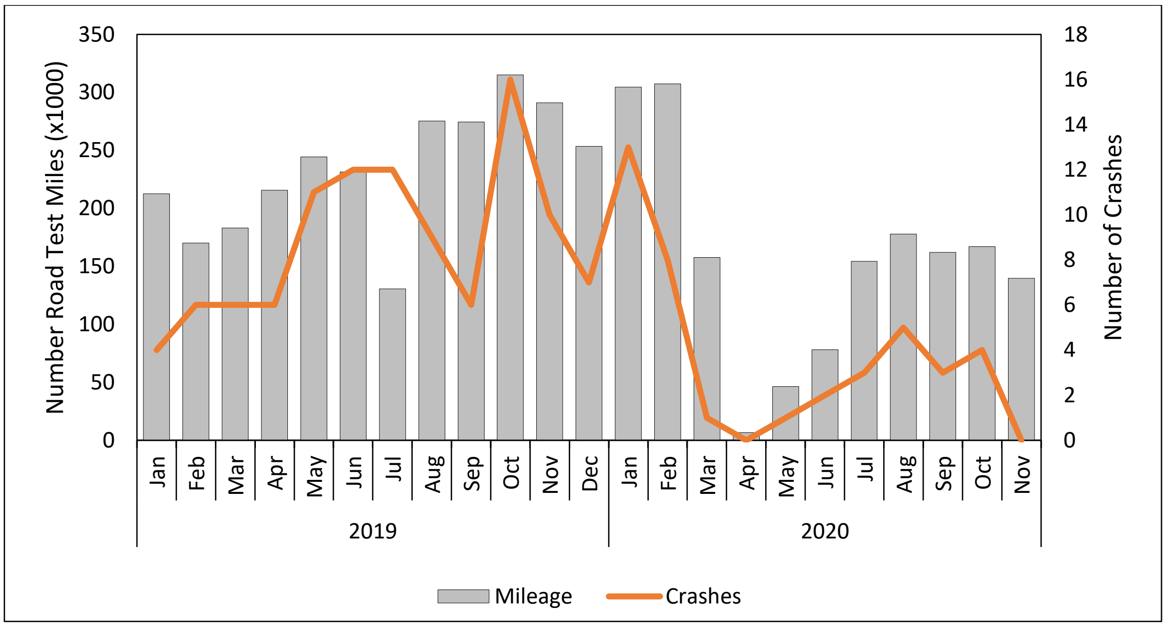


Figure 4.3. Distribution of AV crashes and VMT in 2019 and 2020

AV testing data in 2019 includes 651 unique VIN from 30 AVT permit holders. The tested AVs were driven 2,849,850 miles in 2019 and were involved in 105 crashes. The crash data were manually extracted from the crash reports. Table 4.1 represents the share of companies in testing AVs and AV crashes.

Table 4.1. AV road test under CA AVT program in 2019

Company	Drive Test Mileage		Number of Crashes	
	Miles	Percentage	Count	Percentage
Waymo LLC	1302109.6	46.6%	25	23.8%
CRUISE LLC	875744.5	31.3%	61	58.1%
PONY.AI, INC.	202476.3	7.2%	2	1.9%
Baidu USA LLC	106243.5	3.8%	0	0.0%
Nuro	72146.8	2.6%	0	0.0%
Zoox, Inc	70458.0	2.5%	8	7.6%
Lyft	46864.1	1.7%	6	5.7%
AutoX Technologies, Inc.	40802.0	1.5%	0	0.0%
Mercedes Benz Research & Development North America, Inc.	16011.4	0.6%	0	0.0%
Aurora Innovation, Inc.	13852.1	0.5%	1	1.0%
Apple Inc.	8192.9	0.3%	1	1.0%
NVIDIA	7179.0	0.3%	0	0.0%
AImotive Inc.	6386.0	0.2%	1	1.0%
WeRide Corp	5920.0	0.2%	0	0.0%
SF Motors, Inc.	3453.6	0.1%	0	0.0%
Drive.ai Inc	3201.4	0.1%	0	0.0%
Nissan North America, Inc	2329.4	0.1%	0	0.0%
Nullmax	2201.0	0.1%	0	0.0%
Qualcomm Technologies, Inc.	2182.9	0.1%	0	0.0%
SAIC Innovation Center	2143.9	0.1%	0	0.0%
Toyota Research Institute	2111.0	0.1%	0	0.0%
Phantom AI, Inc.	1125.0	0.0%	0	0.0%
PlusAI, Inc.	962.0	0.0%	0	0.0%
Udelv, Inc	695.1	0.0%	0	0.0%
Valeo North America Inc.	99.6	0.0%	0	0.0%
BMW of North America	21.4	0.0%	0	0.0%
Telenav, Inc.	21.0	0.0%	0	0.0%
Tesla, Inc.	12.2	0.0%	0	0.0%

Since the crash reports do not include the VIN, the daily AVs VMT is estimated using the monthly VMT reports, assuming no variations in daily VMT in a month. The MTC is then approximated using the daily VMT and the time of the crash.

4.3 NDS data

NDS data are collected as part of the second Strategic Highway Research Program (SHRP2) program. In the SHRP2 study, more than 3,100 volunteer drivers in six locations had their cars equipped with cameras, radar, and other sensors to capture data as they went about their usual driving tasks. The six sites where NDS data were collected are Seattle, WA; Bloomington, IN; Buffalo, NY; State College, PA; Durham, NC; and Tampa, FL. The NDS includes the volunteer driver's information, the vehicle they drive, and their trip information, as well as the potential crash and near-crash events that occurred in each trip. The NDS dataset consists of more than 5.4 million miles, more than 1 million hours of recorded videos, and more than 1,500 crashes.

NDS study consists of four types of data and can be requested using the query tool available on the Insight website. Time series or vehicle kinematics data include the data collected from each instrumented vehicle while it is being driven. Video data include the data collected from the cameras installed in the participant's vehicle. Driver survey and questionnaire data include answers to questionnaires, vision test results, and the results of brief physical tests described in the consent agreement. Event data include the crash, near-crash, and baseline event data. This data also includes follow-up investigations of selected crashes with answers to an interview with the driver by one of the SHRP 2 researchers and the police report resulting from the crash.

In this study, the NDS dataset should be consistent with the AV's crash and VMT dataset in terms of the total VMT. A sample of consecutive trips is randomly selected; as such, the total number of miles driven would be equal to 3 million miles. Consequently, 509,338 trips were included in the dataset, and a total number of 130 crashes were observed in these trips. Table 4.2 reports the characteristics of trips and crashes in the NDS data. The length of trips varies from less than a mile to 382.4 miles, with an average of 6.8 miles. The average number of miles before a crash is calculated as 9.9. Given that the time of crashes is known, MTC is calculated using the trip lengths in the NDS dataset. To this end, the total VMT in the sample for consecutive crash reports is accumulated.

Table 4.2. NDS data trips characteristics

	Min	**Max**	**Median**	**Mean**
Trip Length	0.0	382.4	3.0	6.8
Trips Length before a Crash	0.0	215.6	5.0	9.9

4.4 Empirical Study Design

The empirical study is designed in three steps (Figure 4.4). First, despite the challenges in AV crash availability, a database is created by sourcing AV crashes from CA DMV (as discussed in section 4.2) and combining them with conventional vehicle crashes from NDS. In the second step, the proposed methodology in this study is examined using the created database. In this step, the empirical and parametric failure functions for AV and conventional vehicles are estimated. Third, the safety performance of AV and conventional vehicles are evaluated by (1) testing the hypothesis of whether AV failure function is the same as the conventional vehicle failure functions, and (2) estimating the no-crash expectancy of AV and conventional vehicle and the safety effectiveness of AVs.

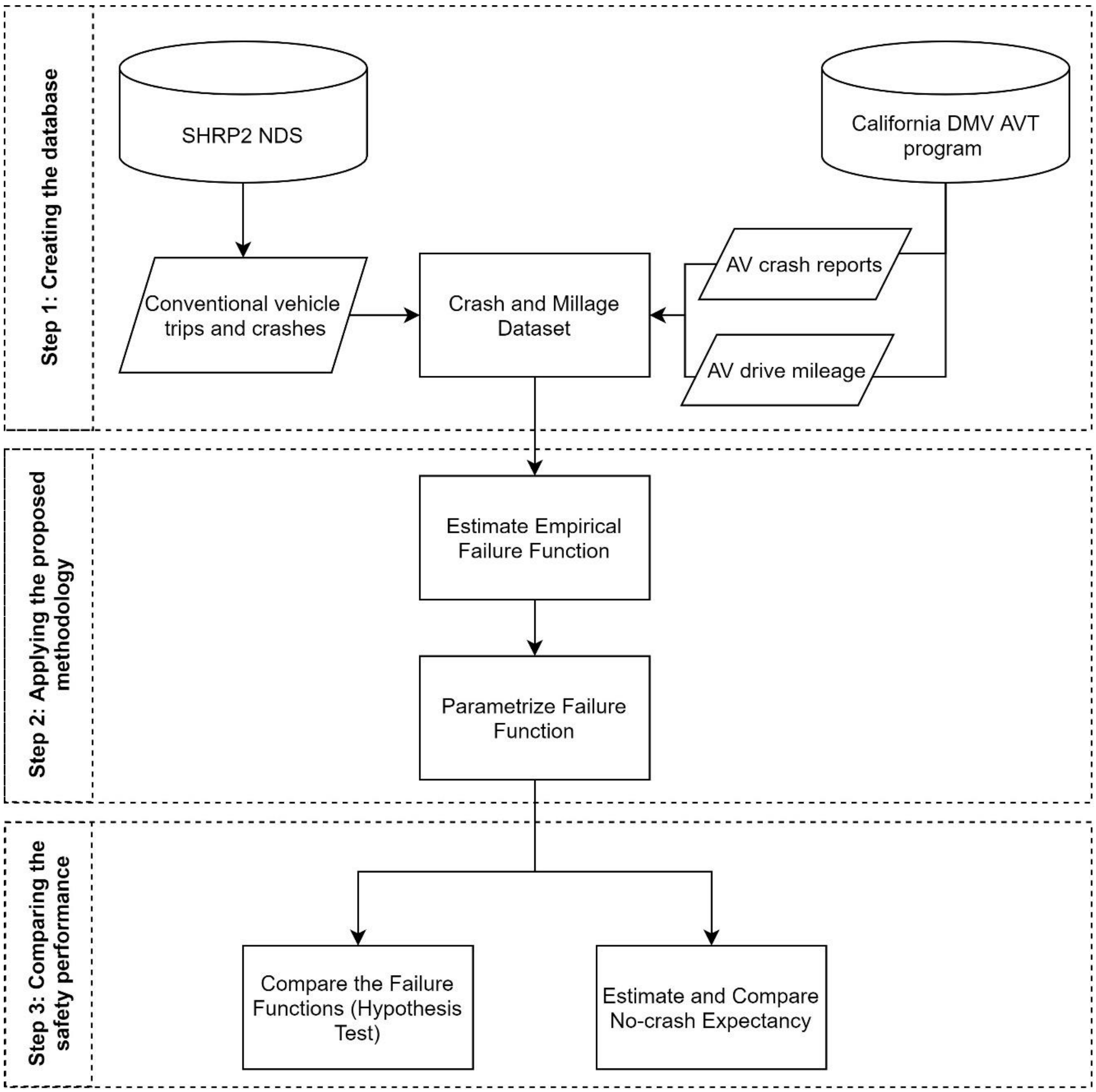


Figure 4.4. The designed empirical study architecture

4.5 Chapter Summary

An empirical study is designed to conduct a fair comparison between AV and conventional vehicle safety using the proposed methodology in this study. Although AVs are testing on United States' public roads in several states, the CA DMV is the only program that mandates AV manufacturers and developers to publicly share their road test data publicly, even though limited data lacks detail about AV test settings and environment. Despite AV road test data availability challenges, a crash database is created by sourcing AV crashes from the CA DMV AVT program and combining it with NDS crashes. Following the empirical study architecture, the created database is further used to examine the proposed AV safety evaluation methodology and evaluate AVs' safety under road tests compared to conventional vehicles. This chapter included the details of the designed empirical study and the datasets used in the empirical study. In the next chapter, the empirical study is analyzed, and the results are reported.

CHAPTER 5

RESULTS OF THE EMPIRICAL STUDY

This chapter documents the results of the 3-step empirical study, described in Chapter Four. The chapter sections follow the steps of the empirical study and report (1) characteristics of the created database, (2) the results of applying the proposed methodology on the empirical data, and (3) the comparison between conventional vehicles and AVs safety.

5.1 Dataset Characteristics

A descriptive analysis of AV and conventional vehicle crash datasets show a higher crash rate was observed for conventional vehicles comparing to AVs. In 2,849,50 miles of driving conventional vehicles, 130 crashes were observed, which is higher than 105 crashes AVs were involved in while driving the same millage. Consequently, the rate of AV crashes is 20% lower than conventional vehicles. The average of AV's MTC is higher than conventional vehicles, where on average, AVs were involved in crashes every 27,399 miles in comparison with 21,634 miles for conventual vehicles. Table 5.1 summarizes crash frequency and MTC statistics.

Table 5.1. Autonomous vehicles and conventional vehicles crash rates

Descriptive Statistics	Conventional Vehicles Crash Dataset	Autonomous Vehicles Crash Dataset
Number of crashes	130	105
Number of miles driven (million miles)	2,849,850	2,849,850
Rate of crashes (per million miles)	45.6	36.8
Mean MTC	21,634	27,399
Minimum MTC	12	4,212
Maximum MTC	112,975	134,023
Median MTC	12,679	15,767

5.2 Apply the Proposed AV Safety Evaluation Method

This section represents the estimated empirical and parametric failure functions for AV and conventional vehicles.

5.2.1 Empirical failure function estimation

The cumulative distribution of MTC in Figure 5.1 represents the empirical failure function, $F(x)$. The likelihood of AV and the conventional vehicle being involved in crashes can be compared using the empirical failure functions. For example, after 50,000 miles, the likelihood of involving in a crash was observed to be 86% and 84% for conventional vehicles and AV, respectively. In this regard, the survival probability, as a complement of failure probability, would be 14% and 16% for conventional vehicles and AVs. The probability of crashes was higher than 50% for AVs after driving 15,000 miles, in contrast with 13,000 for conventional vehicles. Although from Figure 5.1, AV crash likelihood is lower than conventional vehicles at (almost) every mile of driving, the significance of this difference needs to be investigated statistically.

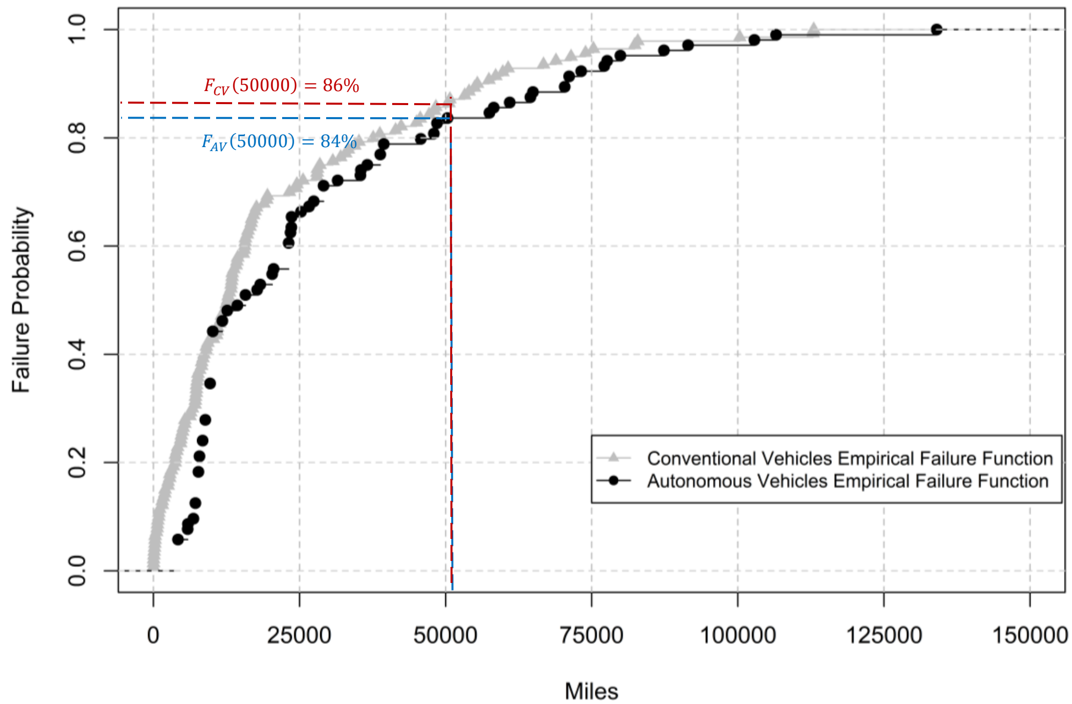


Figure 5.1. Estimated empirical failure function, $F(x)$, for autonomous vehicles and conventional vehicles

5.2.2 Parametric failure functions estimation

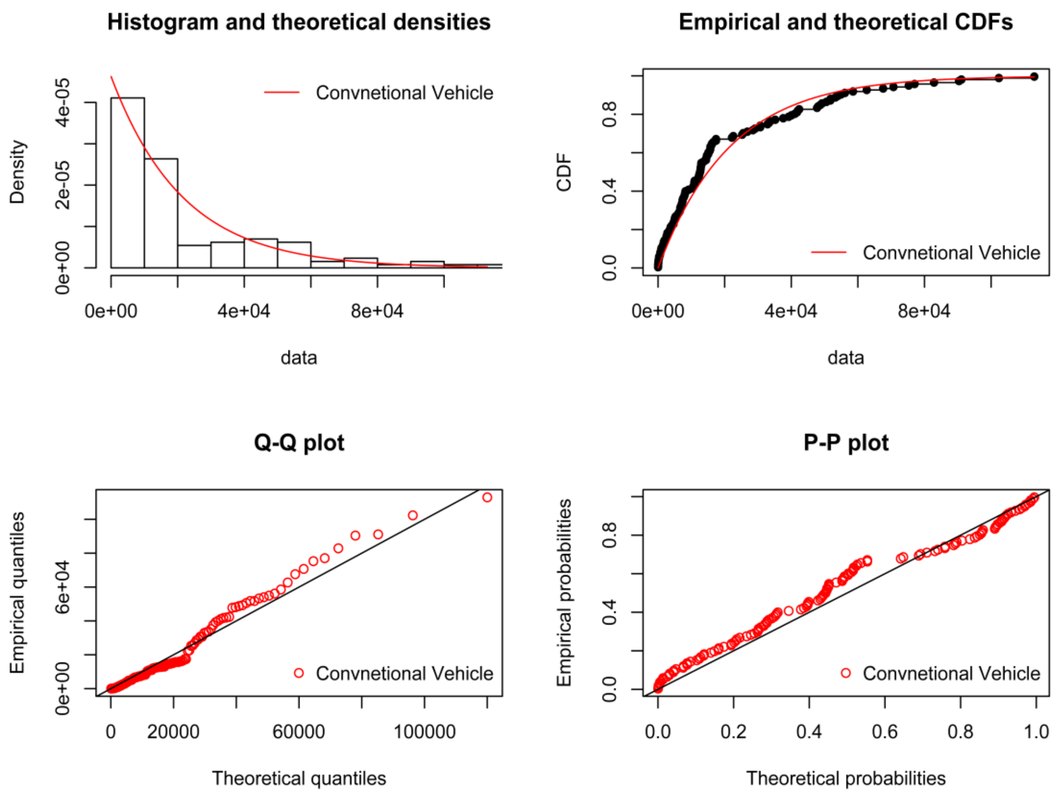
Using the MLE method, we fitted the distribution function to automated and conventional vehicles failure functions. As discussed in section 2.2, the exponential distribution function is fitted to the conventional vehicle failure function to meet the constant hazard rate assumption. For AVs, however, we test five parametric failure functions: exponential, gamma, log-normal, Weibull and log-logistic distribution functions. are consistent with AIC and BIC.

Table 5.2 depicts the estimated parameters and their standard error along with the AIC and BIC of the fitted distribution functions. The results of visual evaluations of distributions' fit considering Q-Q and P-P plots (Figure 5.2) are consistent with AIC and BIC.

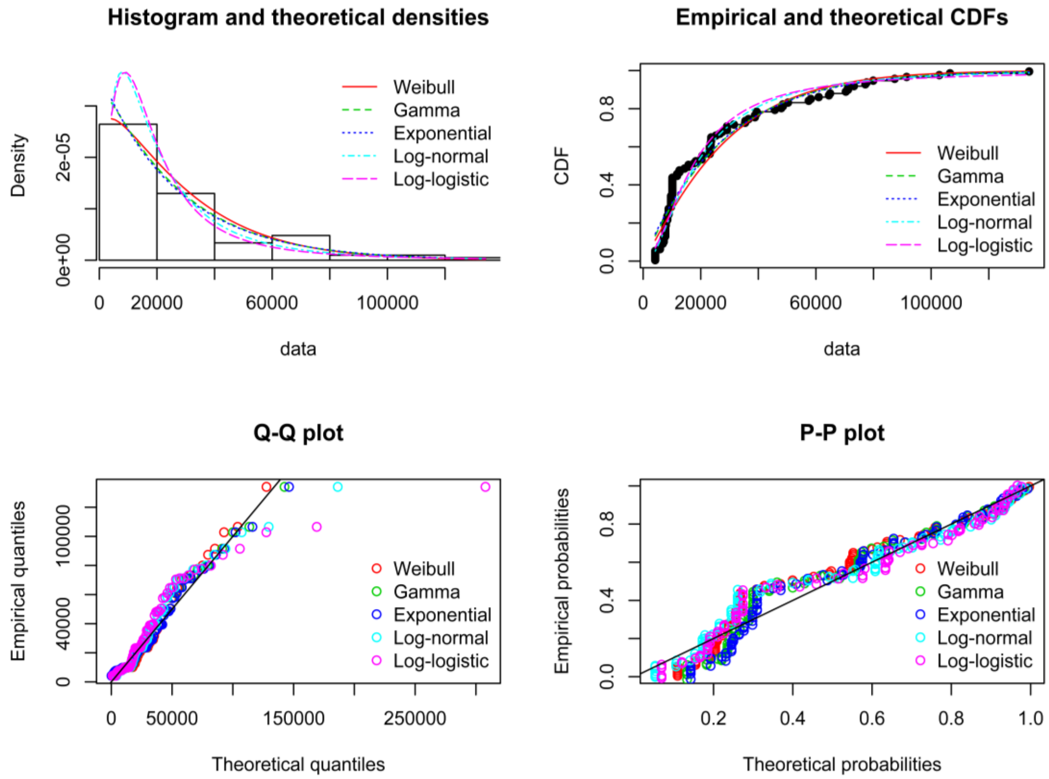
Table 5.2. Estimated parametric failure functions

Sample	**Parametric Distribution**	**Parameters**					**Goodness of fit**	
		Shape (SE)	Scale (SE)	Rate (SE)	Expected Value (SE)	Standard Deviation (SE)	AIC	BIC
Conventional Vehicles Failure	Exponential	NA	NA	4.62×10^{-5} (4.10×10^{-6})	NA	NA	2835.09	2837.95
Autonomous Vehicles Failure	Weibull	1.12 (8.82×10^{-2})	2.87×10^{4} (2.56×10^{3})	NA	NA	NA	2335.03	2340.32
	Gamma	1.05 (2.00×10^{-1})	3.82×10^{-5} (8.13×10^{-6})	NA	NA	NA	2335.93	2341.22
	Exponential	NA	NA	3.65×10^{-5} (3.71×10^{-6})	NA	NA	2335.40	2338.04
	Log-normal	NA	NA	NA	9.8 (8.76×10^{-2})	0.09 (6.22×10^{-2})	2316.36	2321.65
	Log-logistic	1.85 (1.56×10^{-1})	17212.03 (1.58×10^{3})	NA	NA	NA	2324.25	2329.54

NA: Not Applicable


Histogram and theoretical densities
Convnetional Vehicle
Density
4e-05
2e-05
0e+00
0e+00
4e+04
8e+04
data
Empirical and theoretical CDFs
CDF
0.8
0.4
0.0
Convnetional Vehicle
0e+00
4e+04
8e+04
data
Q-Q plot
Empirical quantiles
6e+04
0e+00
Convnetional Vehicle
0
20000
60000
100000
Theoretical quantiles
P-P plot
Empirical probabilities
0.8
0.4
0.0
Convnetional Vehicle
0.0
0.2
0.4
0.6
0.8
1.0
Theoretical probabilities

(a)



(b)

Figure 5.2. Visual comparison of Weibull, Gamma, Exponential, and Log-normal distribution functions goodness-of-fit to failure functions for (a) conventional vehicles and (b) autonomous vehicles

For more accurate evaluations, the K-S and A-D GOF tests are performed with a 95% confidence interval (The estimated one-way sample test statistics for the examined distributions show that the A-D and K-S tests reject the null hypothesis when the data sample follows a specific parametric distribution.

Table 5.3). The estimated one-way sample test statistics for the examined distributions show that the A-D and K-S tests reject the null hypothesis when the data sample follows a specific parametric distribution.

Table 5.3. Parametric distribution functions goodness-of-fit test

Parametric Distribution	**Conventional Vehicles Crashes**		**Autonomous Vehicles Crashes**	
	Test Statistic		Test Statistic	
	K-S (Critical value)	A-D (Critical value)	K-S (Critical value)	A-D (Critical value)
Weibull	NA	NA	0.17** (0.13‡)	3.14** (0.757)
Gamma	NA	NA	0.14** (0.13)	2.87** (0.752)
Exponential	0.12** (0.11)	2.78** (1.32)	0.14** (0.13)	2.97** (1.32)
Log-normal	NA	NA	0.18** (0.13)	2.18** (0.752)
Log-logistic	NA	NA	0.16** (0.11)	2.23** (0.32)

* The test Statistic is lower than the critical value → Cannot reject the null hypothesis; the sample follows the specified distribution

** The test Statistic is larger than the critical value → Reject the null hypothesis

† Calculated for sample size equal to 130

‡ Calculated for samples size equal to 105

5.2.3 Comparing Failure Functions

In the previous section, the parametric distribution functions were fitted to automated and conventional vehicle failure functions. Evaluating the parametric functions GOF and hypothesis tests, it was concluded that failure functions could not be parameterized. This section reports the results of testing the hypothesis that the AV empirical failure function is statistically different from the parametric conventional vehicle failure function with 95% confidence, using two-sample non-parametric K-S and A-D tests. As shown in Table 5.4, the K-S test statistics are higher than the critical value in a 95% confidence interval, which rejects the null hypothesis, the AV failure function $F_{AV}(x)$ is consistent with the AV estimated failure function $F_{CV}(x)$. Similarly, the comparison between the A-D test statistics and critical value in a 95% interval rejects the null hypothesis. Consequently, it can be concluded that the AV failure function is statistically inconsistent with the conventional vehicle failure function and, therefore, with 95% confidence, AV failure probability is lower than conventional vehicles' failure probability.

Table 5.4. Comparing autonomous vehicles and conventional vehicles failure functions

Goodness of Fit Tests	Critical Value for 95% Confidence Interval	Test Statistics
K-S test	0.009	0.261
A-D test	2.492	9.715

5.2.4 Safety effectiveness and no-crash expectancy estimation

The no-crash expectancy for each type of vehicle can be estimated at a given mileage. Table 5.5 compares the automated and conventional vehicle no-crash expectancies at different mileage. The results show that in the next 10,000 miles of driving, the no-crash expectancy for a conventional vehicle is ~4,000 miles while comparing to ~7,500 miles of no-crash expectancy for AVs, an 93% improvement is observed.

In the next 150,000 miles of driving, the no-crash expectancy would be increased by ~6,000 miles when shifting from conventional vehicles to AVs. In other words, in comparison with a conventional vehicle, on average, an AV can drive an additional 6,000 miles before observing a crash which represents a 27% statistically significant advantage in favor of vehicle automation. Consequently, the safety effectiveness of AVs is estimated as 1.27.

Table 5.5. AV and conventional vehicles no-crash expectancy and safety effectiveness

Mileage	Conventional Vehicle No-crash Expectancy (miles)	Automated Vehicle No-crash Expectancy (miles)	Difference in Miles	Difference in Percentage	AV Safety Effectiveness
10,000	3884.3	7499.6	3615.3	93%	-
25000	8026.8	11724.0	3697.2	46%	-
50,000	13730.1	16863.4	3133.3	23%	-
100,000	20256.89	24812.9	4556.01	22%	-
150,000	21611.98	27398.9	5786.92	27%	1.27

5.3 Chapter Summary

The proposed AV safety evaluation methodology is implemented on a designed empirical study, using AV and conventional vehicles' crashes. The primary analysis of AV and conventional vehicle crash datasets showed that AVs' crash rates are 20% lower than convention vehicles. On average, AVs were driven 27,399 miles before being involved in a crash which is higher than the 21,634 miles for conventional vehicles. Also, a comparison of AV and conventional vehicle's empirical distribution of MTC represent the lower failure (crash) probability for AVs. Parametric distributions were fitted to the failure function of AVs and conventional vehicles, and it was shown that no parametric distribution could characterize AV and conventional vehicle failure function. Using two samples K-S and A-D tests, the hypothesis of whether the AV failure function is different from the conventional vehicles' failure function was examined. It can be concluded that the difference between AVs and conventional vehicles' failure function is statistically significant based on the 95% confidence interval level. The findings of the analysis presented in the paper imply that Level 3 of automation testing in California is safer than conventional vehicles with 95% confidence. Also, comparing the no-crash expectancy between the conventional vehicle and AV shows a 27% improvement in a 150,000 miles of operation. The next chapter summarizes this research and highlights the limitations of this work as well as avenues for future research.

CHAPTER 6

SUMMARY AND DISCUSSION

This chapter comprises three sections. First, previous chapters are summarizing, and the conclusions of this research are highlighted. Second, in the discussion section, the researcher discusses the strengths and limitations of this research and underlines its policy implications. Finally, the potential avenues for future research are pointed out.

6.1 Summary and Conclusions

This research targeted the safety evaluation of AVs and sought the answer to three fundamental questions:

1) How can AV safety be validated, and what are the research gaps in the existing safety evaluation methods?
2) What methodologies are required to validate AV safety, evaluate their safety performance and investigate the contributing factors to AV crashes?
3) How safe are the existing AVs in comparison with conventional vehicles?

First, a scoping review methodology was conducted to (systematically) synthesize the AV safety quantification methods. The identified evaluation methods are compared, and the gaps and challenges in AV safety evaluation are uncovered. As a result of the scoping review, the AV safety evaluation methods were categorized into six groups: target crash population, road test analysis, traffic simulation, driving simulator, safety effectiveness estimation, and system failure assessment. We ran two evaluations on the identified approaches. First, we investigated each approach in terms of its input, output, and application to estimate AVs' safety implications at the vehicle, transportation system, and society levels. Second, we qualitatively compared them in terms of three criteria: availability of input data, suitability for evaluating different automation levels, and reliability of estimations. The comparison presented in this review can be used as a guideline for future research when choosing the appropriate AV safety evaluation method based on the study objective and limitations. This review identifies four challenges in AV safety evaluation: (a) shortcomings in methodologies for evaluating and quantifying AV safety, (b) uncertainties in AV implementations and their impacts on AV safety, (c) potential riskier behavior of AV passengers as well as other road users, and (d) emerging safety issues related to AV implementations. These challenges need to be addressed for a clearer perception of AV safety.

In Chapter Two, it was discussed that road test data analysis is the most reliable method for evaluating AV safety since it can address the AV safety challenges. However, the reliability of results is contingent upon the availability of data. Also, the road test data is limited to transportation and system-level safety evaluation, while vehicle-level safety evaluations would be necessary for improving AV technologies. To address the limitations of the road test analysis method for AV safety evaluation, we proposed a new methodology based on survival analysis for evaluating AV safety. As a result, AV and the conventional vehicle's failure function can be estimated, representing the likelihood of involving in a crash at a certain mileage of driving. The inconsistency of failure function can be further analyzed using a statistical method that can support the assessment of analogy between AV and conventional vehicle safety. In addition, on the basis of the vehicles' failure function, a new metric no-crash expectancy is defined that provides a better perception regarding the vehicles' safety performance. Further investigation of theories behind traditional crash prediction models shows that they are not effective for AVs.

We further designed an empirical study to collect and create reliable and comparable AV (Level 3 of automation) and conventional vehicle crash datasets (collected from a naturalistic study for conventional vehicles). As a result of implementing the proposed methodology on AV and conventional vehicle crashes, we showed that AVs are safer than conventional vehicles with 95% confidence. Also, AVs showed a higher no-crash expectancy and 1.27 safety effectiveness comparing to conventional vehicles. This study indicates a safer performance of Level 3 of automation than conventional vehicles applying the proposed method on comparable crash datasets. However, the results of our analysis are subject to the accuracy of AV crash data, the assumptions regarding AV road tests environment, and AV crash mileage. Future research is required to address the limitations of this study and explore our simplifying assumption. Moreover, the proposed methodology can be used to evaluate the safety of AV technologies and different levels of automation and MPR.

6.2 Discussion

The strengths and limitations of the conducted literature review, the proposed methodology, and the empirical study, as well as the potential policy implications of this study, are presented in this section.

6.2.1 Strengths and Limitations

The results of synthesizing AV safety evaluation literature are expected to serve as a stop knowledge point and future research avenues to contribute to AV safety evaluation. However, the conducted review has some limitations. First, this study focuses on AV safety quantification methods; therefore, I did not include the literature that evaluated ADAS safety implications or proposed frameworks and conceptual models for AV safety evaluation rather than quantifying the impacts. Both ADAS safety evaluation methods and proposed frameworks for AV safety evaluation might have the potential to address some of the limitations of the existing quantification methods. Second, the AV safety evaluation methodologies were examined qualitatively and relatively. Future research can provide a more accurate comparison between the methods by running quantitative analyses. Thirds, the literature review only includes peer-reviewed publications and white paper and AV manufacturer reports are not included in our review. Fourth, the focus of this review was on methodologies that quantified AVs' substantive safety rather than the nominal safety and perceived safety. Nominal safety refers to whether or not a vehicle is fulfilling all standards and laws that apply to the vehicle and the nominal safety of AVs needs to be investigated in accordance with standards (Kalra and Paddock, 2016). The perceived safety of a vehicle is how the general public experiences the safety of the vehicle, which the perceived safety of AV was targeted by conducting survey studies (Moody et al., 2020). The safety of vehicles should be evaluated based on three definitions of safety. Even though AV safety can be comparable to that of conventional vehicles, users' degrading perceptions of AV safety may hinder the adoption of this new technology. Future research is required to review the literature and examine the methodologies used for evaluating AV nominal safety and perceived safety for more accurate evaluations and understanding of AV safety. Moreover, this study was not intended to synthesize the results of AV safety quantifications but rather to explore the methodologies.

This study proposed a novel methodological framework for evaluating AV safety in comparison with conventional vehicles. The proposed framework is transferable and can be used to evaluate levels of automation and ADAS. The designed empirical study employs NDS crash data that included both police reportable and non-reportable crashes and, therefore, can be considered to contain all crashes, similar to the AV crash dataset. To the best of the authors' knowledge, to date, no study has conducted such a fair comparison between automated and conventional vehicles safety. The accuracy of road test data analysis is heavily subject to the accuracy of reported crashes from AV manufactures, automation level and operation design domain (ODD) under which AVs were tested, and fallback users' characteristics. Another disadvantage of automated driving road tests is that they may be exposing road users to the risk of crashes from under-developed AVs, as discussed by Kalra (2017). Such a safety issue may limit road tests and, consequently, the applicability of the proposed framework. Analyzing road test data can provide insights into how AV interacts with other road users, including other AVs; however, investigating AV safety in higher levels of MPR can be challenging, given that a limited number of AVs are operating on the roads. Although the proposed method can be used for evaluating the vehicle-level crash contributing factors, the potential heterogeneity in the road test data should be addressed. Drivers' behavior and characteristics, vehicles' characteristics (size, type, mechanical features, etc.), and road test environment are some of the factors that can cause heterogeneity in the road test data. Future models can control for observable factors and account for unobservable factors using more advanced variations of the hazard proportional cox-regression (e.g., random-parameter models (Balusu et al., 2020)).

Our empirical analysis has certain limitations as well. First, some limitations are inherited in road test data analysis. As CA DMV mandates AV manufacturers, it was assumed that they reported all crashes, and the crash dataset consists of all AV crashes tested in California in 2019. According to CA DMV, the AVT program is limited to testing Level 3 of automation, and so the analysis presented in this study can only evaluate the safety of Level 3 of automation. Although there is no information regarding the AV testing ODD, since Level 3 of automation is designed to operate in unlimited ODD (SAE, 2018), it was assumed that AVs were tested on roads with different functional classifications and are comparable with conventional vehicles. Level 3 of automation requires fallback users to intervene in certain situations. The disengagement from the ADS imposes a considerable risk of crashes (Happee et al., 2017). Depending on the experience and awareness of fallback users, this risk can be lower or higher. Since the AV crash dataset is collected as part of the AVT program, it is expected that the fallback users are both experienced and constantly pay attention to AVs performing the DDT. In conclusion, the empirical study may overestimate AV safety since the disengagement risk could not be measured accurately. Second, given that AVs' crash report does not include the time of the crash and the vehicle mileage, the estimated miles-to-crash is rounded up to the miles driven in a day on which the crash occurred. Having access to the exact millage of vehicles would resolve this issue and result in a smoother failure function for AV. Nevertheless, we do not anticipate it would affect the conclusions of this study. Third, the empirical study was conducted using one year of AVs operation, which can be translated into 2.8 million miles.

While 2.8 million miles are expected to be sufficient to capture miles-to-crash and failure probability distribution, future studies can expand their timeframe and examine our expectations in this regard. Fourth, to be able to analyze AV safety using the limited available data, we assumed a homogeneous crash dataset. This assumption needs to be revisited. Fifth, the designed empirical study targeted the substantive safety of AVs in terms of crash frequency. Therefore, our results do not provide insights into the nominal safety or perceived safety of AVs. Sixth, this study explored AV safety in terms of crash frequencies. Nevertheless, AV's contributions to the severity of crashes need to be investigated as well. While the severity of AV crashes was studied in the literature (Xu et al., 2019, Wang and Li, 2019), the proposed framework can be used to compare AV failure function by crash severity once enough road test data is available. Finally, it is expected that NDS data may be impacted by the self-selection bias (participants who are selected to be monitored may change their behavior), which could change the comparison of the risk between AVs and human-driven vehicles. The magnitude of the effects is currently not known but is probably not very large.

6.2.2 Policy and research implications

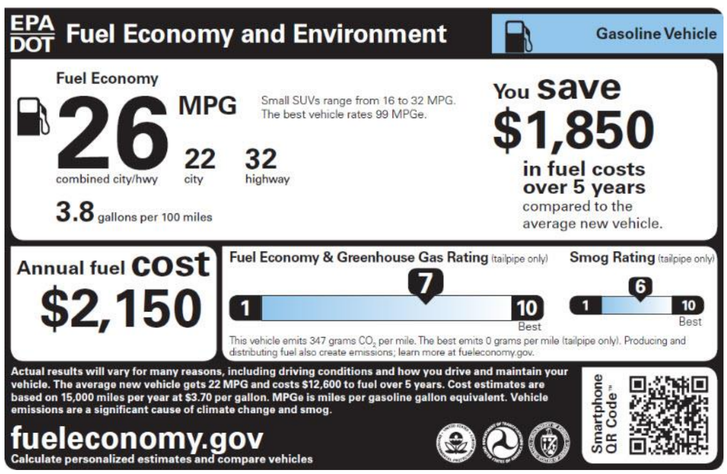
The availability and accuracy of road test data are a fundamental need for evaluating AV safety under different environments and with different market penetration rates, regardless of the evaluation methodology. As discussed previously, conclusions from the proposed methodology in this study were contingent upon the accuracy of AV crash data and could be improved if more detailed data were available. As such, federal, state, and local laws are required to not only support automated driving road tests but also promote the transparency of road test programs. To date, in the United States, no federal law mandates AV manufacturers to publicize their road test results. Although many states allowed AVs to be tested on their roads, a few mandated manufactures to report their data (as discussed above, this may change soon).

Some pitfalls and limitations are inherent in automated driving road tests. Exposure of the road users to the safety risks associated with vehicles under development is one of the pitfalls that look inevitable. Authorizing rules to monitor the road test and increasing the liability of manufactures in time of a crash can be some remedies to lessen the road test disadvantages. In addition, to assure the reliability of the results, road tests must encompass the AVs' ODD while examining real-world safety challenges that AVs might encounter. This incentivized researchers to generate road testing scenarios (Feng et al., 2020b, Feng et al., 2020a).

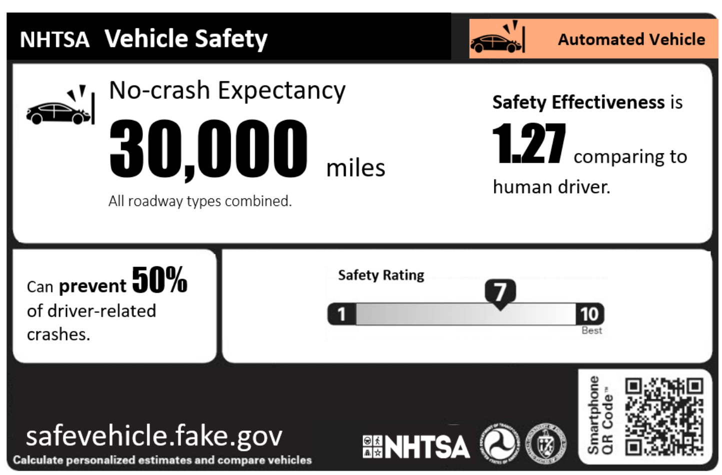
Despite the previous efforts, Milakis et al. (2017) pointed out that AV safety requires further investigations for policy-making purposes. Besides, Pettigrew et al. (2018) showed that the public is not aware of AVs' safety advantages and is quite skeptical regarding AV safety. Standards are required to define to what degree AVs should be safer than conventional vehicles to be able to find their way on the roads. The timing of AV introduction is crucial since postponing it would hinder access to AVs' benefits.

This study proposes rethinking safety evaluations in the era of vehicle automation. While we proposed a new method for evaluating AV safety and touched on the theories behind crash prediction models and how AVs can impact them, future research is required to further investigate the theories proposed in this paper once more data is available. We consider this study as a research agenda for future research on AV safety.

The new AV safety evaluation metric, no-crash expectancy, can be used for comparing the safety performance of different automated cars. Environmental Protection Agency’ (EPA) provides the fuel economy data that is used on the fuel economy label on all new cars and light trucks that can be used for comparing the vehicle fuel economy. Similarly, NHTSA can use the proposed methodology and vehicle safety metrics test vehicles and report the information regarding vehicle safety. Figure 12 illustrates a sample label for AV safety.



(a)



* Only for demonstration

(b)

Figure 3. An illustration of (a) EPA’s fuel economy label and (b) suggested vehicle safety label

6.3 Future Studies

The synthesized literature in this study has various implications for the future direction of AV safety research. First, identified approaches have some shortcomings and limitations that need to be addressed. Mix-traffic issues, system failure, and fallback user errors have not been considered in the target crash population approach. Accounting for these factors can potentially lead to more accurate estimations of AV safety implications. Driving simulators and traffic simulation studies can benefit from ground truth data (e.g., AV road test data) to verify their assumptions and study findings. AV system failure assessments should be revisited using more reliable data on AV system failure rates. Running statistical analyses on a large amount of AV road test data in future studies can provide more reliable conclusions regarding AV safety. Second, AV safety studies do not generally account for uncertainties in AV implementations—i.e., AV MPR and its role in urban areas, trip patterns, and transportation systems. Future research can address this limitation in order to assess the safety impact of AVs at the society level. Third, since a riskier behavior of AV passengers as well as other road users is expected after AV implementation, further investigations of the risk homeostasis hypothesis are needed to measure and govern the potential safety impacts. Fourth, the emerging safety issues related to AVs, including cybersecurity and AVs' reactions during unavoidable crashes, should be studied further. In addition, future studies should address the limitations of this review, namely (a) defining a broader review question, (b) evaluating the identified methodology quantitatively, and (c) investigating AVs' nominal and perceived safety implications. Table 6.1 shows the potential list of future research directions, the study topics, and the level of safety impact these studies can address.

Table 6.1. Suggested future studies

Research topic	Study subtopics	Level of safety impact
Address the limitations of existing AV safety quantification methods	Consider mix-traffic issues, system failure, and the risk associated with fallback-ready user reaction at the time of AV disengagement from ADS in the target crash population approach	Transportation system + Society
	Evaluate traffic simulation and driving simulator results using AV road test data	Vehicle + Transportation system
	Collect and analyze reliable system failure rates	Vehicle
	Perform reliable statistical analysis on a larger AV dataset	Society
Perform full-chain assessment of AVs' safety implications	Account for AV MPR and its influence on urban areas, trip patterns, and transportation systems	Society
Investigate the potential risky behavior of AV users	Examine the risk homeostasis hypothesis	Transportation system + Society
Study the emerging safety issues associated with AV implementations	Address AVs' cybersecurity issues	Vehicle
	Preprogram AVs to follow the best course of action during unavoidable crashes	Vehicle

Future studies are required to address some of the limitations of the conducted empirical study. As extensively discussed in Section 6.2.1, future research is required to (1) evaluate AV safety using more reliable road test data (AV crash data), including details about ODD and fallback-user characteristics, (2) incorporate a higher resolution road operation data for estimating MTC, (3) using a larger database for analysis, (4) including the crash severity into the analysis. Analyzing more reliable AV crash reports (once available) and considering crash severity into analyses are suggested. The researcher suggests examining the safety of other levels of ADS and ADAS using the proposed methodology. This requires sufficient road test data along with information regarding the number of miles driven by the target vehicle before an incident.

From the methodological standpoint, further research is required to (1) explore the applicability of the proposed method and cox proportional regression for investigating the vehicle-level characteristics of AVs in their safety performance, (2) incorporate the unobserved heterogeneity in AV safety performance evaluation, and (3) research the functional form of hazard rate and its impacts on crash prediction models in the era of vehicle automation.

6.3.1 Vehicle-level safety analysis

Although the existing econometrics methodologies for crash predictions can examine the impacts of road characteristics and environmental conditions on AV crashes frequency at the road segment level (Lord et al., 2021), the impacts of vehicle-level factors remain unclear. For instance, the safety impacts of AV technology improvements can be of interest to manufacturers that cannot be investigated using traditional econometrics methodologies in the context of road test analyses. The contribution of AV driver's characteristics (in levels 1, 2, and 3) to crash frequency is another example in which the existing road segment-level methodologies cannot be used for investigating vehicle-level contributing factors. The proposed cox proportional regression model in this study can be used to study the vehicle-level contributing factors―namely, specific safety technologies, vehicle design, fallback-user characteristics, vehicle ODD, etc. This also suggests that the implications of the proposed vehicle-level safety evaluation methodology go beyond AV safety evaluations can be used to explore the safety impacts of other vehicle-level characteristics such as vehicle type, vehicle age, vehicle maintenance status, braking system status and technology, etc.

6.3.2 Heterogeneity in automated vehicle crashes

The crash data is not homogeneous, and the heterogeneity needs to be addressed for efficient estimates and accurate predictions. While accounting for contributing factors can address the heterogeneity to some extent, the unobserved heterogeneity requires further attention by the researcher when sufficient automated driving road test data is available. Random parameter models have been widely used in the literature to address the unobserved heterogeneity (Lord and Mannering, 2010, Washington et al., 2020).

6.3.3 Automated vehicle hazard rate

While the survival functions were assumed as exponential in this study to fulfill the constant hazard rate for vehicles, future studies need to investigate the feasibility of constant hazard rate for vehicle-level safety evaluations. In addition, we discussed the impacts of considering hazard rate as a function of the number of miles driven on the AV crash occurrence process and elaborated on the theories behind crash prediction models in the era of vehicle automation. Further research is required to formulate a non-homogeneous Poisson process for AV crashes.

REFERENCES

Agriesti, S., Studer, L., Gandini, P., Marchionni, G., Ponti, M. & Visintainer, F. 2019. Safety on the Italian Highways: Impacts of the Highway Chauffeur System. *Smart Transportation Systems 2019* (pp. 57-73). Springer, Singapore.

Anderson, T. W. 2011. Anderson-Darling Tests of Goodness-of-Fit. *International Encyclopedia of Statistical Science*, 1**,** 52-54.

Anderson, T. W. & Darling, D. A. 1954. A test of goodness of fit. *Journal of the American Statistical Association*, 49**,** 765-769.

Arksey, H. & O'Malley, L. 2005. Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8**,** 19-32.

Arvin, R., Kamrani, M., Khattak, A. & Rios-Torres, J. Safety Impacts of Automated Vehicles in Mixed Traffic. Transportation Research Board Annual Meeting, 2018, Washington D.C., District of Columbia.

Arvin, R., Khattak, A. & Rios Torres, J. Evaluating Safety with Automated Vehicles at Signalized Intersections: Application of Adaptive Cruise Control in Mixed Traffic. Transportation Research Board Annual Meeting, 2019, Washington D.C., District of Columbia.

Arvin, R., Khattak, A. J., Kamrani, M. & Rio-Torres, J. 2020. Safety evaluation of connected and automated vehicles in mixed traffic with conventional vehicles at intersections. *Journal of Intelligent Transportation Systems***,** 1-18.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F. & Rahwan, I. 2018. The moral machine experiment. *Nature*, 563**,** 59.

Bagloee, S. A., Asadi, M. & Practice 2016a. Crash analysis at intersections in the CBD: A survival analysis model. *Transportation Research Part A: Policy and Practice*, 94**,** 558-572.

Bagloee, S. A., Tavana, M., Asadi, M. & Oliver, T. 2016b. Autonomous vehicles: challenges, opportunities, and future implications for transportation policies. *Journal of Modern Transportation*, 24**,** 284-303.

Bahram, M., Ghandeharioun, Z., Zahn, P., Baur, M., Huber, W. & Busch, F. Microscopic traffic simulation based evaluation of highly automated driving on highways. 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2014, Qingdao, China. IEEE, 1752-1757.

Balusu, S. K., Mannering, F. & Pinjari, A. 2020. Hazard-based duration analysis of the time between motorcyclists' initial training and their first crash. *Analytic Methods in Accident Research,* 28**,** 100143.

Banerjee, S. S., Jha, S., Cyriac, J., Kalbarczyk, Z. T. & Iyer, R. K. Hands off the wheel in autonomous vehicles?: A systems perspective on over a million miles of field data. 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 2018, Luxembourg, IEEE, 586-597.

Berthelon, C. & Gineyt, G. 2014. Effects of alcohol on automated and controlled driving performances. *Psychopharmacology,* 231**,** 2087-2095.

Bhavsar, P., Das, P., Paugh, M., Dey, K. & Chowdhury, M. 2017. Risk analysis of autonomous vehicles in mixed traffic streams. *Transportation Research Record,* 2625**,** 51-61.

Bila, C., Sivrikaya, F., Khan, M. A. & Albayrak, S. 2017. Vehicles of the future: A survey of research on safety issues. *IEEE Transactions on Intelligent Transportation Systems,* 18**,** 1046-1065.

Blommer, M., Curry, R., Kochhar, D., Swaminathan, R., Talamonti, W. & Tijerina, L. The effects of a scheduled driver engagement strategy in automated driving. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2015. SAGE Publications Sage CA: Los Angeles, CA, 1681-1685.

Boggs, A. M., Wali, B. & Khattak, A. J. 2020. Exploratory analysis of automated vehicle crashes in California: a text analytics & hierarchical Bayesian heterogeneity-based approach. *Accident Analysis and Prevention,* 135**,** 105354.

Canis, B. 2020. Issues in Autonomous Vehicle Testing and Deployment. Congressional Research Service. Available: https://fas.org/sgp/crs/misc/R45985.pdf (April 2020).

Caragata Nasvadi, G. & Wister, A. 2009. Do restricted driver's licenses lower crash risk among older drivers? A survival analysis of insurance data from British Columbia. *The Gerontologist,* 49**,** 474-484.

Childress, S., Nichols, B., Charlton, B. & Coe, S. 2015. Using an activity-based model to explore the potential impacts of automated vehicles. *Journal of the Transportation Research Board,* 2493**,** 99-106.

Cohen, S. & Shirazi, S. 2017. Can We Advance Social Equity with Shared, Autonomous and Electric Vehicles? Institute of Transportation Studies at the University of California, Davis. Available:

https://www.transformca.org/sites/default/files/3R.Equity.Indesign.Final_.pdf> (January 2020).

Combs, T. S., Sandt, L. S., Clamann, M. P. & McDonald, N. C. 2019. Automated vehicles and pedestrian safety: exploring the promise and limits of pedestrian detection. *American journal of preventive medicine,* 56**,** 1-7.

Cui, J., Liew, L. S., Sabaliauskaite, G. & Zhou, F. 2019. A review on safety failures, security attacks, and available countermeasures for autonomous vehicles. *Ad Hoc Networks,* 90**,** 101823.

Das, S., Dutta, A. & Tsapakis, I. 2020. Automated vehicle collisions in California: Applying Bayesian latent class model. *IATSS Research*, 44(4), pp.300-308.

Deluka Tibljaš, A., Giuffrè, T., Surdonja, S. & Trubia, S. 2018. Introduction of Autonomous Vehicles: Roundabouts Design and Safety Performance Evaluation. *Sustainability,* 10**,** 1060-1060.

Desmond, P. A., Hancock, P. A. & Monette, J. L. 1998. Fatigue and automation-induced impairments in simulated driving performance. *Transportation Research Record,* 1628**,** 8-14.

Detwiller, M. & Gabler, H. C. Potential reduction in pedestrian collisions with an autonomous vehicle. The 25th ESV Conference Proceedings. NHTSA, 2017, Detroit, Michigan. 1-8.

Fagnant, D. J. & Kockelman, K. 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice,* 77**,** 167-181.

Favarò, F. M., Nader, N., Eurich, S. O., Tripp, M. & Varadaraju, N. 2017. Examining accident reports involving autonomous vehicles in California. *PLoS one,* 12**,** e0184952.

Feng, S., Feng, Y., Sun, H., Bao, S., Zhang, Y. & Liu, H. X. 2020a. Testing scenario library generation for connected and automated vehicles, part II: Case studies. *IEEE Transactions on Intelligent Transportation Systems*, 22(9), 5635-5647.

Feng, S., Feng, Y., Yu, C., Zhang, Y. & Liu, H. X. 2020b. Testing scenario library generation for connected and automated vehicles, part i: Methodology. *IEEE Transactions on Intelligent Transportation Systems*, 22(3), 1573-1582.

Ferrante, A. M., Rosman, D. L. & Marom, Y. 2001. Novice drink drivers, recidivism and crash involvement. *Accident Analysis and Prevention,* 33**,** 221-227.

Fu, H. 2008. Identifying repeat DUI crash factors using state crash records. *Accident Analysis and Prevention,* 40**,** 2037-2042.

Gold, C., Happee, R. & Bengler, K. 2018. Modeling take-over performance in level 3 conditionally automated vehicles. *Accident Analysis & Prevention,* 116**,** 3-13.

Goodall, N. J. 2014. Ethical decision making during automated vehicle crashes. *Transportation Research Record,* 2424**,** 58-65.

Gouy, M., Diels, C., Reed[1], N., Stevens, A. & Burnett, G. 2012. The effects of short time headways within automated vehicle platoons on other drivers. *Human Aspects of Road andRail Transportation*, p.529.

Happee, R., Gold, C., Radlmayr, J., Hergeth, S. & Bengler, K. 2017. Take-over performance in evasive manoeuvres. *Accident Analysis & Prevention,* 106**,** 211-222.

Harhay, M. O., Porcher, R., Cantu, E., Crowther, M. J., Christie, J. D., Thabut, G. & Donaldson, G. C. 2018. An Alternative approach for the analysis of time-to-event and survival outcomes in pulmonary medicine. *American Journal of Respiratory and Critical Care Medicine,* 198**,** 684-687.

Hendrickson, C. T. & Harper, C. 2018. Safety and Cost Assessment of Connected and Automated Vehicles. Carnegie-Mellon University. Technologies for Safe and Efficient Transportation University Transportation Center. Number: 01603608. Available: < https://rosap.ntl.bts.gov/view/dot/35124> (May 2021).

Jafarnejad, S., Codeca, L., Bronzi, W., Frank, R. & Engel, T. A car hacking experiment: When connectivity meets vulnerability. 2015 IEEE Globecom Workshops, 2015, San Diego, California. IEEE, 1-6.

Jäntschi, L. & Bolboacă, S. D. 2018. Computation of probability associated with Anderson–Darling statistic. *Mathematics,* 6**,** 88.

Junietz, P., Wachenfeld, W., Klonecki, K. & Winner, H. Evaluation of Different Approaches to Address Safety Validation of Automated Driving. 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018. IEEE, 491-496.

Kalra, N. 2017. *Challenges and approaches to realizing autonomous vehicle safety,* Santa Monica, CA, RAND Corporation. Available: <https://www.rand.org/pubs/testimonies/CT463.html.> (November 2021)

Kalra, N. & Paddock, S. M. 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice,* 94**,** 182-193.

Katrakazas, C., Quddus, M. & Chen, W. H. 2019. A new integrated collision risk assessment methodology for autonomous vehicles. *Accident Analysis and Prevention,* 127**,** 61-79.

Kockelman, K., Avery, P., Bansal, P., Boyles, S. D., Bujanovic, P., Choudhary, T., Clements, L., Domnenko, G., Fagnant, D. & Helsel, J. 2016. Implications of connected and automated vehicles on the safety and operations of roadway networks: A final report. Center for Transportation Research, University of Texas at Austin. FHWA/TX-16/0-6849-1, Available: <https://library.ctr.utexas.edu/ctr-publications/0-6849-1.pdf> (January 2019).

Koopman, P. & Wagner, M. 2016. Challenges in autonomous vehicle testing and validation. *SAE International Journal of Transportation Safety,* 4**,** 15-24.

Kundinger, T., Riener, A., Sofra, N. & Weigl, K. Drowsiness Detection and Warning in Manual and Automated Driving: Results from Subjective Evaluation. Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 2018, Toronto, Canada. 229-236.

Kusano, K. D. & Gabler, H. C. 2014. Comprehensive target populations for current active safety systems using national crash databases. *Traffic injury prevention,* 15**,** 753-761.

Lee, C. 2017. Grabbing the wheel early: Moving forward on cybersecurity and privacy protections for driverless cars. *Federal Communations Law Journal,* 69**,** 25.

Lee, S., Oh, C. & Hong, S. 2018. Exploring lane change safety issues for manually driven vehicles in vehicle platooning environments. *IET Intelligent Transport Systems,* 12**,** 1142-1147.

Lee, S. C., Yoon, S. H. & Ji, Y. G. 2020. Effects of Non-Driving-Related Task Attributes on Takeover Quality in Automated Vehicles. *International Journal of Human–Computer Interaction***,** 1-9.

Li, R. & Zhai, R. Estimation and Analysis of Minimum Traveling Distance in Self-driving Vehicle to Prove Their Safety on Road Test. Journal of Physics: Conference Series, 2019, Ningbo, China. IOP Publishing, 032101.

Li, T. & Kockelman, K. M. Valuing the safety benefits of connected and automated vehicle technologies. Transportation Research Board 95th Annual Meeting, 2016, Washington D.C., District of Columbia.

Litman, T. 2017. Autonomous vehicle implementation predictions. Victoria Transport Policy Institute, Victoria, Canada. Available: <https://www.vtpi.org/avip.pdf> (January 2019).

Liu, X. 2012. *Survival analysis: models and applications*, John Wiley & Sons.

Lord, D., Qin, X. & Geedipally, S. R. 2021. *Highway Safety Analytics and Modeling*, Elsevier Publish Ltd, Amsterdam, Netherlands.

Lubbe, L., Jeppsson, H., Ranjbar, A., Fredriksson, J., Bärgman, J. & Östling, M. Predicted road traffic fatalities in Germany: The potential and limitations of vehicle safety technologies from passive safety to highly automated driving. Proceedings of IRCOBI conference. 2018.

Mannering, F. 1993. Male/female driver characteristics and accident risk: some new evidence. *Accident Analysis and Prevention,* 25**,** 77-84.

Martínez-Díaz, M. & Soriguera, F. 2018. Autonomous vehicles: Theoretical and practical challenges. *Transportation Research Procedia,* 33**,** 275-282.

Matysiak, A. & Razin, P. Analysis of the advancements in real-life performance of highly automated vehicles' with regard to the road traffic safety. MATEC Web of Conferences, 2018. EDP Sciences, 05003.

Milakis, D., Van Arem, B. & Van Wee, B. 2017. Policy and society related implications of automated driving: A review of literature and directions for future research. *Journal of Intelligent Transportation Systems,* 21**,** 324-348.

Montanaro, U., Dixit, S., Fallah, S., Dianati, M., Stevens, A., Oxtoby, D. & Mouzakitis, A. 2018. Towards connected autonomous driving: Review of use-cases. *Vehicle System Dynamics,* 57**,** 1-36.

Moody, J., Bailey, N. & Zhao, J. 2020. Public perceptions of autonomous vehicle safety: An international comparison. *Safety Science,* 121**,** 634-650.

Morando, M. M., Tian, Q., Truong, L. T. & Vu, H. L. 2018. Studying the Safety Impact of Autonomous Vehicles Using Simulation-Based Surrogate Safety Measures. *Journal of Advanced Transportation,* 2018, 6135183.

Mousavi, S. M., Lord, D., Dadashova, B. & Reza Mousavi, S. Can Autonomous Vehicles Enhance Traffic Safety at Unsignalized Intersections? International Conference on Transportation and Development, 2020, Reston, Virginia. American Society of Civil Engineers, 194-206.

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A. & Aromataris, E. 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology,* 18**,** 143.

NHTSA 2018. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. U.S. Department of Transportation. Available: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812506> (March 2020).

Papadoulis, A., Quddus, M. & Imprialou, M. 2019. Evaluating the safety impact of connected and autonomous vehicles on motorways. *Accident Analysis and Prevention,* 124**,** 12-22.

Petrović, Đ., Mijailović, R. & Pešić, D. 2020. Traffic Accidents with Autonomous Vehicles: Type of Collisions, Manoeuvres and Errors of Conventional Vehicles' Drivers. *Transportation Research Procedia,* 45**,** 161-168.

Pettigrew, S., Talati, Z. & Norman, R. 2018. The health benefits of autonomous vehicles: Public awareness and receptivity in Australia. *Australian and New Zealand journal of public health,* 42**,** 480-483.

PTV 2018. *PTV VISSIM Manual*. Available: < https://usermanual.wiki/Document/Vissim20102020Manual.1098038624.pdf> (November 20201)

Qin, Y. & Wang, H. 2019. Influence of the feedback links of connected and automated vehicle on rear-end collision risks with vehicle-to-vehicle communication. *Traffic Injury Prevention,* 20**,** 79-83.

Rahman, M. S., Abdel-Aty, M., Lee, J. & Rahman, M. H. 2019. Safety benefits of arterials' crash risk under connected and automated vehicles. *Transportation Research Part C: Emerging Technologies,* 100**,** 354-371.

Raj, A., Kumar, J. A. & Bansal, P. 2019. A Multicriteria Decision Making Approach to Study the Barriers to the Adoption of Autonomous Vehicles. *arXiv preprint arXiv:.12051*.

Rau, P., Yanagisawa, M. & Najm, W. G. Target crash population of automated vehicles. In 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV), 2015. 1-11.

Razali, N. M. & Wah, Y. B. 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics,* 2**,** 21-33.

Reynolds, P., Hurley, S., Torres, M., Jackson, J., Boyd, P. & Chen, V. W. 2000. Use of coping strategies and breast cancer survival: results from the Black/White Cancer Survival Study. *American Journal of Epidemiology,* 152**,** 940-949.

Rodrigue, J.-P., Comtois, C. & Slack, B. 2016. *The geography of transport systems*, Routledge.

Royston, P. & Parmar, M. K. 2013. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Medical Research Methodology,* 13**,** 1-15.

SAE 2018. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. SAE International Warrendale, Pennsylvania. Available: <https://www.sae.org/standards/content/j3016_201806/> (January 2019).

Schoettle, B. & Sivak, M. 2015. A preliminary analysis of real-world crashes involving self-driving vehicles. The University of Michigan, Transportation Research Institute. UMTRI-2015-34,

Sener, I. N., Zmud, J. & Williams, T. 2019. Measures of baseline intent to use automated vehicles: a case study of Texas cities. *Transportation Research Part F: Traffic Psychology and Behaviour,* 62**,** 66-77.

Sinha, A., Chand, S., Wijayaratna, K. P., Virdi, N. & Dixit, V. 2020. Crash Severity and Rate Evaluation of Conventional Vehicles in Mixed Fleets with Connected and Automated Vehicles. *Procedia Computer Science,* 170**,** 688-695.

Sohrabi, S. & Khreis, H. 2020. Burden of disease from transportation noise and motor vehicle crashes: Analysis of data from Houston, Texas. *Environment International***,** 105520.

Sohrabi, S., Khreis, H. & Lord, D. 2020. Impacts of Autonomous Vehicles on Public Health: A Conceptual Model and Policy Recommendations. *Sustainable Cities and Society,* 63**,** 102457.

Sohrabi, S., Khodadadi, A., Mousavi, S.M., Dadashova, B. and Lord, D., 2021. Quantifying the automated vehicle safety performance: A scoping review of the literature, evaluation of methods, and directions for future research. Accident Analysis & Prevention, 152, p.106003.

Sousa, N., Almeida, A., Rodrigues, J. C. & Jesus, E. N. 2017. Dawn of autonomous vehicles: review and challenges ahead. *Proceedings of the ICE-Municipal Engineer,* 171**,** 1-12.

Strand, N., Nilsson, J., Karlsson, I. M. & Nilsson, L. 2014. Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transportation research part F: traffic psychology and behaviour*, 27**,** 218-228.

Taeihagh, A. & Lim, H. S. M. 2018. Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport Reviews,* 39**,** 103-128.

Tainio, M. 2015. Burden of disease caused by local transport in Warsaw, Poland. *Journal of Transport and Health,* 2**,** 423-433.

Teoh, E. R. & Kidd, D. G. 2017. Rage against the machine? Google's self-driving cars versus human drivers. *Journal of Safety Research,* 63**,** 57-60.

Tiwari, G., Bangdiwala, S., Saraswat, A. & Gaurav, S. 2007. Survival analysis: Pedestrian risk exposure at signalized intersections. *Transportation research part F: traffic psychology and behaviour,* 10**,** 77-89.

Train, K. E. 2009. *Discrete choice methods with simulation*, Cambridge university press.

US DOT 2018. Preparing for the Future of Transportation: Automated Vehicles 3.0. US Department of Transportation. Available: < https://www.transportation.gov/av/3> (November 2021)

Virdi, N., Grzybowska, H., Waller, S. T. & Dixit, V. 2019. A safety assessment of mixed fleets with Connected and Autonomous Vehicles using the Surrogate Safety Assessment Module. *Accident Analysis and Prevention,* 131**,** 95-111.

Wang, L., Zhong, H., Ma, W., Abdel-Aty, M. & Park, J. 2020. How many crashes can connected vehicle and automated vehicle technologies prevent: A meta-analysis. *Accident Analysis and Prevention,* 136**,** 105299.

Wang, S. & Li, Z. 2019. Exploring the mechanism of crashes with automated vehicles using statistical modeling approaches. *PloS One,* 14.

Wilde, G. J. 1998. Risk homeostasis theory: an overview. *Injury Prevention,* 4**,** 89-91.

Xie, K., Ozbay, K., Yang, H. & Yang, D. 2019. A new methodology for before–after safety assessment using survival analysis and longitudinal data. *Risk Analysis,* 39**,** 1342-1357.

Xu, C., Bao, J., Liu, P. & Wang, W. 2018. Investigation of contributing factors to extremely severe traffic crashes using survival theory. *International journal of injury control and safety promotion,* 25**,** 141-153.

Xu, C., Ding, Z., Wang, C. & Li, Z. 2019. Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes. *Journal of Safety Research,* 71**,** 41-47.

Yanagisawa, M., Najm, W. G. & Rau, P. Preliminary Estimates Of Target Crash Populations For Concept Automated Vehicle Functions. 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV), 2017, Detroit.

Yang, J., Ward, M. & Akhtar, J. 2017. The development of safety cases for an autonomous vehicle: A comparative study on different methods. SAE Technical Paper. 0148-7191, Available: <https://saemobilus.sae.org/content/2017-01-2010> (January 2019).

Ye, L. & Yamamoto, T. 2019. Evaluating the impact of connected and autonomous vehicles on traffic safety. *Physica A: Statistical Mechanics and its Applications,* 526**,** 121009-121009.

Young, W., Sobhani, A., Lenné, M. G. & Sarvi, M. 2014. Simulation of safety: A review of the state of the art in road safety simulation modelling. *Accident Analysis and Prevention,* 66**,** 89-103.

Yun, H. & Yang, J. H. 2020. Multimodal warning design for take-over request in conditionally automated driving. *European Transport Research Review,* 12**,** 1-11.

Zhang, W., Guhathakurta, S., Fang, J. & Zhang, G. 2015. Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach. *Sustainable Cities and Society,* 19**,** 34-45.

APPENDIX

The appendix provides more details regarding the literature review.

Table A1. Summary of target crash population studies (Reprinted with permission from Sohrabi et al. 2021)

Author	ADS/ADAS	Target Crashes	ODD					Effectiveness	Significant Results
			Road Type	Road Surface Condition	Weather Condition	Lighting Condition	Speed		
Kusano and Gabler (2014)	FCW, PCAM, LDW	18 pre-crash scenarios	NA	NA	NA	NA	NA	NA	• Safety systems can mitigate 20% and 26% of serious injury and fatal crashes, respectively.
Lee and Kockelman (2016)	CACC, LKA, ESC	37 pre-crash scenarios	NA	NA	NA	NA	NA	NA	• Reduction of crash costs by 126 million annually. • Reduction of functional human-years lost by nearly 2 million (per year).
Detwiller and Gabler (2017)	AEB	Transportation-related pedestrian crashes	Urban area	NA	NA	NA	✓	100%	• Employing two different sets of rules resulted in a reduction or mitigation of 40% and 95% of crashes, respectively.
Yanagisawa and Rau (2017)	Level 2 to Level 4	37 pre-crash scenarios	Intersection, ramp, highway, work zone	✓	✓	✓	✓	100%	• L2 to L4 can address 35–250 billion dollars in comprehensive costs and 1100–11,000 fatal crashes annually.
Hendrickson and Harper (2018)	BSM, LDW, and FCW	Lane-change crashes, lane-departure crashes, and rear-end collision	NA	NA	NA	NA	✓	100%	• All technologies together can mitigate 1.3 million crashes annually, including 133,000 injury and 10,000 fatal crashes.
Lubbe et al. (2018)	AEB, LCW, LKA*, ESC	30 pre-crash scenarios	NA	✓	✓	NA	✓	100%	• Fatality reduction from 12–13% (using passive safety systems only) to 45–63% (using advanced ADAS and assuming cautious driving). • Reduction of vulnerable road user fatalities by 33–41%.

Agriesti et al. (2019)	Level 3	Distracted driving, insufficient safety distance, speeding, skidding, road departure	Highways	✓	✓	✓	✓	100%	• 66% of crashes involving AVs and 6.6% of crashes involving conventional vehicles (considering 10% MPR) could be avoided.
Combs et al. (2019)	Pedestria n detection	Transportation-related pedestrian crashes	Urban/rural, intersection/not intersection, freeway/not freeway	NA	NA	NA	✓	100% except for adverse condition (20%)	• Different combinations of sensors can lead to a 30% to 90% reduction of fatal pedestrian crashes.

Note: NA = Not Applicable.

* LKA: Lane Keeping Assistant

Table A2. Summary of AV road test data analysis studies (Reprinted with permission from Sohrabi et al. 2021)

Study	Type of Analysis	Data Source	Approach	Significant Results
Schoettle and Sivak (2015)	Frequency; characteristics of the incident	CA DMV (2014–2015) and Google self-report (2012–2014) (11 crashes)	• Comparing AV and conventional cars' crash rates after adjusting for underreporting. • Descriptive analysis of crash characteristics (vehicle motion at the time of the crash, crash type, and crash severity) and comparison to conventional vehicles.	• Most of the crashes happened while the AV's speed was less than 5 mph. • The rate of rear-end crashes in AVs is higher than conventional cars. • The severity of AV crashes is lower than conventional cars. • The rate of AV crashes is 8 times higher than conventional vehicles.
Kalra and Paddok (2016)	Reliability	Accident rates in the US (2013)	• Estimating number of failure-free miles AVs should drive to reach conventional cars' failure rate using survival analysis. • Estimating the required total number of miles driven to evaluate AVs' failure rate. • Estimating the total number of miles AVs need to drive to demonstrate their failure rate is statistically lower than conventional cars.	• AVs need to drive 1.6 million miles failure-free to be as safe as conventional cars. • AVs need to drive 51 and 61 million miles to be able to test their failure rate and statistically examine if their failure rate is lower than conventional cars, respectively.
(Teoh and Kidd, 2017)	Frequency	CA DMV (2009–2015)*	• Comparing AV (police-reportable) crash rate to conventional cars' crash rate.	• Google self-driving cars are safer than conventional human-driven passenger vehicles (2.19 vs. 6.06 per million VMT).
Favarò et al. (2017)	Frequency; characteristics of the incident; reliability	CA DMV (September 2014 to March 2017) (5326 disengagements and 26 accidents)	• Descriptive analyses of crashes by collision type, location, and manufacturer. • Comparing AV crash rate and number of miles driving until an accident to conventional cars' crash rate and number of miles driving until an accident.	• The rate of crashes was lower for AVs than conventional cars, and AVs will drive longer before an accident (~42,000 vs. 500,000 miles). • Most of the AV crashes happened at intersections. • Rear-end crashes are higher for AVs than for conventional cars.
Matysiak and Razin (2018)	Frequency	CA DMV (2015–2017)	• Comparing AVs' disengagement data to injury and fatal crashes in Europe and US.	• AVs' crash rate is 2 to 3 times higher than conventional cars. • AVs should drive more than 442 million km fatal-free to be considered safer than human-driven cars.
Banerjee et al. (2018)	Frequency; characteristics of the incident; reliability	CA DMV (September 2015 to November 2017)	• Comparing AVs' disengagement rate to conventional cars' accident rates. • Analyzing the cause of disengagement from manufacturer report (after excluding unknown causes). • Comparing to other safety-critical autonomous systems.	• Conventional vehicles were 15−4000 times less likely (depending on the AV manufacturer) than AVs to have an accident. • 64% of disengagements were the result of problems in, or untimely decisions made by, the machine learning system. • In terms of reliability per mission, AVs are 4.22 times worse than airplanes and 2.5 times better than surgical robots.

Xu et al. (2019)	Characteristics of the incident	CA DMV (January 2015 and June 2018) (73 crashes)	• Using bootstrap-based binary logistic regressions to investigate the factors contributing to the collision type and severity of CAV-involved crashes.	• Rear-end and sideswipe crashes are the two predominant collision types, which account for 57.5% and 28.8% of CAV-involved crashes, respectively. • AV driving mode, collision location, roadside parking, rear-end collision, and one-way road are the main factors contributing to the severity level of CAV-involved crashes. • CAV driving mode, CAV stopped or not, CAV turning or not, normal vehicle turning or not, and normal vehicle overtaking or not are the factors affecting the collision type of CAV-involved crashes.
Wang and Li (2019)	Characteristics of the incident	CA DMV (2017 to 2018) (107 crashes) NTSB (2017 to 2018) (6 crashes)	• Investigating the factors contributing to AV crash collision types and severity using logistic regression and decision tree.	• The highway and automated driving mode were identified as the location where severe injuries are likely to happen due to high travel speed. • Collision types of AV-related crashes depend upon the driving mode, location, and whether crashes are associated with yielding to pedestrians/cyclists. • Both ordinal logistic regression and the decision tree models show consistent results.
Li and Zhai (2019)	Reliability	The accident rate on China highways (2008–2015)	• Finding the minimum fault-free distance of AVs to be as safe as conventional cars by inferring the overall distribution from the sample distribution and calculating how much sample size is needed at minimum.	• With a 95% confidence interval, AVs need to drive fault-free for ~226 million km and should be tested for 115,972 million km to be considered as safe as conventional cars.
Petrović et al. (2020)	Characteristics of the incident	CA DMV (2015–2017) (53 accidents)	• Analyzing the type of collision frequencies using descriptive statistics of crash data.	• The rear-end type of collision is statistically more significantly frequent in traffic accidents with AVs.
Boggs et al. (2020)	Characteristics of the incident	CA DMV (2014–2018) (113 crashes)	• Frequentist and Bayesian binary logit model to examine the factors contributing to the AV crashes.	• Speed of conventional vehicle, missing speed, on-street parking, speed limit, driving through arterial and collector, and intersections were positively associated with AV crash assurance. • The number of lanes marked with a centerline and clear weather conditions increase the risk of crashes.
Das et al. (2020)	Characteristics of the incident	CA DMV (2014–2019) (151 crashes)	• Bayesian latent class model to classify AV crashes and to examine the factors. contributing to each class of crashes. • Text mining of AV crash narratives.	• Six classes of AV crashes were identified and associated with turning, multivehicle collisions, dark lighting conditions with streetlights, and sideswipe. • More detailed collision narratives are required to draw reliable conclusions.

* Only Google self-driving car crashes.

Table A3. Summary of traffic simulation studies (Reprinted with permission from Sohrabi et al. 2021)

Authors	**Simulation Information**					**Driving Behavior Model**		**SSM**	**Results**
	Facility Type	**Length**	**Software**	**Technology**	**MPR**	**Conventional Vehicle**	**AV**		
Bahram et al. (2014)	Four-lane highway	6000 m	PELOPS	Highly automated vehicles (HAVs)	0%, 50%, and 100%		The model of HAV controller developed in Simulink; the model is coupled via Xface2 to the interface in PELOPS	• TTC • TET (lower values represent safer situations) • TIT (lower values are associated with higher level of safety)	• TTC = 3.0 sec: 1440, 729, and 16 conflicts for MPRs of 0%, 50%, and 100%, respectively. • At 50% MPR of HAV, the critical situation < 1.5 sec increased remarkably compared to the base scenario. • MPRs of 0%, 50%, and 100% are associated with the TET of 144.1 sec to 72.9 sec and 1.6 sec, respectively. • By increasing the MPR from 0% to 50% and 100%, the TIT changed from 66 to 76.29 and 1.10 sec^2, respectively. • MPR of 50% is not as safe of the other cases since AVs tend to follow other vehicles closely.
Zhang et al. (2015)	Four-lane freeway	7 km	VISSIM	CAV	0%, 10%, 20%, and 30%	Wiedemann 99	Car-following and lateral lane-change decisions coded in C++	• TET • TIT • TERCRI LCC	• Compared to the base scenario: - Providing 1 or 2 exclusive lanes led from −1.8% to −87.1% and −2.1% to −85.3% of lateral conflicts. - Installing 1 or 2 exclusive lanes resulted in +42.4% to −52.90% and +45.7% to −55.2% of longitudinal risk. • Only MPRs of 10% and demands < 6000 veh/h providing exclusive lanes had mainly adverse effects on longitudinal conflicts ranging from 1.8 to −40.4, but for other scenarios with different MPRs and traffic demands, the overall safety improved.
Kockelman et al. (2016)	• Intersection • Freeway on/off-ramp	NA	VISSIM	AV	25%, 50%, 75%, and 100%	NA	NA	TTC	• Bottleneck: 40–88% reduction in the number of conflicts by increasing the AV MPR from 0% to 100%. • 4-leg intersection: 4% reduction in the number of conflicts by increasing the AV MPR from 0% to 100%. • 77% and 31% reduction in the number of conflicts for two other intersections. • 17% increase for another intersection. • Freeway on-ramps/off-ramps: 49% reduction in the conflicts by increasing the MPR from 0% to 100.
Deluka et al. (2018)	Roundabout	NA	VISSIM	AV	0%, 10%, 25%, and 50%	Wiedemann 74	Calibrated Wiedemann 74	TTC and PET	• By increasing the AV MPR from 0% to 50%: - Omisalj roundabout: number of conflicts increased from 0 to 45; the majority of them were rear-end conflicts. • Malinska roundabout: the conflicts increased from 2 to 5, with all the conflicts being rear-end.
Morando et al. (2018)	• Signalized intersection • Roundabout	NA	VISSIM	AV Level 4	0%, 25%, 50%, 75%, and 100%	Wiedemann 99 car-following model with default parameters	Modified Wiedemann 99	TTC	• Intersection: AVs reduced the number of conflicts by 20% to 65%, with an AV MPR of between 50% and 100%. • Roundabout: the number of conflicts was reduced by 29% to 64% with 100% AV penetration rate.
Arvin et al. (2018)	Intersection	NA	SUMO	AV Levels 3 and 5	0%, 7%, 15%, 40%, 60%, 80%, and 100% (for MPR 100, different combinations of AV Level 3 and AV Level 5 were used)	Wiedemann 74	Modified Wiedemann 74	TTC	• Cases with human-driven vehicles, Level 3 and Level 5 AVs: the average crashes decreased from 9 to 0 by increasing the MPR from 0% to 100%. • Cases with AV Level 5 and human-driven vehicles: at low AV MPR (below 40%), the number of crashes increased from 9 to 10. • Cases with AV Level 5 and human-driven vehicles: by increasing the AV MPR (over 40%), the number of crashes reduced from 10 to 0.
Papadoulis et al. (2019)	Three-lane motorway section	4.27 km	PTV VISSIM 9.0 and API	CAV	0%, 25%, 50%, 75%, and 100%	Wiedmann 99	External CAV driver model API written in C++	• TTC PET	• Reduction in conflicts by 12–47%, 50–80%, 82–92%, and 90–94% for MPRs of 25%, 50%, 75%, and 100%, respectively.
Arvin et al. (2019)	Intersection	NA	VENTOS	HAVs and low-level AVs (LAVs)	Various combinations of conventional vehicles, LAVs, and HAVs	ACC model	Wiedemann	• TTC • Driving volatility	• For AV MPR of 0%, an average of 9.43 conflicts was observed. • At AV MPR of 100%, there was a 90.1% improvement compared to the baseline. • Where all the vehicles were HAVs: the intersection became conflict-free. • By increasing the MPR of LAVs and HAVs, the volatility decreased from 8.5 to 5.5 for acceleration. • By increasing the MPR of LAVs and HAVs, the speed volatility decreased from 6.9 to 3.8.

Katrakazas et al. (2019)	A section of highway	4.52 km	VISSIM	AV	NA	Wiedemann 99	NA	• TTC	• The artificial and the real-world datasets indicated that: - If the network-level, real-time collision risk indicates a situation as conflict-prone traffic, the probability of detecting if a vehicle poses a threat to an AV increases by 10%. - When traffic conditions were marked as safe, the prediction did not improve the probability of a road user being a threat for the ego-vehicle. • By using disaggregated traffic data (i.e., 30 seconds), the probability of a traffic participant posing a threat to the ego-vehicle was enhanced by about 6%. • The proposed method allows AVs to change their trajectory, reduce their speeds, or even prompt a passenger to take the controls to ensure safety even when other sensor systems fail since network-level predictions utilize data at a higher temporal interval than the sampling frequency.
Virdi et al. (2019)	Intersection	NA	VISSIM	CAV	0% to 100% (10% incremental)	Wiedemann 74 and Wiedemann 99	Virdi CAV control protocol algorithm	• TTC • PET	• The first 20% MPR of CAVs resulted in: - +22% change in conflicts at signalized intersections. - 87% reduction in conflicts at priority intersections. - −62% change in conflicts at roundabouts. - 33% increase in conflicts at DDI intersections. • At high CAV MPR, a global reduction in conflicts occurred such that the 90% CAV MPR was accompanied by: - −48% change in conflicts at signalized intersections. - 100% reduction in near-miss events at priority intersections. - −98% change in near-crash events at roundabouts. • 81% reduction in conflicts at DDI intersection.
Rahman et al. (2019)	• Arterial segment • Intersection	3.8 miles	VISSIM	CV and CV lower-level automation (CVLLA) (two automated features such as automated braking and lane-keeping assistance)	0%, 40%, 60%, 80%, and 100%	Wiedemann	C++ programming	• TTC • TET • TIT • TERCRI • LCC • NCJ	• Segment: by increasing the MPR from 0% to 100%: - TET decreases from approximately 1750 to 1450 and 1370 for CV and CVLLA, respectively. - TIT decreases from 445 to 345 and 310 for CV and CVLLA, respectively. - TERCRI reduces from 390 to 308 and 265 for CV and CVLLA, respectively. - LCC decreases from 520 to 455 and 405 for CV and CVLLA, respectively. • Intersection: for different evaluated values of TTC and PET thresholds: - Total number of conflicts were decreased by 21–24% for CV technologies compared to base scenario. • Total number of conflicts were reduced by 31–34% for CVLLA compared with that of base condition.
Ye and Yamamoto (2019)	Two-lane road segment	10 km	NA	CAV	10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%	User-defined	User-defined	• TTC • Acceleration rate • Velocity difference	• Reduction in the number of dangerous situations by increasing the MPR depends on traffic density and TTC. • By increasing the MPR from 0% to 100%, the reduction in the dangerous situations falls within 0% and 97%.
Qin and Wang (2019)	Freeway	20 km	MATLAB	CAV	Different MPRs	NA	NA	• TET • TIT	• Average reduction of 75% to 95% depending on the number of feedback links by increasing the CAV MPR. • By increasing the feedback links from 1 to 2, average reduction in collision risks changes from 75% to 95%. • There is not a significant reduction in the number of conflicts between 2, 3, and 4 links.
Mousavi et al. (2020)	Urban unsignalized intersections	NA	VISSIM	AV	0% and 100%	Wiedemann 74	Modified Wiedemann 74	• TTC	• Overall, regardless of the traffic LOS, AVs are capable of decreasing the total number of conflicts by 3.16. • The higher the traffic congestion, the better the performance of AVs compared to conventional vehicles.
Sinha et al. (2020)	Freeway	NA	VISSIM	CAV	0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100%	Wiedemann 99	User-defined driving behavior	• TTC, PET, relative speed	• Manual vehicle–manual vehicle crash rate decreased from 0.9 to 0.0 by increasing the CAV MPR from 0% to 100%. • CAV–manual vehicle crash rate started escalating to 0.3 by increasing the CAV MPR to 90%. • The overall crash rate dropped from 0.9 to 0.0 by increasing the CAV MPR from 0% to 100%.

Arvin et al. (2020)	Intersection	NA	VISSIM	AV CAV	0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100%	Wiedemann 74	AVs: ACC and cooperative ACC (CACC) models CAVs: modified Wiedemann	• Number of longitudinal conflicts • Driving volatility	• Number of conflicts: - Implementing only AVs resulted in a reduction in the number of conflicts from 10 to zero by increasing the MPR from 0% to 100%. - At AV MPR of 10%, the number of conflicts increased compared to the AV MPR of 0%. - By adding coordination into AVs, the number of conflicts decreased steadily from 10 to zero for CAV MPRs of 0% and 100%. - Overall, from MPR of 10% to 90%, the CAV scenarios had fewer number of conflicts than the AV scenarios. • Speed volatility: - For the AV environment, the speed volatility experienced two peaks at MPRs of 40% and 80%. - AV MPRs of 0% and 100% experienced seven and zero conflicts. - For the CAV environment, the number of conflicts decreased constantly from MPR of 0% to 100%. • Speed volatilities in the CAV environments were lower than the AV environments.

Note: NA = Not Applicable.

Table A4. Summary of driving simulator studies (Reprinted with permission from Sohrabi et al. 2021)

Author	Participants' Information			Experiment Factors				AV Challenge	Scenario Parameters	Statistical Tool	Response Variable	Significant Results
	Age	Annual mileage	Driving experience	Facility	Speed	Traffic	Repetition					
Desmond et al. (1998)	18–27	×	2 to 8 years of driving experience	Not considered	80 km/h	×	✓	Fatigue	Perturbing events	ANOVA	• Physical fatigue items, perceptual fatigue items, boredom/apathy • Lateral control such as heading error, deviation of the vehicle	• A similar level of workload. • Better performance recovery in manual driving. • Automated driving results in undermobilizing driver's effort.
Gouy et al. (2012)	20–63	2000–56,000 km	Experience with a driving simulator, at least 1 year of driving experience	Three-lane highway	90 km/h	✓	✓	Platoon environment	Time headway within the platoons	ANOVA	Time headway	• Smaller average and minimum time headway when driving adjacent to AV platoons with short time headway.
Bertholen and Gineyt (2014)	21–29	×	At least 2 years of driving experience	Three-lane highway	Highway: 110 km/h Urban scenario: 70–90 km/h	With/ without	×	Drunk driving	Driving environment (urban area, car following, highway), different alcohol concentration	ANOVA	Number of collisions, mean speed	• Lateral and longitudinal control of the AV is more likely to be impaired compared to strategies adopted in evasive situation.
Strand et al. (2014)	24–65	>10,000 km	No automated driving experience & > 5 yr driving experience	Two-lane undivided rural road	70 km/h	✓	✓	System failures	Automation level, extent of system failure (moderate/ severe/completely)	ANOVA, ANCOVA, Fisher's exact tests	Minimum TTC, minimum time headway, response time, point-of-no-return, number of collisions	• Further automation leads to lower performance of driver. • Drivers performed better at controlling the lower extent of system failure.
Blommer et al. (2015)	40 (24 < 45 yr and 16 > 45 yr)	×	No experience of automated driving	Four-lane undivided roadway	50–70 mph	Light traffic	×	Disengagement	Continuous and scheduled automated driving, secondary tasks	ANOVA	Response time, eye glance behavior, percent eyes-on-road time	• Radio listeners responded significantly faster. • The scheduled driver engagement strategy performed better when visual distraction was used.
Happee et al. (2017)	33.5 (SD = 9)	×	Familiarity with the driving simulator	Three-lane highways	120 km/h	With (30 veh/km)/ Without traffic	✓	Disengagement	Time budget, lane driven, traffic density, secondary tasks	Linear regression, Fisher's exact tests	In total, 19 performance metrics in terms of risk, braking, and steering, such as TTC, clearance toward the obstacle and the roadside, peak accelerations, overshoot, etc.	• AV can cause delayed initial steering and braking, lower TTC, and stronger braking or steering. • No difference between cognitive and visual distraction. • The precision of maneuver remained unaffected.
Gold et al.* (2018)	19–79	×	At least 1 year of driving experience	Three-lane highways	120 km/h	0, 10, 20, 30 veh/km	✓	Disengagement	Time budget, lane driven, traffic density, secondary tasks, repetition of the experiment	Generalized linear regression	TOT, TTC, crash, brake application	• Traffic density (negatively), repetition (positively), and time budget were highly influential. • TOT, TTC, and crash probability showed reliable results.
Lee et al. (2018)	23 below and 7 above 50 years of age	×	Not considered	Three-lane highway	100 km/h	✓	×	Platoon environment	Platoon size, different MPR	ANOVA, logistic regression	Steering magnitude, steering velocity, lane-change duration, lane-change (success/failure)	• Smaller average and minimum time headway when driving adjacent to AV platoons with short time headway.
Kundiger et al. (2018)	18–64	×	Not considered	Three-lane highway	MV: 120 km/h AV: 110 km/h	Light traffic	×	Drowsiness	Age group, different time of the day, different sleepiness category	ANOVA	Karolinska Sleepiness Scale	• Time and driving mode have a significant effect on the development of drowsiness.
(Yun and Yang, 2020)	22–33	×	At least 6 months of driving experience	Four-lane highway	100 km/h	×	✓	Disengagement	Diverse warning combinations (visual, auditory, haptic), disengagement scenarios (planned/unplanned)	MANOVA	• Human behavior metrics: response time, TTL • Vehicle control metrics: SDLP, SRR • Psychological metrics: SCR, AHR	• The multimodal warning method showed superiority over unimodal warnings. • Each modality is preferred in a specific situation (e.g., haptic and auditory modality elicits a more immediate and stable warning, respectively). • Response time in unplanned disengagement is faster than planned events.
(Lee et al., 2020)	25–39	×	More than 1 year of driving experience		×	×	×	Disengagement	Different secondary tasks with different physical/visual/cognitive loads	Non-parametric ANOVA	• Mean longitudinal/lateral acceleration, maximal longitudinal/lateral acceleration, maximum speed, minimum speed, DTC, TTC, SDLP	• Resource allocation associated with each of the non-driving-related tasks did not significantly affect the take-over quality. • The cognitive load of the non-driving-related tasks more effectively affect the longitudinal and lateral control than their physical and visual attributes.

* This study used a series of driving simulator experiments with the same design.