Normalizing Crash Risk of Partially Automated Vehicles under Sparse Data

Noah J. Goodall

Virginia Transportation Research Council, USA, noah.goodall@vdot.virginia.gov

This is an original manuscript of an article published by Taylor & Francis in *Journal of Transportation Safety & Security* on March 1, 2023, available online: http://dx.doi.org/10.1080/19439962.2023.2178566.

Abstract

The safety of increasingly automated vehicles is of great concern to regulators, yet crash rates are generally reported by manufacturers using proprietary metrics with limited source data. Without consistent definitions of crashes and exposure, automated vehicle crash rates cannot be meaningfully compared with baseline datasets. The objective of this study was to establish methods to normalize automated vehicle crash rates using one manufacturer's crash reports as a case study. The manufacturer's quarterly crash rates for vehicles using SAE Level 2 and Level 3 automation were compared. Road type was controlled for using data from a naturalistic driving study with the same model vehicles, while driver age was controlled for using demographic ownership surveys. Although Level 3 vehicles were claimed to have a 43% lower crash rate than Level 2 vehicles, their improved was only 10% after controlling for different rates of freeway driving. Direct comparison with general public driving was impossible due to unclear crash severity thresholds in the manufacturer's reports, but analysis showed that controlling for driver age would increase reported crash rates by 11%. These results establish the need for detailed crash data, crash definitions, and exposure and demographic data in order to accurately assess automated vehicle safety.

Disclaimer: *The views and opinions expressed in the article are those of the author and do not necessarily reflect the official policies or positions of any agency of the Commonwealth of Virginia.*

1 Introduction

The driving task is becoming increasingly automated, with computerized control not only of the throttle but also of brake and steering. Vehicles that can combine these functions are referred to as automated vehicles.

Establishing the safety of automated driving systems, both those conducted using on-road tests with trained operators as well as those being sold to the public, is of critical importance. California has established licensing programs for automated vehicle testing and requires companies to submit reports of all crashes as well as mileage both in automated and manual control (California Department of Motor Vehicles, 2018). Other states have various laws relating to automated vehicle operation, but few mandate companies report crashes or crash rates (National Conference of State Legislatures, 2019). The National Highway Traffic and Safety Administration (NHTSA) recently issued a standing general order requiring AV developers to report crashes within one day of occurrence (National Highway Traffic Safety Administration, 2021b). Before this order, companies testing outside California generally reported crashes and autonomous miles-traveled on a voluntary basis using metrics and thresholds of their own choosing (National Highway Traffic Safety Administration, 2021a).
When data is available, it can be difficult to compare self-reported statistics to baseline crash rates. Many automated vehicle crashes reported in California, for example, fall below the threshold for police-reportability and cannot be compared to national-level crash data which relies largely on police reports. An estimated 15% of injury crashes and 24% of property damage-only crashes are never reported to police (M. Davis and Company, Inc., 2015), while 9% of injury crashes and 24% of property damage-only crashes are reported but not logged (Blincoe et al., 2015). Even with robust data, establishing the statistical significance of automated vehicle safety can be expensive. Kalra and Paddock (2016) demonstrated that establishing that an AV has a fatal crash rate equivalent to the national average with 95% confidence would require driving a fleet of 100 vehicles continuously for 12.5 years. Considering that surveys suggest customers may require automated vehicles that crash four to five times less often than human-driven vehicles (Liu et al., 2019), the ability to directly compare the safety of different vehicle technologies quickly and accurately is essential to building public trust.

Tesla is a manufacturer of battery-powered electric vehicles. Beginning in 2014, new vehicles came installed with hardware allowing combined steering and throttle control. The automated driving system, named Autopilot, was officially introduced in 2015 and has undergone several updates via over-the-air software patches. Autopilot is considered an example of SAE Level 2 automation (SAE International, 2018) as it combines steering, brake, and throttle control, yet requires a human to monitor the system at all times and take control with little to no notice. In addition to Autopilot, many Tesla vehicles are equipped with active safety features such as automatic emergency braking, forward collision warning, blind spot warning, and lane departure assistance (Tesla, Inc., 2022). Active safety features remain active when Autopilot is not engaged, unless directly deactivated by the driver. As each of these features control a single aspect of vehicle control, they are generally classified as SAE Level 1 automation.

Beginning in 2018, Tesla began publishing short safety reports on a quarterly basis (Tesla, Inc., 2021). The reports list the average miles between observed crashes for vehicles using Autopilot, vehicles using active safety features only, and vehicles using neither Autopilot nor using safety features. Although the reports provide a critical insight into the crash rates of partially automated vehicles in use by the general public, they are difficult to interpret due to the lack of separate crash counts and baseline mileages, as well as unclear definitions of the criteria for incidents to be considered crashes. Given that Autopilot until recently was recommended for freeway use only, and given the demographics of their vehicle owners, crash rates of their vehicles may not allow a direct comparison to baseline crash rates without controlling for driver age and road usage.

The objective of this study is to establish a methodology for evaluating automated vehicle developers' safety reports in comparison to different samples, using the manufacturer's safety reports as a case study. This evaluation is based on Tesla crash reports from July 2018 through March 2021, and results may change as the automated driving system continues to be refined.[1] This is intended as a planning level analysis that will assist manufacturers, regulators, and researchers who may have access to more detailed, timely, and accurate data regarding vehicle safety and performance.

2 Literature review

Several studies have investigated automated vehicle crash rates. Schoettel and Sivak (2015) analyzed crash records of three companies approved to test automated vehicles in California and required to submit crash records to the California Department of Motor Vehicles. They found that although automated vehicles crashed at higher rates than the national average, the crash rates were within the 95% confidence interval in part due to small sample sizes.

Blanco et al. (2016) performed a more in depth study of Waymo (Google, at the time) automated vehicle crashes using Waymo's own internal records. To obtain national rates of minor crashes that did not meet police reporting thresholds, data from the Second Strategic Highway Research Program Naturalistic Driving Study (SHRP 2

[1] Tesla's most recent report as of this writing, for April through June of 2021, uses different safety metrics that previous reports (Kane, 2021). Previous reports provided crash rates driving with "Autopilot engaged," "without Autopilot but with our active safety features," and "without Autopilot and without our active safety features" (Tesla, Inc., 2021) The latest report uses different terminology, distinguishing between driving "in which drivers were using Autopilot technology (Autosteer and active safety features)" and "drivers who were not using Autopilot technology (no Autosteer and active safety features)" (Tesla, Inc., 2021) It is not clear from their wording if how the prior category of active safety features only maps to the new categories. Without further guidance, these new crash rates cannot be compared to previously reported crash rates.

NDS) (Antin et al., 2019) was used for comparison. When comparing crash rates of similar severities, the researchers found that Waymo automated vehicles crashed at lower rates than national estimates.

Dixit et al. (2016) compared Waymo's automated vehicle crash rate as reported to the California Department of Motor Vehicles from September 2014 to November 2015 with California Highway Patrol's (CHP) estimated statewide crash rates. The study was limited due to CHP's coverage of State, U.S., and Interstate roads when Waymo testing at the time occurred primarily on local roads outside CHP's jurisdiction.

Teoh and Kidd (2017) analyzed Waymo automated vehicle crashes in both the California Department of Motor Vehicle reports as well as testing performed outside California and discussed in Waymo's monthly activity reports. Waymo crashes were limited to those that met the criteria for police reporting, but were compared to general public crashes that were actually reported to the police. This created some challenges as a significant number of general public crashes that meet the criteria for a police report are either never reported, or are reported but not logged into police databases (Blincoe et al., 2015; M. Davis and Company, Inc., 2015). This was evident in several Waymo crashes where police were called but did not respond. Waymo crash rates were also compared to the SHRP 2 NDS dataset, but the naturalistic data was biased towards higher risk drivers at both ends of the age spectrum and did not appear to be age-weighted in the analysis. The authors found that SHRP 2 NDS crash rates were higher than Waymo's at a 95% confidence interval.

Goodall (2021) investigated struck-from-behind crashes of automated vehicles using age-weighted crash rates from SHRP 2 NDS database as a baseline. Automated vehicles were struck from behind at five times the rate of human-driven vehicles, although much of the difference could be attributed to higher rates of urban driving experienced in automated vehicle testing.

Other studies have used automated vehicle crash records to model crash severity (Wang & Li, 2019), analyze crash narratives using text mining (Alambeigi et al., 2020; Boggs et al., 2020), and perform exploratory analysis (Das et al., 2020; Leilabadi & Schmidt, 2019; McCarthy, 2021).

In independent research, Templeton (2020) compared Tesla's stated crash rates with Autopilot enabled and not enabled by attempting to control for increased use of Autopilot on relatively safer freeways. To compare human-driven crash rates of freeways and non-freeways, Templeton used fatality rates, which may overestimate crash rates on freeways as higher speeds increase crash severity according to a fourth power law (Evans, 1994). When controlling for road type, the crash rate benefits of Autopilot narrowed substantially. Templeton, however, normalized based on national fatal crash rates rather than Tesla's metric of total crash rates.

A review of the literature establishes a clear need for methods to isolate environmental and demographic factors when evaluating automated vehicle safety using raw crash statistics.

3 Materials and methods

Data sources and normalization procedures are discussed in the following sections. Table 1 provides a summary of data and sources used in this study.

Table 1. Summary of normalization data and sources.

Data	Tesla	National
Unadjusted Crash Rates	Tesla Safety Reports [a]	SHRP 2 NDS [c]
Road Use Ratio, Freeway vs. Non-Freeway	MIT AVT [b]	SHRP 2 NDS [c]
Crash Rates adjusted for Road Type	Calculated in this study	SHRP 2 NDS [c]

[a] (Tesla, Inc., 2021)

[b] Massachusetts Institute of Technology Advanced Vehicle Technology Tesla naturalistic driving study (Fridman et al., 2018), with ratios supported by studies of other SAE Level 1 and 2 systems (Gershon et al., 2021; Reagan et al., 2019)

[c] Strategic Highway Research Program 2 Naturalistic Driving Study (Antin et al., 2019)

-

3.1 Manufacturer data

Tesla's stated crash rates for vehicles with Autopilot (AP) and active safety features engaged were compared to Tesla's stated crash rates with active safety features-only (ASO) engaged. Manufacturer's safety rates were obtained from their self-published quarterly safety reports (Tesla, Inc., 2021) which list crash rates in terms of miles per crash but do not specify total mileages nor total crash count as separate values. Tesla defines a crash as an incident where "the crash alert indicated an airbag or other active restraint" was deployed (Tesla, Inc., 2021). According to the report, this "correlates to nearly any crash at about 12 mph (20 kph) or above, depending on the crash forces generated" (Tesla, Inc., 2021). Crash rates when no active safety features were in use were also reported, but these are excluded from this analysis as active safety features are engaged as the default. The use of ASO is considered as operating under manual control when comparing with other datasets.

The ratio of Tesla miles traveled with Autopilot and active safety only on freeways was obtained from an unpublished 2018 study (Fridman et al., 2018). The report discussed a naturalistic driving study of 28 vehicles with automated driving features, of which 25 were Tesla models equipped with Autopilot. Vehicles had traveled a total of 323,384 miles at the time the report was written. The study focused primarily on driver attentiveness, but a breakdown of distances traveled on roads with various speed limits were shown in Figures 3 and 4 in that report (Fridman et al., 2018). Distance by speed limit from that report is reproduced here in **Error! Reference source not found.**.

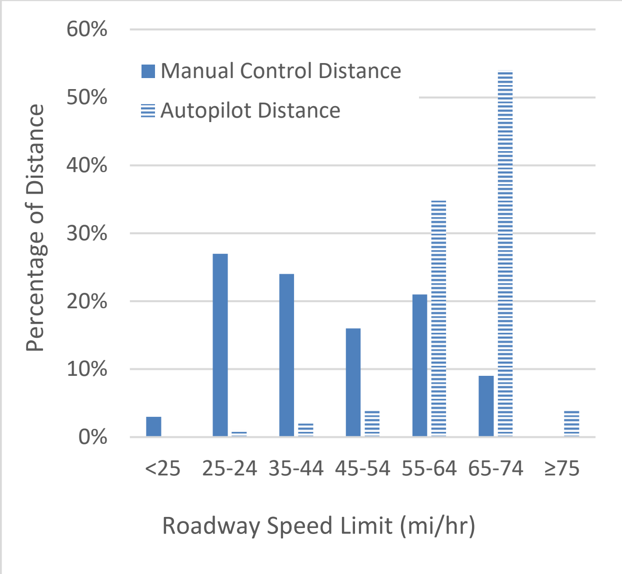

Figure 1. Distance traveled on different roads by speed limit (Fridman et al., 2018).

An author has confirmed in personal correspondence that all mileages in the study are accurate (Reimer, 2021). Data from the same naturalistic driving study has been used in several other papers (Ding et al., 2020; Morando et al., 2020; Reagan et al., 2019). Furthermore, the ratios of freeway-to-non-freeway mileages are similar to those found in other studies of ACC and SAE Level 2 automation in different model vehicles (Morando et al., 2020; Reagan et al., 2019).

3.2 National data

National crash rates and adjustment factors were obtained from the SHRP 2 NDS. Between 2012 and 2014, video, audio, and kinematic data were collected from 3,500 drivers using their personal vehicles (Antin et al., 2019). Although data was continuously recorded and later deleted, it was only saved permanently when either triggered by a change of acceleration threshold or other crash indicator, or at random intervals to collect baseline driving data. Events were analyzed and categorized by a team of trained analysts.

SHRP 2 NDS used several tiers of crash severity, of which three appear to correlate to the manufacturer's crash definition. The first is any crash that results in the activation of an airbag. The second is referred to as Crash Severity I – Most Severe ("Severe" for short), which includes any airbag crashes as well as any crashes that result in physical injury, roll over, or a change in speed of the instrumented vehicle of greater than 20 mph (Virginia Tech Transportation Institute, 2015). Vehicle safety research refers to this change of speed as delta-v. The third is referred to as Crash Severity II – Police Reportable, and includes any crash that does not meet the definition of Severe but either causes an estimated minimum of $1500 in damages or reaches acceleration on any axis of 1.3g (Virginia Tech Transportation Institute, 2015). The combined severe/police reportable (Crash Severities I and II) metric was selected for use in the analysis as part of this effort, as it appears closest the manufacturer's crash definition. Airbag crashes are included in severe crashes.

The SHRP 2 NDS also records baseline mileage data and crash events by age of the driver and roadway type. Roads are classified based on observation from video rather than speed. Events classified as "Interstate/bypass/divided highway with no traffic signals" (Virginia Tech Transportation Institute, 2015) are considered freeways, while all other events are considered non-freeway.

Each entry in the database includes two separate events, as many crashes have an initial event (e.g., a rear-end collision) followed by a second event (e.g., a curb strike). SHRP 2 NDS data was queried for Crash Severity 1 = I or II OR Crash Severity 2 = I or II. Records were then filtered by locality and airbag deployment.

3.2.1 Age-weighting

The sample of drivers in the SHRP 2 NDS dataset was biased towards younger and older drivers as they represent higher risk groups. Mileages and crash counts for all age groups were reweighted based on ages of United States licensed drivers in Blanco et al. (2016). Their calculations are reproduced in Table 2. Age-weighted SHRP 2 NDS data are used throughout the study.

Table 2. Age group sample weights for SHRP 2 Naturalistic Driving Study data (Blanco et al., 2016).

Age	Weight	Percentage in SHRP 2 NDS	Percentage of US Licensed Drivers	Million miles driven	Weighted million miles driven
16-24	0.32	37	12	12.9	4.1
25-39	1.53	17	26	6.4	9.8
40-54	2.33	12	28	4.6	10.7
55-74	1.35	20	27	6.3	8.5
75+	0.5	14	7	3.4	1.7
Totals	-	100	100	33.6	34.8

The analysis in this paper relies on the assumption that the freeway-to-non-freeway and age group crash ratios found in the SHRP 2 NDS are consistent with the manufacturer's data, as there are no roadway-specific nor age-related factors in the manufacturer safety reports. Crash rates by severity are shown in Table 3. Although the ages of drivers in Tesla's crash rate data are unknown, their ages could be estimated from a 2018 demographic survey of 424 Tesla owners (Hardman et al., 2019).

Table 3. SHRP 2 NDS age-weighted crash rates by severity.

	Crashes per 100 million miles		
Severity	Freeway	Non-Freeway	Combined
Airbag deployment	29	26	27
Severe, incl. airbags	160	293	257
Police-reportable, not severe	226	482	412
Severe and Police-reportable	386	776	668
Minor	569	2455	1934
All	955	3231	2602

3.3 Adjustment methods

Crash rates can be adjusted to account for differences in environment and demographics in different data sets. A sample dataset with a crash rate *r* is exposed to some variable *i* at a different proportion *p* than the comparison dataset. In the case study, for example, vehicles running Autopilot were driven on freeways (i) 93% of the time, resulting in p_i = 0.93. In the SHRP 2 NDS, only 28% of vehicle mileage was recorded on freeways, i.e $p_i = 0.28$. In the SHRP 2 NDS data, vehicles on non-freeways crashed 2.01 times more often per mile than vehicles on freeways. The observed Autopilot crash rate can be adjusted to reflect national driving rates to reflect the crash rate that might be observed if 28% of Autopilot mileage was on freeways and 72% were on non-freeways, assuming that the 2.01 ratio holds for Autopilot.

The first step is to find the odds ratios for freeway and non-freeway driving in the SHRP 2 NDS data. The odds ratio is defined as the crash rate under the risk factor divided by the crash rate for all driving:

$$OR = \frac{r_i}{r_t} \tag{1}$$

where r_i is the crash rate for either freeway or non-freeway miles, and r_t is the crash rate per mile of all driving.

The crash rate r_i can be obtained by solving a system of equations. First, the sum of the crash rate r for each risk factor i multiplied by the proportion of exposure spent in that risk factor p_i is equivalent to the total crash rate r_t.

$$\sum_{i=0}^{n} r_i p_i = r_t p_t \tag{2}$$

In the second equation, the odds ratio formula is rewritten as:

$$r_t = \frac{r_i}{OR_i} \tag{3}$$

Given that r_t remains constant for a given sample, this yields the following equation:

$$r_t = \frac{r_o}{OR_o} = \frac{r_1}{OR_1} = \cdots = \frac{r_{n-1}}{OR_{n-1}} = \frac{r_n}{OR_n} \tag{4}$$

The value of any crash rate r_i can be expressed in terms of any other crash rate r_j as:

$$r_i = r_j \frac{OR_i}{OR_j} \tag{5}$$

Inserting this equation into the crash rate summation equation yields:

$$r_o p_0 + r_1 p_1 + \cdots + r_{n-1} p_1 + r_n p_n = r_t p_t \tag{6}$$

$$r_o p_o + r_0 \left(\frac{OR_1}{OR_0}\right) + r_0 \left(\frac{OR_2}{OR_0}\right) + \cdots + r_0 \left(\frac{OR_{n-1}}{OR_0}\right) + r_0 \left(\frac{OR_n}{OR_0}\right) = r_t p_t \tag{7}$$

$$r_i \left(\sum_{i=0}^{n} \frac{OR_j}{OR_i} p_j \right) = r_t p_t \tag{8}$$

$$r_i = \frac{r_t p_t}{\sum_{i=0}^{n} \left(\frac{OR_j}{OR_i} p_j \right)} \tag{9}$$

When p_i is expressed as a percentage of the total, then $p_t = 1$. In applying the methodology, the odds ratio should be calculated from the more detailed database. For this case study, odds ratios were calculated from the SHRP 2 NDS and the Autopilot NDS.

Once individual crash rates for each risk factor are calculated, they can be combined into an updated total crash rate using the formula:

$$\sum_{i=0}^{n} r_i p_{i,adj.} = r_{t,adj} p_t \tag{10}$$

but replacing p_i with the comparison ratios, e.g. 28% freeway miles and 72% non-freeway miles.

4 Results

Crash rates using Autopilot and active safety features only were compared. In the manufacturer's reports, Autopilot had higher average distances between crashes, ranging between 1.46 and 2.35 times greater average miles between crashes than active safety only (Tesla, Inc., 2021). Autopilot is used predominately on freeways, as surface street functionality was only released recently to select users and remains in beta. A naturalistic driving study has shown that 93% of Autopilot usage is on roads with speed limits of 55 mi/hr or greater. In contrast, of the miles traveled using active safety only, only 30% occurred on roads with speed limits only greater than 55 mi/hr (Fridman et al., 2018).

Because the SHRP 2 NDS dataset does not include speed limits but rather road classifications, roads with speed limits greater than 55 mi/hr were classified as freeways to allow for direct comparison with SHRP 2 NDS. Although this classification is not ideal, as some non-freeways have posted speed limits of 55 mi/hr, other studies have investigated usage of SAE Levels 1 and 2 technologies on different road types and found similar ratios. Drivers of Cadillac CT6 vehicles used Super Cruise SAE Level 2 automation mostly on freeways, with 99% of all Super Cruise usage occurring on freeways and 1% on non-freeways (Gershon et al., 2021). Similarly, combined usage of adaptive cruise control and Super Cruise was also predominantly freeway, at 93% freeway usage and 7% non-freeway usage.[2] Another naturalistic driving study of the Range Rover Evoque and Volvo S90 found that ACC (SAE Level 1) and Pilot Assist (SAE Level 2) saw between 81% and 93% of their usage on freeways (Reagan et al., 2019). These ratios are similar to MIT's Tesla Autopilot usage rates of 93% freeway and 7% non-freeway.

Combined severe and police-reportable crashes were used to calibrate the manufacturer data. From the SHRP 2 NDS, the observed ratio of freeway to non-freeway crashes was 2.01. Assuming that this ratio holds for the manufacturer data, then the crash rates can be adjusted using the methods in the previous section. Variables for Q1 2021 are provided in Table 4. The crash rate after adjusting for freeway and non-freeway use is shown in Table 5 and Figures 2 and 3. Although Autopilot was claimed to have a 43% lower crash rate than active safety only, this improvement reduces to 10% (average of all quarterly reports) after controlling for different rates of freeway driving.

Table 4. Case study crash adjustment variables.

Variable	Freeway	Non-Freeway
OR	0.58	1.17
p, Autopilot	0.93	0.07
p, Active Safety Only	0.30	0.70
p, SHRP 2 NDS	0.28	0.72

---

[2] These figures were reconstructed from (Gershon et al., 2021) as follows. There were 22,108 total miles, of which 62% (13,707 miles) were on freeways ("limited access highways"). Of freeway miles, 40% (5,483 miles) used Super Cruise, 10% (1,371 miles) used adaptive cruise control, and 50% (6,853 miles) were under manual control. The non-freeways miles were calculated by subtracting freeway miles from total miles, with total mileage reported in the paper as 5,514 miles using Super Cruise, 1,891 miles using adaptive cruise control, and 14,702 miles under manual control.

Table 5. Autopilot and active safety only crash rates adjusted for road usage.

	Autopilot Crashes per 100 million miles		Active Safety Only Crashes per 100 million miles	
Date	Unadjusted	Adjusted	Unadjusted	Adjusted
Q3 2018	29.9	48.4	52.1	52.8
Q4 2018	34.4	55.5	63.3	64.2
Q1 2019	34.8	56.3	56.8	57.6
Q2 2019	30.6	49.4	45.7	46.3
Q3 2019	23.0	37.2	37.0	37.6
Q4 2019	32.6	52.6	47.6	48.3
Q1 2020	21.4	34.5	50.3	51.0
Q2 2020	22.1	35.7	44.1	44.7
Q3 2020	21.8	35.2	41.3	41.9
Q4 2020	29.0	46.8	48.8	49.5
Q1 2021	23.9	38.6	48.8	49.5

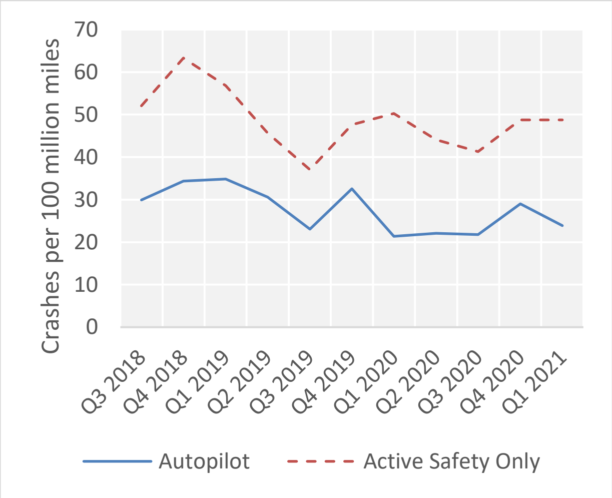

Figure 2. Unadjusted Tesla crashes per 100 million miles.

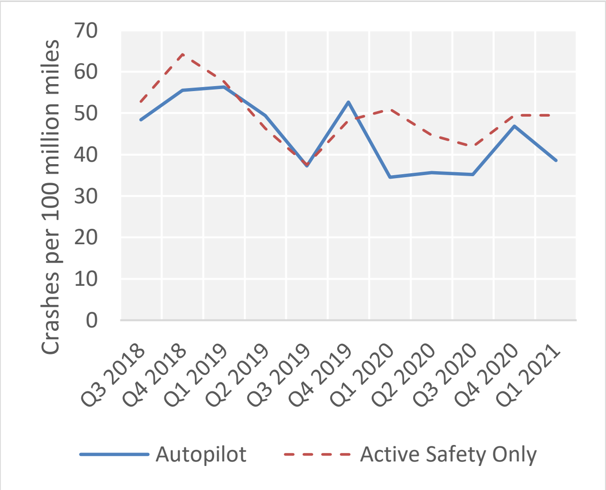

Figure 3. Tesla crashes per 100 million miles, adjusted for road classification.

The same procedure was used to adjust for driver demographics. Tesla owners are concentrated in the 50 to 70-year-old age brackets and underrepresented among drivers ages 16–39 and 75 and older (Hardman et al., 2019). **Error! Reference source not found.** lists the percentage of drivers in each bin for both samples. The results compare the crashes per 100 million miles for Autopilot and active safety in unadjusted numbers, adjusted for road type, and adjusted for both road type and owner age. Crash rates adjusted for road classification and owner age are shown in **Error! Reference source not found.**. Averaging over each quarter, controlling for driver age increased the reported crash rates by approximately 11%.

Table 6. Comparison of driver demographics.

Age	Tesla Owners[1]	Percentage of US Licensed Drivers[2]	Weighted Most Severe and Police Reportable Crashes per 100 million miles[3]	Odds Ratio
16-24	0.7%	12%	1413	2.11
25-39	12.1%	26%	531	0.79
40-54	29.3%	28%	501	0.75
55-74	47.5%	27%	604	0.90
75+	10.3%	7%	1059	1.58
Totals	100.0%	100.0%	669	1.00

[1] Statistics obtained from Hardman et al. (2019).
[2] Statistics obtained from Blanco et al. (2016).
[3] Statistics obtained from Transportation Research Board of the National Academy of Sciences (2013).

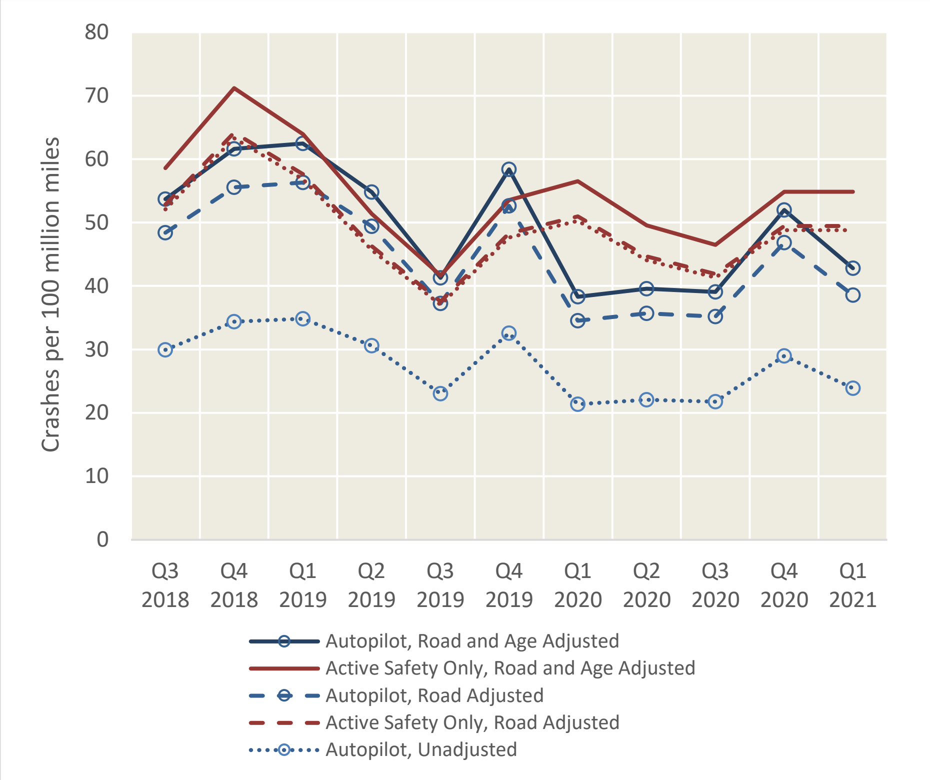

Figure 4. Manufacturer crash rates adjusted for road use and owner age.

5 Discussion

By correcting for roadway usage differences between the Autopilot and active safety only data, much of the crash reduction seen by vehicles using Autopilot appears to be explained by lower crash rates experienced on freeways. While the raw crash rate shows an average 43% reduction in crash rate for Autopilot compared to active safety only, this improvement is only 10% after controlling for different rates of freeway driving. Correcting for age demographics likewise produced an 11% increase in the estimated crash rate.

Several other factors may explain difference in safety rates of new vehicle technologies based on who is using them, where they are being used, and when they are being used. Some safety features cannot be used in rain or snow, for example, which may bias the data towards clear weather and lower crash rates generally. In another example, drivers often disengage Autopilot to change lanes or prepare to exit a freeway (Morando et al., 2020), both areas of increased crash risk. While Tesla includes crashes where Autopilot was deactivated less than five seconds prior to impact, there remains a potential for Autopilot use to be biased towards safer, within-lane driving. With more data regarding not only the crashes but also the use of vehicle technologies can allow for more accurate and thorough assessments of vehicle safety benefits.

5.1 Limitations

Comparing limited data across inconsistent samples as done in this study requires several assumptions that may limit the accuracy of the results. There are five main assumptions that may affect the results.

First, it is assumed that ratios between freeway and non-freeway driving found in the SHRP 2 NDS are maintained in the manufacturer data set. If drivers crash twice as often on non-freeways as freeways, it is assumed that drivers in the manufacturer's data will also crash twice as often, although the absolute rate may be different.

Second, the MIT data is assumed to be accurate and applicable. The data was collected in early 2018 before manufacturer safety reports were published, so it is possible that driver travel patterns and usage of Autopilot has changed significantly in the interim, thereby affecting the results. Much of the naturalistic data was collected in and around Boston, Massachusetts. This data may not be applicable to the United States generally.

Third, it was assumed that vehicles on the same type of road were exposed to the same crash risk. Not only type of road but also conditions on the road may differ for Tesla drivers resulting in different crash rates. Congested traffic, for example, might result in more crashes per mile traveled. If Tesla vehicles were more likely to be driven along urban corridors during peak periods, then they would likely have encountered more congestion. These data, while not available, could have a significant impact on observed crash rates.

Fourth, there is some uncertainty over what classifies a road as freeway. In the SHRP 2 NDS, camera images were evaluated to determine road classification. In the MIT data, an assumption was made that roads with speed limits of 55 mi/hr or greater were freeways. There are, of course, roads with 55 mi/hr speed limits that are not freeways, although it is unclear to what extent these roads are prevalent in the Boston area where much of the naturalistic data was collected. A similar naturalistic study of drivers using Cadillac CT6 vehicles with Super Cruise SAE Level 2 automation in the Boston area reported usage by road type and found similar rates of freeway vs. non-freeway driving (Gershon et al., 2021). In the study, researchers found that 99% of Super Cruise mileage was on freeways, and 93% of combined Super Cruise and adaptive cruise control mileage was on freeways. This suggests that the MIT data usage figures are accurate, although these ratios should be updated as more data becomes available.

Fifth, it was assumed that the SHRP 2 NDS crash severities of severe and police-reportable, when combined, captured the same types of crashes as those recorded by Tesla. While these definitions appear most similar of the choices available, it is unlikely they are a perfect match as SHRP 2 NDS data relied on some human judgement and the manufacturer crash rate definitions were dependent on active safety features with unclear thresholds. Because of uncertainty of crash definitions, it was impossible to directly compare manufacturer crash rates with national estimates. The use of SHRP 2 NDS crash rates to calibrate freeway vs. non-freeway driving, however, can be updated with clearer crash definitions.

5.2 Best practices

When comparing crash rates among different vehicles with different driver demographic and usage patterns, some best practices may help facilitate evaluations. Anonymized driver or owner demographics may be used to isolate safety impacts, as highly educated middle-aged drivers are both more likely to purchase luxury cars and simultaneously crash at lower rates than the overall population. Similarly, specific details regarding built-in restrictions on when vehicle technologies can and cannot be used allows researchers to better design baseline metrics that match the technology's operational design domain. Manufacturers should record additional anonymized data on vehicle driving environments, rates of usage of automation modes, and details on when and where the modes are used. Selective sampling of short segments of vehicle trips could be used to develop statistics on baseline behavior to compare against crash records.

Finally, consistent crash definitions and thresholds are becoming increasingly necessary in order to evaluate new safety technologies. The standard metric of police-reportable has different meanings in different jurisdictions—this creates confusion when crashes that meet the threshold for reporting are either not reported or reported but not logged into police databases. More consistent definitions could combine such metrics as maximum jerk, maximum acceleration, changes in subject vehicle relative speed. When metrics such as airbag deployment or vehicle crash sensing are employed, measurements of changes in vehicle velocity over time can help to further refine crash classification and provide insight into a crash's severity.

6 Declaration of interests statement

Declaration of interest: None.

7 Data availability statement

Data used in this analysis are hosted at the Open Science Framework repository and available for download at https://doi.org/10.17605/OSF.IO/5HKMD.

8 References

Alambeigi, H., McDonald, A. D., & Tankasala, S. R. (2020). Crash themes in automated vehicles: a topic modeling analysis of the California Department of Motor Vehicles automated vehicle crash database. *ArXiv:2001.11087 [Stat]*. http://arxiv.org/abs/2001.11087

Antin, J. F., Lee, S., Perez, M. A., Dingus, T. A., Hankey, J. M., & Brach, A. (2019). Second strategic highway research program naturalistic driving study methods. *Safety Science*, *119*, 2–10. https://doi.org/10.1016/j.ssci.2019.01.016

Blanco, M., Atwood, J., Russell, S., Trimble, T., McClafferty, J., & Perez, M. (2016). *Automated Vehicle Crash Rate Comparison Using Naturalistic Data*. Virginia Tech Transportation Institute.

Blincoe, L. J., Miller, T. R., Zalashnja, E., & Lawrence, B. A. (2015). *The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised)* (DOT HS 812 013). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812013

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 & Prevention*, *135*, 105354. https://doi.org/10.1016/j.aap.2019.105354

California Department of Motor Vehicles. (2018, December 19). *Title 13, Division 1, Chapter 1, Article 3.7 – Testing of Autonomous Vehicles*. https://www.dmv.ca.gov/portal/uploads/2020/06/Adopted-Regulatory-Text-2019.pdf

Das, S., Dutta, A., & Tsapakis, I. (2020). Automated vehicle collisions in California: Applying Bayesian latent class model. *IATSS Research*. https://doi.org/10.1016/j.iatssr.2020.03.001

Ding, L., Glazer, M., Wang, M., Mehler, B., Reimer, B., & Fridman, L. (2020). MIT-AVT clustered driving scene dataset: evaluating perception systems in real-world naturalistic driving scenarios. *2020 IEEE Intelligent Vehicles Symposium (IV)*, 232–237. https://doi.org/10.1109/IV47402.2020.9304677

Dixit, V. V., Chand, S., & Nair, D. J. (2016). Autonomous vehicles: disengagements, accidents and reaction times. *PLOS ONE*, *11*(12), e0168054. https://doi.org/10.1371/journal.pone.0168054

Evans, L. (1994). Driver injury and fatality risk in two-car crashes versus mass ratio inferred using Newtonian mechanics. *Accident; Analysis and Prevention*, *26*(5), 609–616.

Fridman, L., Brown, D. E., Kindelsberger, J., Angell, L., Mehler, B., & Reimer, B. (2018). *Human Side of Tesla Autopilot: Exploration of Functional Vigilance in Real-World Human-Machine Collaboration*. https://www.coursehero.com/file/68234521/Case-Study-1pdf/

Gershon, P., Seaman, S., Mehler, B., Reimer, B., & Coughlin, J. (2021). Driver behavior and the use of automation in real-world driving. *Accident Analysis & Prevention*, *158*, 106217. https://doi.org/10.1016/j.aap.2021.106217

Goodall, N. J. (2021). Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data. *Accident Analysis & Prevention*, *154*, 106056. https://doi.org/10.1016/j.aap.2021.106056

Hardman, S., Lee, J. H., & Tal, G. (2019). How do drivers use automation? Insights from a survey of partially automated vehicle owners in the United States. *Transportation Research Part A: Policy and Practice*, *129*, 246–256. https://doi.org/10.1016/j.tra.2019.08.008

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. https://doi.org/10.1016/j.tra.2016.09.010

Kane, M. (2021, October 1). *Tesla Vehicle Safety Report Q2 2021: Numbers Are Down Again*. InsideEVs. https://insideevs.com/news/537818/tesla-autopilot-safety-report-2021q2/

Leilabadi, S. H., & Schmidt, S. (2019). In-depth Analysis of Autonomous Vehicle Collisions in California. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 889–893. https://doi.org/10.1109/ITSC.2019.8916775

Liu, P., Yang, R., & Xu, Z. (2019). How safe is safe enough for self-driving vehicles? *Risk Analysis*, *39*(2), 315–325. https://doi.org/10.1111/risa.13116

M. Davis and Company, Inc. (2015). *National Telephone Survey of Reported and Unreported Motor Vehicle Crashes* (DOT HS 812 183). National Highway Traffic Safety Administration. http://www-nrd.nhtsa.dot.gov/Pubs/812183.pdf

McCarthy, R. L. (2021). Autonomous vehicle (AV) accident data analysis: California OL 316 reports: 2015-2020. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*. https://doi.org/10.1115/1.4051779

Morando, A., Gershon, P., Mehler, B., & Reimer, B. (2020). Driver-initiated Tesla Autopilot disengagements in naturalistic driving. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 57–65. https://doi.org/10.1145/3409120.3410644

National Conference of State Legislatures. (2019, March 19). *Autonomous Vehicles | Self-Driving Vehicles Enacted Legislation*. National Conference of State Legislatures. http://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx

National Highway Traffic Safety Administration. (2021a, May 31). *Voluntary Safety Self-Assessment* [Text]. https://www.nhtsa.gov/automated-driving-systems/voluntary-safety-self-assessment

National Highway Traffic Safety Administration. (2021b, June 29). *Standing General Order 2021-01: Incident Reporting for Automated Driving Systems and Level 2 Advanced Driver Assistance Systems* [Text]. https://www.nhtsa.gov/sites/nhtsa.gov/files/2021-06/Standing_General_Order_2021_01-digital-06292021.pdf

Reagan, I. J., Hu, W., Cicchino, J. B., Seppelt, B., Fridman, L., & Glazer, M. (2019). Measuring adult drivers' use of Level 1 and 2 driving automation by roadway functional class. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2093–2097. https://doi.org/10.1177/1071181319631225

Reimer, B. (2021). *Personal communication*.

SAE International. (2018). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* (SAE J3016_201806).

Schoettle, B., & Sivak, M. (2015). *A Preliminary Analysis of Real-World Crashes Involving Self-Driving Vehicles* (UMTRI-2015-34). University of Michigan

Transportation Research Institute. http://www.umich.edu/~umtriswt/PDF/UMTRI-2015-34_Abstract_English.pdf

Templeton, B. (2020, October 28). *New Tesla Autopilot Statistics Show It's Almost As Safe Driving With It As Without*. Forbes. https://www.forbes.com/sites/bradtempleton/2020/10/28/new-tesla-autopilot-statistics-show-its-almost-as-safe-driving-with-it-as-without/

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. https://doi.org/10.1016/j.jsr.2017.08.008

Tesla, Inc. (2021). *Tesla Vehicle Safety Report*. https://www.tesla.com/VehicleSafetyReport

Tesla, Inc. (2022). *Safety*. Tesla. https://www.tesla.com/safety

Transportation Research Board of the National Academy of Sciences. (2013). *The 2nd Strategic Highway Research Program Naturalistic Driving Study Dataset*. https://insight.shrp2nds.us/

Virginia Tech Transportation Institute. (2015). *SHRP2 Researcher Dictionary for Video Reduction Data Version 3.4*. Virginia Tech Transportation Institute. https://drive.google.com/file/d/0B0WDAFahhsCGX05hM3U4OVVjZUk/view?usp=sharing&usp=embed_facebook

Wang, S., & Li, Z. (2019). Exploring the mechanism of crashes with automated vehicles using statistical modeling approaches. *PLoS ONE*, *14*(3). https://doi.org/10.1371/journal.pone.0214550