

Towards the assessment of automated-vehicle safety with duration modeling

Soheil Sohrabi

Zachry Department of Civil and Environmental Engineering
Texas A&M University
Texas, USA
Email: sohrabi.s@tamu.edu

Dominique Lord

Professor, Zachry Department of Civil and Environmental Engineering
Texas A&M University
Texas, USA
Email: dlord@civil.tamu.edu

Bahar Dadashova

Texas A&M Transportation Institute
Texas A&M University
Texas, USA
Email: b-dadashova@tti.tamu.edu

Fred Mannering

Professor of Civil and Environmental Engineering
Executive Director, Center for Urban Transportation Research
University of South Florida
Florida, USA
E-mail: flm@usf.edu

ABSTRACT

Ideally, the evaluation of automated vehicles would involve the careful tracking of individual vehicles and recording of observed crash events. Unfortunately, due to the low frequency of crash events, such data would require many years to acquire and potentially place the motoring public at risk if defective automated technologies were present. To acquire information on the safety effectiveness of automated vehicles more quickly, this paper demonstrates an approach that uses the collective crash histories of a group of automated vehicles. To achieve this, a duration modeling approach is applied where the duration is studied as the number of miles driven between crashes in the group of vehicles. Anderson-Darling and Kolmogorov-Smirnov tests are also conducted to statistically assess differences between automated and conventional vehicles. The findings indicate that automated driving was safer, with a higher number of miles between crashes, relative to their conventional vehicle counterparts. The results indicate that the number of miles between crashes would be increased by roughly 27% when switching from conventional vehicles to automated vehicles. Despite limited data, this study can be considered a reasonable initial approximation of automated vehicle safety. With the increasing availability of more complete automated vehicle data, the proposed methodology has the potential to study future automated vehicle innovations and upgrades.

Keywords: Automated Vehicles; Safety; Crash Rate; Duration Analysis; Road Test

1. Introduction

Safety concerns are one of the major issues in developing automated vehicles. Research has argued that a reliable approach to evaluate automated vehicle safety is to assess data gathered from six sources; traffic simulations, driving simulators, target crash populations, road test data analyses, safety effectiveness estimations, and failure risk assessments (Sohrabi et al., 2021). Such a comprehensive approach can potentially account for many elements of automated-vehicle safety, including the possibility that automated vehicle users may exhibit riskier behavior as they overestimate the effectiveness of automated-vehicle technologies, various issues relating to automated vehicle interactions with conventional vehicles, and various automated-vehicle system failure risks (Sohrabi et al., 2021). However, in practice, automated vehicle safety is often evaluated more simply by benchmarking their safety performance against conventional vehicles (Schoettle and Sivak, 2015; Teoh and Kidd, 2017; Favarò et al., 2017; Matysiak and Razin, 2018; Banerjee et al., 2018; Goodall, 2021; Sohrabi et al., 2021). In the extant literature, automated vehicle crash rates have been estimated as either the number of crashes per vehicle-mile traveled (Schoettle and Sivak, 2015, Teoh and Kidd, 2017, Favarò et al., 2017, Goodall, 2021) or the number of disengagements for automated driving systems per vehicle-mile traveled (Matysiak and Razin, 2018, Banerjee et al., 2018). Automated vehicle incident rates have then typically been compared with 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). However, given the disparities in the analogy between automated and conventional vehicles crash rates, mixed conclusions relating to automated vehicle safety have often been drawn.

Despite the advantages of analyzing road test data relative to other automated vehicle safety evaluation approaches, such an analysis has a number of limitations. First, to achieve reliable

evaluations of automated vehicle safety, extensive automated vehicle road tests are required under a variety of traffic/environmental conditions (Kalra and Paddock, 2016, Li and Zhai, 2019). This has been problematic in practice because the potential safety concerns associated with automated vehicle operation, such as imposing crash risks on other road users, have hindered the widespread use of immense automated vehicle road tests (Kalra, 2017). Second, using automated vehicle crash rates reported by manufacturers and comparing them against conventional vehicle crashes rates from police reports (which are plagued by under-reporting) can lead to biased results (Favarò et al., 2017, Banerjee et al., 2018, Matysiak and Razin, 2018). Despite the attempts of researchers to extract police-reported crashes of automated vehicles for such an analysis, and to adjust conventional vehicle crash rates for underreporting to overcome this limitation (Toeh and Kidd, 2017; Schoettle and Sivak, 2015), the analogy between automated and conventional vehicle crash rates can suffer from a false equivalency. Another issue in road test analyses is the reliance solely on the descriptive assessments of automated vehicle crashes, which provides limited information on the likelihood of crash occurrence.

Given the above limitations, there is a need to provide some insights into automated vehicle safety with perhaps more limited but more readily available data. The intent of the current study is to explore this possibility by applying a duration modeling approach to statistically assess the safety of automated vehicles relative to conventional vehicles with limited road-test data. Duration modeling traditionally refers to statistical methods that consider time-to-event. Here, instead of the time-to-event, accumulated miles between events is used (with the event being a collision) as a means for comparing automated and conventional vehicles. Automated vehicle safety is assessed relative to conventional vehicle safety by evaluating functions determining the accumulated miles (durations) between successive crashes in groups of automated and conventional vehicles, instead

of assessing individual vehicles as would typically be done in duration modeling. On the basis of the proposed duration modeling approach, a crash-free expectancy is computed as a means of comparing automated and conventional vehicle safety. To demonstrate the approach, an empirical assessment is undertaken using two comparable sources of data. For conventional vehicles, police, and non-police-reportable crashes were collected from the Second Strategic Highway Research Program's naturalistic driving study (reflecting both major and minor crashes) and, for automated vehicle crashes, data from the California Department of Motor Vehicles Autonomous Vehicle Tester program is used since it provides the same minor and major crash information.

2. Methodological Approach

The likelihood of a crash is typically modeled as the frequency of crashes over some specified time period, and count-data modeling methods such as Poisson Regression and its variants (negative binomial and others) are widely used in the conventional vehicle crash literature (Lord and Mannering, 2010; Mannering and Bhat, 2014). Interestingly, these various count data methods have an implied distribution of the time between crash occurrences. For example, the Poisson regression implies that the time between crash occurrences is exponentially distributed. Thus, an alternative to assessing crash likelihoods as the frequency of crashes would be to directly consider the time until an event occurrence. Herein, we modify this concept somewhat by studying duration between crashes as the number of miles between crash events (instead of time) to implicitly account for crash exposure over time.

Duration/survival modeling approaches seek to not only provide statistical insights into the distribution of durations between events but also account for the possibility that the probability of an event occurring may change over time (or, in the case of this paper, over accumulated mileage).

That is, the probability of a crash occurring may be different given that a crash has not occurred up to some mileage m than it is if a crash has not occurred up to some mileage $m + 1$.

Over the years, duration/survival models have been widely used in other areas of transportation data analysis (Hensher and Mannering, 1994).¹ Duration analysis in the context of traffic safety and crash analysis can be traced back to the early accident occurrence work of Jovanis and Chang (1989) and Chang and Jovanis (1990). Following this, 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 by studying differences in the time until crashes between male and female drivers. Since then, duration modeling has been used in a variety of traffic safety analyses, including, investigating drinking and driving 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 driving risk for older drivers (Caragata and Wister, 2009), comparing intersection crashes (Bagloee et al., 2016), assessing contributing factors to motorcycle crashes (Chen et al., 2018, Balusu et al., 2020) and assessing contributing factors to severe crash events (Xu et al., 2018). More recently, Xie et al. (2019) have used duration analysis for evaluating the impacts of safety treatments in before and after studies.

While the duration-modeling literature in the context of traffic safety has largely focused on individual observations (for example, tracking the crash histories of drivers or vehicles over an extended period of time), we argue here that insights can be extracted from duration analyses using

¹ Examples include studies of the duration of highway closures (Jones et al., 1991; Nam and Mannering, 2000; Chung, 2010), the duration of roadway incidents (Nam and Mannering, 2000; Stathopoulos and Karlaftis, 2002; Kang and Fang, 2011; Hojati et al., 2013), the duration until roadway pavement failure (Loizos and Karlaftis, 2005; Anastasopoulos and Mannering, 2015), the duration between trips (Mannering and Hamed, 1990; Hamed and Mannering, 1993; Ettema et al., 1995; Bhat, 1996a,b; Niemeier and Morita, 1996; Wang, 1996; Kitamura et al., 1997; Kharoufeh and Goulias, 2002; Bhat et al., 2004), the duration until crisis evacuation (Hasan et al., 2013), the duration of travel-generating activities (Kim and Mannering, 1997; Van den Berg et al., 2012), the duration of vehicle overtaking Maneuvers (Vlahogianni, 2013; Bella and Gulisano, 2020), the duration of vehicle deceleration time (Haque and Washington, 2015; Bella and Silvestri, 2016; Bella and Silvestri, 2018), and many others.

group-level data where the miles driven between crash events in a group of automated vehicles can be compared with the miles driven between crash events in a group of conventional vehicles.² To formulate this approach in the context of duration/survival analysis methods, the miles driven between crashes is specified with a survivor function which gives the probability that duration between successive vehicle crashes in the vehicle group will be greater than or equal to some mileage x . With this, let X be the number of miles between group crashes with a distribution function $f(x)$. The probability of survival (being crash-free) beyond x miles is shown as (Washington et al., 2020):

$$S(x) = \Pr(X > x) = \int_x^{\infty} f(x)dx \quad (1)$$

where $S(x)$ is the survivor function. Since X is a continuous variable, the survivor function is a strictly decreasing function. The cumulative distribution function of number of miles between crashes can then be written as:

$$F(x) = 1 - S(x) \quad (2)$$

The survivor function represents the safety reliability of vehicles, and thus the cumulative distribution $F(x)$ represents the failure function and is strictly increasing. The first derivative of this cumulative distribution with respect to distance gives the density function $f(x) = dF(x)/dx$. With this, the instantaneous rate of failure (crashes) is represented by the hazard function, $h(x)$, as (Washington et al., 2020):

² Tracking individual users or vehicles can lead to a large number of right-censored observations, that is, where a crash is not observed during the study period. For example, in the Balusu et al. (2020) study of the time until motorcyclists' first crash only 3095 of the 43,856 motorcyclists studied were observed to have crashes 5 years after passing a motorcycle training course, thus leaving 40,761 right-censored observations. Although right censoring can be readily handled in duration modeling (Washington et al., 2020), with the grouped vehicle data used in the current study only the very last observation will be right censored, thus reducing the occurrence of this issue.

$$h(x) = f(x)/[1 - F(x)] \quad (3)$$

where $h(x)$ is the conditional probability that a crash will occur in the vehicle group between accumulated crash-free mileage x and $x + dx$, given that a crash has not occurred up to mileage x (since the last crash). In words, $h(x)$ gives the rate at which crash-free durations are ending at mileage x (the crash-free state that would end with the occurrence of a crash), given that there has not been a crash up to mileage x since the last crash occurred.

In duration modeling, to account for the effect of explanatory variables (variables that increase or decrease the accumulated mileage between crashes), proportional hazards and accelerated failure approaches are commonly used (Washington et al., 2020). The commonly used accelerated failure approach has the explanatory variables accelerate accumulated mileage in a baseline survivor function, S_0 (the survivor function when the effect of all explanatory variables are zero). In accelerated failure models, the natural logarithm of the accumulated mileage between crashes can be expressed as a linear function of explanatory variables as,

$$\ln(X) = \boldsymbol{\beta}\mathbf{V} + \varepsilon \quad (4)$$

where \mathbf{V} is a vector of explanatory variables, $\boldsymbol{\beta}$ is a vector of estimable parameters, and ε is an error term, giving the conditional survivor and hazard functions as, (Washington et al., 2020),

$$S(x | \mathbf{V}) = S_0[xe^{\boldsymbol{\beta}\mathbf{V}}] \quad (5)$$

$$h(x | \mathbf{V}) = h_0[xe^{\boldsymbol{\beta}\mathbf{V}}]e^{\boldsymbol{\beta}\mathbf{V}} \quad (6)$$

where $S_0(x)$ and $h_0(x)$ are underlying baseline survivor and hazard functions, respectively (when all elements of the explanatory variable vector \mathbf{V} are zero). Estimation of survivor and hazard functions in a fully parametric context requires a distributional assumption which, depending on the statistical fit, can include commonly used distributions such as the exponential, Weibull and

log-logistic.³ The estimation process, undertaken with standard maximum likelihood methods (Washington et al., 2020), includes the estimation of both the parameters of the distribution (for example, shape and scale parameters) and the vector of estimable parameters β associated with the explanatory variable vector \mathbf{V} . The choice of a distribution has implications for the shape of the hazard functions. For example, the exponential distribution implies a constant hazard where the likelihood of a crash does not change as crash-free mileage accumulates ($dh(x)/dx = 0$), the Weibull distribution allows the hazard to increase ($dh(x)/dx > 0$) or decrease ($dh(x)/dx < 0$) monotonically over accumulated mileage, and the log-logistic distribution allows the hazard to be non-monotonic (for example increasing up to some mileage x and then decreasing thereafter).⁴

Regarding the choice of distributions, in the more common application, where individual driver/vehicles are tracked, one might expect the hazard to; increase as mileage is accumulating (meaning a crash becomes more likely the longer the crash-free duration extends), decrease as mileage is accumulating (meaning a crash becomes less likely the longer the crash-free duration extends), or increase up to some mileage and then decrease afterward. The logic behind such non-constant hazards is that drivers' risk profiles and experience may change over time and affect their crash probabilities. As an example, Balusu et al. (2020) found that many motorcyclists had increasing hazards (crash likelihoods) initially after passing their motorcycle training courses but then decreasing hazards after a month or so after passing the training course as their accumulated experience made them less likely to crash. In contrast to tracking individual driver/vehicles, in

³ There are also estimation approaches where no distributional assumption is made (non-parametric approaches) or where only the parametric effect of the explanatory variables is assumed and the underlying distribution is not assumed (semi-parametric), with the Cox proportional hazards model being the most commonly used semi-parametric approach. While at first thought non-parametric and semi-parametric approaches seem appealing, they often seriously limit the inferences that can be drawn from the estimation process.

⁴ There are numerous extensions of parametric models that add mixing distributions to account for potential unobserved heterogeneity in the data. These include the Weibull model with gamma distributed heterogeneity (Nam and Mannering, 2000) and a full random parameters modeling approach with possible correlation among parameters (Washington et al., 2020; Balusu et al., 2020).

tracking a group of conventional vehicles, the expectation would be that the hazard would be constant because one would not expect a learning effect to influence the time between crashes of any two vehicles in the vehicle group which would likely be geographically separated. For tracking individual automated vehicles, one might also expect a non-constant hazard rate because learning algorithms within the automated systems and component failures will change over time. However, in the group context, a constant hazard would also be expected since there should not be a relationship between the crashes of geographically separated vehicles.

It is also possible to assess the suitability of various distributions of the miles between crashes for automated and conventional vehicles, and to statistically compare automated and conventional vehicle failure function by using Kolmogorov-Smirnov and Anderson-Darling tests (Razali and Wah, 2011). Kolmogorov-Smirnov and Anderson-Darling tests are two sample tests that can be used for comparing both parametric and non-parametric failure functions. For instance, in the case of comparing automated and conventional vehicles, the null hypothesis herein would be that the automated vehicle parametric failure function $F_{AV}(x)$ is the same as the conventional vehicle failure function $F_{CV}(x)$. The Kolmogorov-Smirnov test statistic is based on the maximum distance of two distributions,

$$D_{1,2} = \text{Sup}_x |F_{AV}(x) - F_{CV}(x)|, \quad (7)$$

where ‘‘Sup’’ stands for Supremum function. The Kolmogorov-Smirnov test statistic $D_{1,2}$ is then compared to critical values $D_{1,2,\alpha}$ for desired significance level α , with the critical values are estimated from the Kolmogorov distribution (Razali and Wah, 2011). If $D_{1,2}$ exceeds the $D_{1,2,\alpha}$, then the null hypothesis (H_0) can be rejected and one conclude the $F_{AV}(x)$ and $F_{CV}(x)$ are different distributions.

Unlike the Kolmogorov-Smirnov test, the Anderson-Darling test finds the difference between two distributions giving more weight to the differences between the tails of the distributions $F_{AV}(x)$ and $F_{CV}(x)$. While the test hypothesis is defined similar to the Kolmogorov-Smirnov test, the difference between the distributions is defined as (Anderson, 2011)

$$W_n^2 = n \int_{-\infty}^{\infty} [F_{AV}(x) - F_{CV}(x)]^2 \psi(F_{CV}(x)) dF_{CV}(x) \quad (8)$$

where $\psi(z)$ is the weight function such that $\psi(z) > 0$ and $\psi = [F_{CV}(z)(1 - F_{CV}(z))]^{-1}$. 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 8 can be written as:

$$A_n^2 = -n - \frac{1}{n} \sum_{j=1}^n (2j - 1) [\log u_{(j)} + \log(1 - u_{(n-j+1)})] \quad (9)$$

where $u_{(j)} = F_{CV}(x_{(j)})$ and $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ is the ordered sample. The Anderson-Darling test critical values are estimated for different distribution functions (Jäntschi and Bolboacă, 2018). Similar to Kolmogorov-Smirnov test, null hypothesis that the distributions are equal can be rejected if Anderson-Darling test statistics (A_n^2) exceed the critical values.

Other methods of comparing automated and conventional vehicle safety would include comparisons of the mileage between crashes distributions as well as an assessment of Restricted Mean Survival Time (RMST), which has been commonly used in duration/survival data comparisons (Royston and Parmar, 2013, Harhay et al., 2018). The RMST of a random variable X , $\mu(x^*)$, is the expected value of $\min(X, x^*)$ (the area under the survivor curve $S(x)$ up to x^*)

$$\mu(x^*) = E(\min(X, x^*)) = \int_0^{x^*} S(x) dx = \int_0^{x^*} (1 - F(x)) dx \quad (10)$$

Since X is the number of miles between crashes, RMST can be interpreted as the crash-free expectancy until x^* miles. For example, automated vehicle's crash-free 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 automated vehicle operation. The crash-free expectancy can be defined similarly for conventional vehicles. Comparing the conventional and automated vehicle RMST at x^* provides insight into how safe a conventional vehicle is in comparison to an automated vehicle.

Another useful metric for evaluating automated vehicle safety can be estimating the ratio of the restricted mean survival time of automated and conventional vehicles. For the entire automated and conventional vehicle operations, when $X \rightarrow \infty$, the calculated ratio gives the relative safety effectiveness (SE) of automated vehicles relative to conventional vehicles as

$$SE = \frac{RMST_{AV}}{RMST_{CV}} = \frac{\mu_{AV}(X)}{\mu_{CV}(X)} = \frac{E_{AV}(X)}{E_{CV}(X)} = \frac{AV \text{ crash free expectancy}}{CV \text{ crash free expectancy}} \quad (11)$$

where X is miles between crashes, $E_{AV}(X)$ is expected value of automated vehicles' miles between crashes and $E_{CV}(X)$ is the expected value of conventional vehicle miles between crashes. A value of SE larger than 1 implies that automated driving is safer and a value of less than one implies conventional vehicles are safer.

3. Empirical Study

The empirical study is undertaken to evaluate automated vehicle safety using comparable automated and conventional vehicle data. Specifically, the issue of false equivalency between automated and conventional vehicles crashes, stemming from the limitations in the availability of conventional vehicles' non-police-reported crashes (see Mannering and Bhat, 2014 for issues relating to the under-reporting inherent in police-reported crashes), is addressed. This is achieved by sourcing conventional vehicle crashes from a naturalistic driving study database, which

includes both minor and major crashes (hence making it comparable to automated vehicle crashes reported by auto manufacturers). Unfortunately, comparable available data are quite limited due to the proprietary data associated with automated vehicles. This data availability issue will necessitate a limited model without extensive explanatory variables influencing the durations (miles) between crashes. This makes sense in the group-vehicle context adopted herein, where relevant explanatory variables would be limited to global factors affecting the vehicle-group as a whole.⁵ However, the possibility of having extensive vehicle-based data in the future will enable the use of a wide range of vehicle specific explanatory variables that can affect hazard and survivor functions, as previous studies have used in studying driver-based data (Mannering, 1993; Balusu, 2020).

⁵ During the various model estimations, a variety of mixing distribution were used to account for possible unobserved heterogeneity in the data (Mannering et al., 2016). However, the limitations of the data (particularly in terms of number of observations) prevented these heterogeneity models from producing estimation results that were statistically different from the conventional fixed-parameter models.

3.1 Automated Vehicle Safety Evaluation Framework

A three-step framework is proposed to evaluate automated vehicle safety based on duration modeling and the available data (Figure 2). First, despite the challenges in automated vehicle crash availability, a database is created by sourcing automated vehicle crashes from the California Department of Motor Vehicles and combining them with conventional vehicle crashes from the naturalistic driving study. In the second step, the proposed methodology in this study is examined using the created database. In this step, the miles driven between successive crashes in automated and conventional vehicle groups is studied, assuming an exponential distribution in an accelerated failure modeling approach. Third, the safety performance equivalency of automated and conventional vehicle groups is tested by comparing miles-between-crash cumulative distribution functions (failure functions). Finally, the relative safety of automated and conventional vehicle groups is assessed using each group's crash-free expectancy.

3.2 Automated Vehicle Crash Data

Automated vehicle road tests have become more common in recent years, with the National Highway Traffic Safety Administration (NHTSA) indicating that 25 automated vehicle manufacturers and developers are testing their cars on United States public roads as of May 2021⁶. Although automated vehicle 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 public⁷. For example, the California Department of Motor Vehicles automated vehicle tester program was established in 2014 with the aim of testing automated vehicles with fallback users (test vehicles require a human in the driver seat who can take control of the vehicle at any time).

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

⁷ There are currently federal laws or rules that are under consideration by NHTSA for the mandatory reporting of all crashes involving automated vehicles occurring in the US (https://www.nhtsa.gov/sites/nhtsa.gov/files/2021-06/Standing_General_Order_2021_01-digital-06292021.pdf).

Under this California vehicle tester program, all manufacturers testing automated vehicles on public roads are mandated to report crashes involving an automated vehicle within ten days after the collision.

For the current study, automated vehicle crash data are sourced from the California autonomous vehicle tester program. Crashes that occurred over a year of automated vehicles operation on California public roads are investigated. As of November 2020, 59 permit holders were testing their automated vehicles under this program. California autonomous vehicle tester program defined automated vehicles 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.”⁸ According to this definition, vehicles equipped with one or more advanced driver assist systems (Levels 1 and 2 of automation) are not tested in this program, and the autonomous vehicle tester program is limited to testing Level 3 of automation.⁹

Based on California autonomous vehicle tester regulations, automated vehicle companies are mandated to report automated vehicle-involved crashes in fewer than ten days after the time of the crash, and it can thus be assumed that the crash reports contain all automated vehicle-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 automated driving systems 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 and

⁸ Sourced from <https://www.dmv.ca.gov/portal/uploads/2020/06/Adopted-Regulatory-Text-2019.pdf> (May 2021)

⁹ Level 3 operation is defined as a level of automation that enables the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions and in those conditions to rely heavily on the vehicle to monitor for changes in those conditions requiring transition back to driver control. The driver is expected to be available for occasional control, but with sufficiently comfortable transition time (Mannering and Washburn, 2020).

the number of miles it was operated during each month. No information regarding the environment under which automated vehicles were tested is publicly available.

As of the time of developing this study, automated vehicle road test data were available until November 2020. Figure 1 represents the distribution of crashes and vehicle miles travelled from January 2019 to November 2020. The data for the year 2019 is used, given the restriction in automated vehicle testing, and road traffic in general, because of the global pandemic in 2020. Automated vehicle testing data in 2019 includes 651 unique vehicles from 30 autonomous vehicle tester permit holders. The tested automated vehicles were driven 2,849,850 miles in 2019 and were involved in 105 crashes. The crash data were manually extracted from the crash reports. However, because the crash reports do not include the vehicle identification number, the daily automated vehicles vehicle miles travelled is estimated using the monthly vehicle miles travelled reports, assuming no variations in daily vehicle miles travelled in a month. The miles between crashes is then approximated using the daily vehicle miles travelled and the time of the crash.

3.3 Naturalistic driving study data

The naturalistic driving study data are collected as part of the Strategic Highway Research Program 2 (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 the naturalistic driving study data were collected are Seattle, WA; Bloomington, IN; Buffalo, NY; State College, PA; Durham, NC; and Tampa, FL. The naturalistic driving study includes the driver information, the vehicle driven, and driver trip information, as well as the potential crash and near-crash events that occurred in each trip. The naturalistic driving study data consist of more than 5.4 million miles, more than 1 million hours of recorded videos, and more than 1,500 crashes.

The naturalistic driving study has four types of data and can be requested using an available query tool. 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 participants' vehicles. 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. These data also include 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 naturalistic driving study data are compiled to be consistent with the automated vehicle crash and vehicle miles travelled data. A sample of consecutive trips is randomly selected so that 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. For the naturalistic driving study 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 driven on an individual trip before a crash is 9.9 miles (with a minimum less than 1 mile and a maximum trip length until a crash of 215.6 miles). Given that the time of crashes is known, miles driven between crashes is calculated using the trip lengths in the naturalistic driving study dataset. To this end, the total miles of travel in the sample for consecutive crash reports is determined.

3.4 Failure rate, miles between crashes, and empirical failure function estimation

A descriptive analysis of automated and conventional vehicles crash datasets shows a higher crash rate was observed for conventional vehicles compared to automated vehicles. In 2,849,50 miles of driving conventional vehicles, 130 crashes were observed, which is higher than 105 crashes automated vehicles were involved in while driving the same mileage. Consequently,

the rate of automated vehicle crashes is 20% lower than conventional vehicles. The average of automated vehicle miles-to-crash is higher than conventional vehicles where, on average, automated vehicles were involved in crashes every 27,399 miles in comparison with 21,634 miles for conventional vehicles. Table 1 summarizes crash frequency and miles-to-crash statistics.

The cumulative distribution of miles between crashes represents the empirical failure function, $F(x)$ (see Equation 2) and are shown in Figure 3. The likelihood of automated and conventional vehicles being involved in crashes can be compared using these functions. For example, after 50,000 miles, the likelihood of being involved in a crash was observed to be 86% and 84% for automated and conventional vehicles, respectively. In this regard, the survival probability, as a complement of failure probability, would be 14% and 16% for conventional and automated vehicles. The probability of crashes was higher than 50% for automated vehicles after driving 15,000 miles, in contrast with 13,000 for conventional vehicles. Although from Figure 3, automated vehicle crash likelihood is lower than conventional vehicles at (almost) every mile of driving, the significance of this difference needs to be investigated statistically.

3.5 Parametric duration model estimation

Using standard maximum likelihood estimations, various parametric functions were fitted to the automated and conventional vehicle data. As discussed previously, the exponential distribution function, with its constant hazard rate, makes the most theoretical sense since one would not expect the hazard functions to increase or decrease over time as they would if individual drivers were being followed in conventional vehicles (where drivers accumulating experiences may affect the probability of a crash as crash free miles are accumulated) or if individual vehicles were followed in the case of automated vehicles (where software updates and changes in sensor

accuracy over accumulating crash free miles may affect the probability of a crash). With parameter $\lambda > 0$, the exponential density function is,

$$f(x) = \lambda e^{-\lambda x}, \quad (12)$$

with a hazard function,

$$h(x) = \lambda \quad (13)$$

Estimation of this model using the log of the mileage between crashes (see Equation 6) gives the estimation results shown in Table 2. Models are presented in this table; one with just a constant term, and the other with a constant term and an indicator variable that is equal to one if the crash-free duration data was from the automated vehicle group and zero otherwise. Table 2 shows that the automated vehicle indicator is statistically significant. The appropriate likelihood ratio test for this is,

$$X^2 = -2[LL(\boldsymbol{\beta}_c) - LL(\boldsymbol{\beta}_{ci})] \quad (14)$$

where $LL(\boldsymbol{\beta}_c)$ is the log-likelihood at convergence of the model with the constant term only and $LL(\boldsymbol{\beta}_{ci})$ is the log-likelihood at convergence of the model with both the constant and the automated vehicle indicator. The statistic in Equation 14 is χ^2 distributed with one degree of freedom (the additional automated vehicle indicator estimated) and has a value of 4.42. This indicates that the null hypothesis that the two models are the same (which is equivalent to saying that there is no difference in the safety of automated and conventional vehicles) can be rejected with 97 percent confidence, suggesting automated vehicles are safer, with the positive value of the automated vehicle indicator meaning a longer duration (miles driven) between crashes.¹⁰

¹⁰ As an alternative test, the data could be split and separate models for automated and conventional vehicles could be estimated with just a constant term. The likelihood ratio test would then be $X^2 = -2[LL(\boldsymbol{\beta}_T) - LL(\boldsymbol{\beta}_a) - LL(\boldsymbol{\beta}_c)]$ where $LL(\boldsymbol{\beta}_T)$ is the log-likelihood at convergence of the data with all vehicles, $LL(\boldsymbol{\beta}_a)$ is the log-likelihood at convergence using only automated vehicle between-crash mileage, and $LL(\boldsymbol{\beta}_c)$ is the log-likelihood at convergence using only

However, because access to detailed data is not available and many important explanatory variables are missing, the data may not fit the exponential distribution well. To test the suitability of the exponential distribution, we employed Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests with a 95% confidence interval (Table 3). The estimated one-way sample test statistics for the examined distributions show that the Anderson-Darling and Kolmogorov-Smirnov tests reject the null hypothesis that the data samples follow an exponential distribution. In fact, a variety of other parametric distributions were also tested and rejected, again reflecting the effect of important missing explanatory variables an issue that will afflict any group-level duration analysis.

3.5 Comparing automated and conventional vehicle failure function

Given the above, non-parametric tests are conducted to determine if the automated vehicle empirical failure function is statistically different from the parametric conventional vehicle failure function using two-sample non-parametric Kolmogorov-Smirnov and Anderson-Darling tests and a 95% confidence level. The Kolmogorov-Smirnov test (see the previous description) produced a test statistic of 0.261 which is higher than the critical value of 0.009 at the 95% confidence level, thus indicating that the null hypothesis that automated and conventional vehicle failure functions are the same can be rejected with high confidence. Similarly, the Anderson-Darling test statistic was 9.715, which is higher than the critical value of 2.492 at the 95% confidence level, again indicating that the null hypothesis that automated and conventional vehicle failure functions are the same can be rejected with high confidence. Consequently, it can be concluded that the automated vehicle failure function is statistically inconsistent with the conventional vehicle failure

conventional vehicle between-crash mileage. However, with the model estimated herein, it can be readily shown that this test will produce the exact same test statistic $X^2 = 4.42$ and thus identical conclusions are drawn.

function and, therefore, with 95% confidence, automated vehicle failure probability is lower than conventional vehicles' failure probability.

3.6 Crash-free expectancy

Table 4 compares the automated and conventional vehicle crash-free expectancies at different mileage (see Equations 10 and 11). The results show that in the next 10,000 miles of driving, the crash-free expectancies in the vehicle groups are roughly 4,000 miles for conventional vehicles and 7,500 for automated vehicles. In the next 150,000 miles of driving, the crash-free expectancy would be increased to nearly 6,000 miles when shifting from conventional vehicles to automated vehicles. In other words, in comparison with a conventional vehicle, on average, the automated vehicle group can drive an additional 6,000 miles before observing a crash which represents a 27% advantage in favor of vehicle automation. Consequently, the safety effectiveness of automated vehicles is estimated as 1.27 relative to conventional vehicles (see Equation 11).

4. Discussion

This study formulates a duration analysis for automated vehicle safety evaluation (using miles between crashes) and demonstrates the necessity of rethinking the approach to crash modeling for evaluating automated vehicle safety. The preliminary analysis of automated and conventional vehicle crash data (based on groups of vehicles) sets indicated that automated vehicles' crash rates are 20% lower than conventional vehicles. On average, automated vehicles 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 automated and conventional vehicles' empirical distribution of the miles driven between crashes represents the lower failure (crash) probability for automated vehicles. Using the Kolmogorov-Smirnov and Anderson-Darling tests, the hypothesis of whether the automated vehicle failure function is different from the conventional vehicles'

failure function was examined, and it was concluded that the difference between automated and conventional vehicles' failure functions are different at the 95% confidence interval level. The findings of the proposed statistical analysis herein imply that Level 3 of automation testing in California is safer than conventional vehicles. Also, comparing the crash-free expectancy between the conventional vehicle and automated vehicle shows a 27% improvement in 150,000 miles of road operation.

While the empirical analysis presented herein is a demonstration of the methodology at the group-of-vehicles level, tracking individual vehicles and drivers (once such data become readily available) has the potential to greatly expand the insights and inferences drawn by allowing many explanatory variables to influence the accumulated mileage between crashes and, perhaps more importantly, such individual data will allow the exploration of various hazard function shapes to determine how crash probabilities vary conditionally on the number of accumulating crash-free miles.

It is important to mention the empirical analysis herein is not without its limitations. To begin, some limitations are inherent in all road test data analysis in that it was assumed that automated vehicle manufacturers reported all crashes, as mandated. According to the California Department of Motor Vehicles, the autonomous vehicle tester 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 automated vehicle testing operational design domain, since Level 3 of automation is designed to operate in various (although limited) operational design domain (SAE, 2016), it was assumed that automated vehicles were tested on roads with different functional classifications and are comparable with conventional vehicles. In addition, because Level 3 automation requires fallback users to intervene in certain

situations, the disengagement from the automated driving systems potentially 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. Because the automated vehicle crash dataset is collected as part of the autonomous vehicle tester program, it is expected that the fallback users are both experienced and constantly pay attention to automated vehicles performing the dynamic driving task. However, it is possible that our findings may overestimate automated vehicle safety since fallback drivers may be a selected group of drivers with low disengagement risk. Along these same lines, there is also the possibility that the overall risk profiles of automated vehicle drivers differs in a fundamental way from that of the naturalistic driving data drivers. That is, because we do not observe the same drivers using both technologies (automated and conventional), there is a potential for a selectivity bias which could influence the safety conclusions (if automated drivers are inherently safer the safety effectiveness of automated vehicles will be overestimated and if they are less safe it will be underestimated). This issue could be addressed with selectivity bias correction methods if data on individual vehicles/drivers were available.¹¹

There is also the issue that the automated vehicle crash reports used herein do not include the time of the crash and the vehicle mileage, necessitating that the estimated miles between crashes be 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 automated vehicle. Nevertheless, we do not anticipate this issue would have a significant effect on the conclusions of this study.

¹¹ This issue is similar to the general problem with crash data in general in that the riskiest drivers are over-represented in the sample and the safest drivers are under-represented. This creates an identification problem as discussed in Mannering et al. (2020).

Another limitation is that the empirical study was conducted using only one year of automated vehicles operation, which translated into 2.8 million miles. While 2.8 million miles are expected to be sufficient to capture the cumulative distribution of the miles between crashes, future studies can expand this timeframe to further assess the impact of additional accumulated miles.

Finally, the current study focused on crash occurrence but not on crash severity (resulting injuries or property damage). While the severity of automated vehicle crashes has been studied in the literature (Xu et al., 2019, Wang and Li, 2019), the proposed framework can be used to compare automated vehicle failure function by crash severity once enough road test data are available. In fact, competing risk duration models would be a natural methodological approach for addressing injury severity in a duration model context (Washington et al., 2020). This would be a fruitful direction for future work.

Acknowledgments

This study was partially supported by Robert Wood Johnson Foundation Grant ID 76489 *Quantifying the Harms and Benefits of Connected and Automated Vehicle Technologies on Public Health and Equity* and the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation's University Transportation Centers Program (Grant No. 69A3551747115). The authors gratefully acknowledge the support provided by the Center for Teaching Old Models New Tricks (TOMNET) and the Center for Transportation, Environment, and Community Health (CTECH), University Transportation Centers sponsored by the US Department of Transportation through Grant No. 69A3551747116 and Grant No. 69A3551747119, respectively. The authors would like to thank Dr. Sue Chrysler and Dr. Johanna Zmud for their support and insights.

REFERENCES

- Anastasopoulos, P., Mannering, F., 2015. Analysis of pavement overlay and replacement performance using random parameters hazard-based duration models. *Journal of Infrastructure Systems*, 21(1), 04014024.
- Anderson, T., 2011. Anderson-Darling Tests of Goodness-of-Fit. *International Encyclopedia of Statistical Science* 1, 52-54.
- Anderson, T., Darling, D., 1954. A test of goodness of fit. *Journal of the American Statistical Association* 49, 765-769.
- Bagloee, S., Asadi, M., 2016. Crash analysis at intersections in the CBD: A survival analysis model. *Transportation Research Part A* 94, 558-572.
- Balusu, S., 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., Jha, S., Cyriac, J., Kalbarczyk, Z., Iyer, R., 2018. 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. IEEE, 586-597.
- Bella, F. and Gulisano, F., 2020. A hazard-based model of the motorcyclists' overtaking duration. *Accident Analysis and Prevention* 141, 1-12.
- Bella, F., Silvestri, M., 2016. Driver's braking behavior approaching pedestrian crossings: a parametric duration model of the speed reduction times. *Journal of Advanced Transportation* 50(4), 630-646.
- Bella, F., Silvestri, M., 2018. Survival Model of Drivers' Speed Reduction Time at Bicycle Crossroads: A Driving Simulator Study. *Journal of Advanced Transportation*, 4738457.

- Bhat, C., 1996a. A generalized multiple durations proportional hazard model with an application to activity behavior during the evening work-to-home commute. *Transportation Research Part B* 30(6), 465-480.
- Bhat, C., 1996b. A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity. *Transportation Research Part B* 30(3), 189-207.
- Bhat, C., Frusti, T., Zhao, H., Schönfelder, S., Axhausen, K., 2004. Intershoping duration: an analysis using multiweek data. *Transportation Research Part B* 38(1), 39-60.
- Caragata N., 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(4), 474-484.
- Chang, H.-L., Jovanis, P., 1990. Formulating accident occurrence as a survival process. *Accident Analysis and Prevention* 22(5), 407-419.
- Chen, S.-J., Chen, C.-Y., Lin, M.-R. 2018. Risk factors for crash involvement in older motorcycle riders. *Accident Analysis and Prevention* 111, 109-114.
- Chung, Y., 2010. Development of an accident duration prediction model on the Korean Freeway Systems. *Accident Analysis and Prevention* 42(1), 282-289.
- Ettema, D., Borgers, A., Timmermans, H., 1995. Competing risk hazard model of activity choice, timing, sequencing, and duration. *Transportation Research Record* 1493, 101-109.
- Favarò, F. M., Nader, N., Eurich, S., Tripp, M., Varadaraju, N. 2017. Examining accident reports involving autonomous vehicles in California. *PLoS one*, 12, e0184952.
- Ferrante, A. M., Rosman, D. L., Marom, Y., 2001. Novice drink drivers, recidivism and crash involvement. *Accident Analysis and Prevention* 33(2), 221-227.

- Fu, H., 2008. Identifying repeat DUI crash factors using state crash records. *Accident Analysis and Prevention* 40(6), 2037-2042.
- Goodall, N 2021. Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data. *Accident Analysis and Prevention* 154, 106056.
- Hamed, M., Mannering, F., 1993. Modeling travelers' post work activity involvement: toward a new methodology. *Transportation Science* 27(4), 381-394.
- Haque, M., Washington, S., 2015. The impact of mobile phone distraction on the braking behavior of young drivers: a hazard-based duration model. *Transportation Research Part C*, 50, 13-27.
- Happee, R., Gold, C., Radlmayr, J., Hergeth, S., Bengler, K. 2017. Take-over performance in evasive manoeuvres. *Accident Analysis and Prevention* 106, 211-222.
- Harhay, M. O., Porcher, R., Cantu, E., Crowther, M., Christie, J., Thabut, G., Donaldson, G., 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(5), 684-687.
- Hasan, S., Mesa-Arango, R., Ukkusuri, S., 2013. A random-parameter hazard-based model to understand household evacuation timing behavior. *Transportation Research Part C* 27, 108-116.
- Hensher, D., Mannering, F., 1994. Hazard-based duration models and their application to transport analysis. *Transport Reviews* 14(1), 63-82.
- Hojati, A., Ferreira, L., Washington, S., Charles, P., 2013. Hazard based models for freeway traffic incident duration. *Accident Analysis and Prevention* 52, 171-181.

- Jäntschi, L., Bolboacă, S., 2018. Computation of probability associated with Anderson–Darling statistic. *Mathematics* 6(6), 88.
- Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. *Accident Analysis and Prevention* 23(4), 239-255.
- Jovanis, P., Chang, H.-L. 1989. Disaggregate model of highway accident occurrence using survival theory. *Accident Analysis and Prevention* 21(5), 445-458.
- Kalra, N., 2017. Challenges and approaches to realizing autonomous vehicle safety. RAND Corporation. CT-463, Available: <https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/Te%20stimony-Kalra-DCCP-Hrg-Self-Driving-Cars-2017-02-14.pdf> (January 2021).
- Kalra, N., Paddock, S., 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A* 94, 182-193.
- Kang, G., Fang, S., 2011. Applying survival analysis approach to traffic incident duration prediction. In *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation*, 1523-1531.
- Kharoufeh, J., Goulias, K., 2002. Nonparametric identification of daily activity durations using kernel density estimators. *Transportation Research Part B* 36(1), 59-82.
- Kim, S.-G., Mannering, F., 1997. Panel data and activity duration models: Econometric alternatives and applications. In *Panels for Transportation Planning: Methods and Applications*. Golob, T., Kitamura, R., Long, L., eds., Kluwer, Norwell, MA, 349-373.
- Kitamura, R., Chen, C., Pendyala, R., 1997. Generation of synthetic daily activity-travel patterns. *Transportation Research Record* 1607, 154-162.

- Klugman, S., Panjer, H., Willmot, G., 2012. Loss models: from data to decisions, John Wiley and Sons, New York, NY.
- 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. IOP Publishing, 032101.
- Lin, T.-D., Jovanis, P., Yang, C.-Z. 1993. Modeling the safety of truck driver service hours using time-dependent logistic regression. *Transportation Research Record* 1407, 1-10.
- Liu, X. 2012. Survival analysis: models and applications, John Wiley and Sons, New York, NY.
- Loizos, A., Karlaftis, M., 2005. Prediction of pavement crack initiation from in-service pavements: A duration model approach. *Transportation Research Record* 1940, 38-42.
- Lord, D., Mannering, F. 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A* 44(5), 291-305.
- Lord, D., Qin, X., Geedipally, S., 2021. Highway Safety Analytics and Modeling, Elsevier Publish Ltd, Amsterdam, Netherlands.
- Lord, D., Washington, S., Ivan, J.. 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention* 37(1), 35-46.
- Mannering, F., 1993. Male/female driver characteristics and accident risk: some new evidence. *Accident Analysis and Prevention* 25(1), 77-84.
- Mannering, F., Bhat, C., 2014. Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research* 1, 1-22.

- Mannering, F., Bhat, C., Shankar, V., Abdel-Aty, M., 2020. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. *Analytic Methods in Accident Research* 25, 100113.
- Mannering, F., Hamed, M., 1990. Occurrence, frequency, and duration of commuters' work-to-home departure delay. *Transportation Research Part B* 24(2), 99-109.
- Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research* 11, 1-16.
- Mannering, F., Washburn, S., 2020. Principles of highway engineering and traffic analysis. Seventh Edition. John Wiley and Sons, New York, NY.
- Matysiak, A., Razin, P. 2018. 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.
- Nam, D., Mannering, F., 2000. An exploratory hazard-based analysis of highway incident duration. *Transportation Research Part A* 34(2), 85-102.
- Niemeier, D. and Morita, J., 1996. Duration of trip-making activities by men and women. *Transportation* 23(4), 353-371.
- Razali, N., Wah, Y., 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling Analytics* 2, 21-33.
- Reynolds, P., Hurley, S., Torres, M., Jackson, J., Boyd, P., Chen, V., 2000. Use of coping strategies and breast cancer survival: results from the Black/White Cancer Survival Study. *American Journal of Epidemiology* 152(1), 940-949.

- Royston, P., Parmar, M., 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 2016. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. SAE International Warrendale, PA. 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.
- Sohrabi, S., Khodadadi, A., Mousavi, S., Dadashova, B., Lord, D., 2021. Quantifying Autonomous Vehicle Safety: A Scoping Review of the Literature, Evaluation of Methods, and Directions for Future Research. *Accident Analysis and Prevention* 152, 105982.
- Stathopoulos, A., Karlaftis, M., 2002. Modeling duration of urban traffic congestion. *Journal of Transportation Engineering* 128(6), 587-590.
- Teoh, E., Kidd, D., 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* 10(2), 77-89.
- Van den Berg, P., Arentze, T., Timmermans, H., 2012. A latent class accelerated hazard model of social activity duration. *Transportation Research Part A* 46(1), 12-21.
- Vlahogianni, E., 2013. Modeling duration of overtaking in two lane highways. *Transportation Research Part F* 20, 135-146.

- Wang, J., 1996. Timing utility of daily activities and its impact on travel. *Transportation Research Part A* 30(3), 189-206.
- Wang, S., Li, Z., 2019. Exploring the mechanism of crashes with automated vehicles using statistical modeling approaches. *PloS One* 14(3), e0214550.
- Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., 2020. *Statistical and econometric methods for transportation data analysis*. Third Edition, CRC Press, Taylor and Francis Group, New York, NY.
- 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(6), 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(2), 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.

Table 1. Automated and conventional vehicles crash rates

| Descriptive Statistics | Conventional-vehicle crash data | Automated-vehicle crash data |
|--|--|---|
| 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 miles-to-crash | 21,634 | 27,399 |
| Minimum miles-to-crash | 12 | 4,212 |
| Maximum miles-to-crash | 112,975 | 134,023 |
| Median miles-to-crash | 12,679 | 15,767 |

Table 2. Exponential distribution duration model parameter estimates of the mileage driven between crashes in automated- and conventional-vehicle groups. (*t* statistics in parentheses).

| Variable Description | All Vehicles | All Vehicles with Vehicle-Type Indicator |
|--|---------------------|---|
| Constant | 3.207 (52.68) | 3.074 (38.77) |
| Automated vehicle indicator (1 if the vehicle is automated, 0 otherwise) | – | 0.275 (2.21) |
| Exponential parameter, λ | 0.040 (16.46) | 0.041 (16.28) |
| Number of observations | 235 | 235 |
| Log likelihood at convergence | –392.92 | –390.71 |

Table 3. Exponential distribution function hypothesis test

| Conventional Vehicles Crashes (130 observations) | | Automated Vehicles Crashes (105 observations) | |
|---|------------------|--|------------------|
| Test Statistic | | Test Statistic | |
| Kolmogorov-Smirnov | Anderson-Darling | Kolmogorov-Smirnov | Anderson-Darling |
| (Critical value) | (Critical value) | (Critical value) | (Critical value) |
| 0.12** (0.11) | 2.78** (1.32) | 0.14** (0.13) | 2.97** (1.32) |

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

Table 4. Automated and Conventional vehicles crash-free expectancy and safety effectiveness

| Mileage | Conventional Vehicle Crash-free Expectancy (miles) | Automated Vehicle Crash-free Expectancy (miles) | Difference in Miles | Difference in Percentage |
|----------------|---|--|--------------------------------|-------------------------------------|
| 10,000 | 3884 | 7,500 | 3616 | 93% |
| 25000 | 8027 | 11,724 | 3697 | 46% |
| 50,000 | 13,730 | 16,863 | 3133 | 23% |
| 100,000 | 20,257 | 24,813 | 4556 | 22% |
| 150,000 | 21,612 | 27,399 | 5787 | 27% |

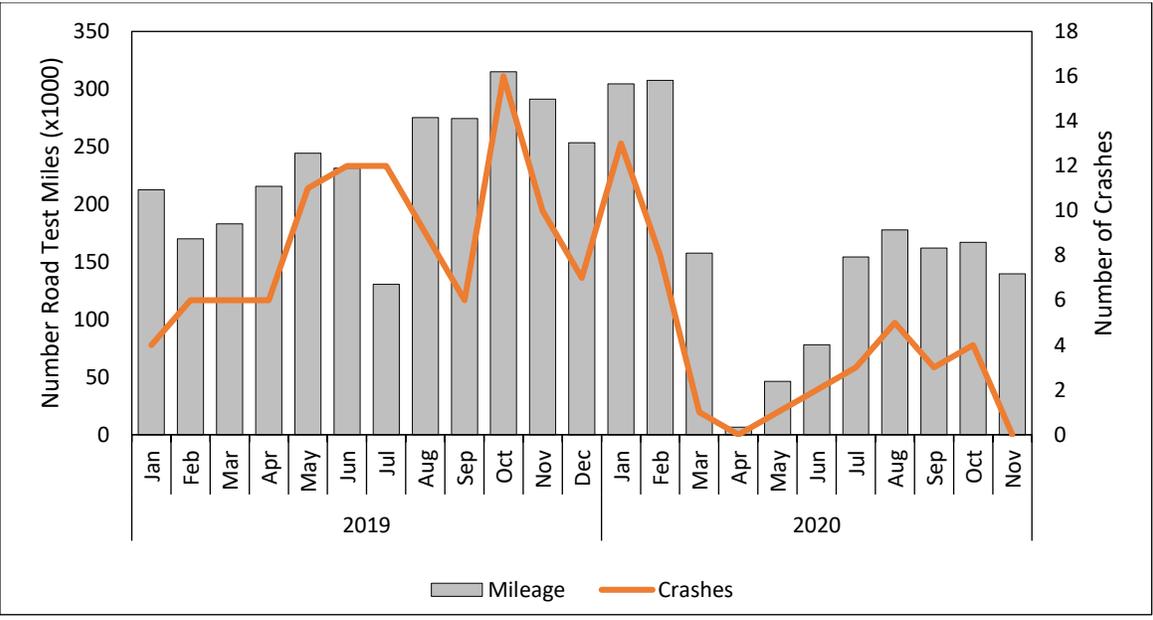


Figure 1. Distribution of automated vehicle crashes and vehicle miles travelled in 2019 and 2020.

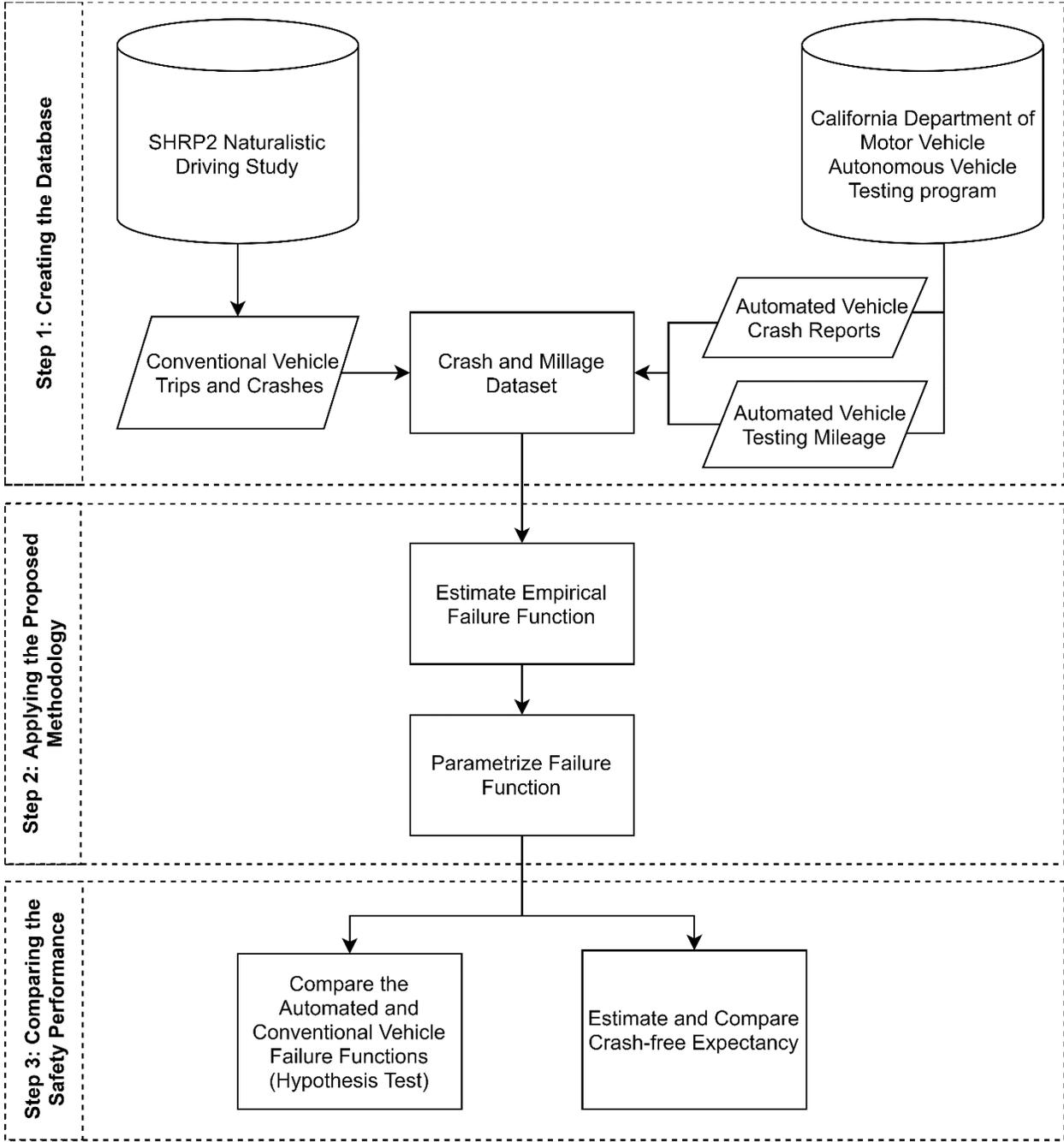


Figure 2. Automated Vehicle Safety Evaluation Framework

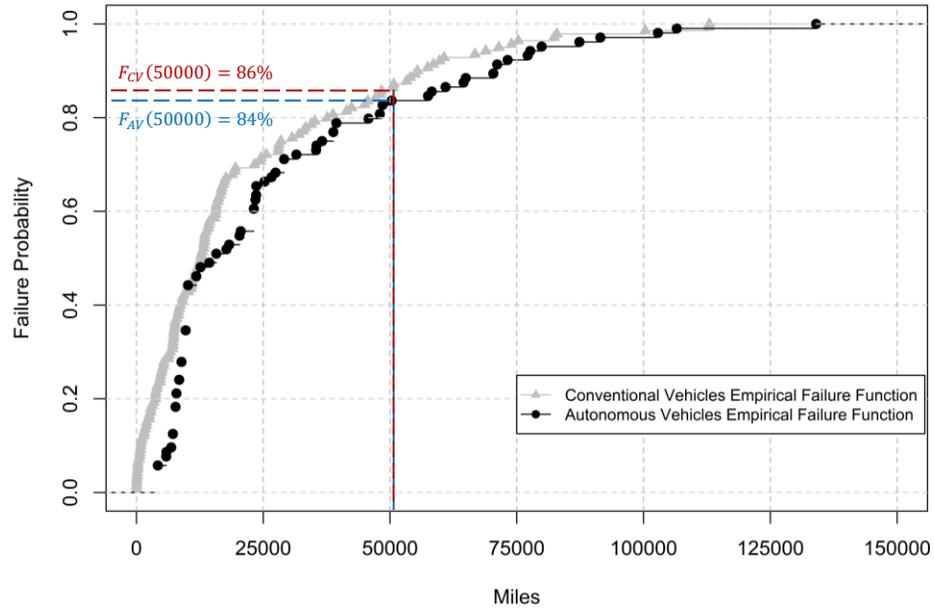


Figure 3. Estimated empirical failure function $F(x)$, for automated vehicles and conventional vehicles.