Do Autonomous Vehicles Outperform Latest-Generation Human-Driven Vehicles? A Comparison to Waymo's Auto Liability Insurance Claims at 25 Million Miles

Luigi Di Lillo [a,b,c] Luigi_DiLillo@swissre.com · Tilia Gode [d,e] ·
Xilin Zhou [a] · John Scanlon [d] · Ruoshu Chen [d,e] · Trent Victor [d]

a Swiss Reinsurance Company, Ltd, Switzerland
b Zardini Lab (research affiliate), Massachusetts Institute of Technology, USA
c Autonomous Systems Laboratory (research affiliate), Stanford University, USA
d Waymo LLC, USA
e Casualty Actuarial Society, USA

Abstract

Understanding the safety impact of Automated Driving Systems (ADS) is crucial for their widespread adoption, yet robust real-world evaluation remains a critical area of development. By leveraging insurance industry third-party auto liability claims, we analyze the Waymo ADS performance across 25.3 million fully autonomous miles (with no human in the driver seat) and demonstrate a consistent and scalable framework for ongoing ADS safety assessment. This study not only updates previous findings with expanded geographical coverage and increased mileage but also establishes a methodology for comparing ADS performance to evolving subsets of human-driven vehicles (HDV). We introduce two key innovations: a novel "latest-generation HDV" benchmark representing drivers of newer, technologically advanced vehicles (2018-2021), and a robust insurance dataset for ADS safety assessment. Results demonstrate that the Waymo ADS significantly outperformed both the *overall driving population* (88% reduction in property damage claims, 92% in bodily injury claims), and outperformed the more stringent *latest-generation HDV* benchmark (86% reduction in property damage claims and 90% in bodily injury claims). This substantial safety improvement[1] over our previous 3.8-million-mile study not only validates ADS safety at scale but also provides a new approach for ongoing ADS evaluation. Our findings provide crucial insights into the evolving safety landscape of ADS technology and have far-reaching implications for transportation safety policies, insurance risk assessments, and public acceptance of ADS technology. This methodology establishes a foundation for future research into the safety impact potential of ADS and offers a framework for assessment as these systems continue to scale and develop.

Key findings

- Waymo's low frequency of auto liability insurance claims, despite increased mileage and smaller confidence intervals, underscores its consistent safety performance as it scales across diverse operational domains.
- The introduction of the *latest-generation HDV* benchmark offers a more forward-looking and challenging comparison for ADS performance evaluation.

[1] The research partnership between Swiss Re and Waymo is a non-paid and non-profit collaboration (Waymo and Swiss Re's new research collaboration).

- Compared to the *overall driving population* there was an 88% reduction in property damage claims and a 92% reduction in bodily injury claims. The *overall driving population* benchmark was expected to have 78 property damage and 26 bodily injury claims.
- Compared to the *latest-generation HDVs* there was an 86% reduction in property damage claims and 90% reduction in bodily injury claims. The *latest-generation HDV* benchmark would be expected to result in 63 property damage and 21 bodily injury claims.
- The two Waymo's bodily injury claims over 25.3M miles represent 0.8% of the total 241 collisions reported to NHTSA under the Standing General Order.
- The use of insurance claims data provides a robust and scalable framework for ongoing ADS safety assessment.

Introduction

Automated Driving Systems (ADS) operating without humans in the driver's seat (SAE Level 4+) have been the subject of numerous safety impact investigations. Many studies have focused on crash and injury prevention capabilities (Abdel-Aty et al., 2024; Cummings, 2023; Chen and Shladover, 2024; Kusano et al., 2024). In addition, Di Lillo et al. (2024a) utilized insurance claims data for safety evaluation, offering a distinct perspective on ADS performance. This body of research is crucial given the persistent safety challenges in road transportation, with the United States reporting nearly forty-one thousand fatalities in 2023 (NHTSA, 2024).

The evolution of ADS safety assessment builds upon a long history of motor vehicle safety research and insurance practices. Helm (1968) provides valuable historical context on motor vehicle liability insurance, which has shaped current approaches to risk assessment. Modern ADS safety studies have employed various methodologies, from analyzing crash reports (Favarò et al., 2017) to discussing the mileage required to demonstrate ADS reliability (Kalra & Paddock, 2016).

A recent collaborative work by multi-sector researchers has identified key challenges in executing retrospective safety assessments of ADS technology (Scanlon et al., 2024b). To address these challenges, they proposed the RAVE Checklist, promoting quality, validity, transparency, and proper interpretation in ADS safety studies (Scanlon et al., 2024b). The RAVE Checklist is useful in assessing the limitations of recent studies.

Several recent studies have attempted to compare ADS performance with human drivers, each with its own methodological approach and limitations: Chen and Shladover (2024) compared Waymo's ADS crash rate in San Francisco against rideshare benchmarks. While innovative in comparing similar operational domains, the study didn't account for reporting threshold differences. Cummings (2024a) assessed multiple ADS providers using Standing General Order (SGO)-reported collisions in California, comparing them to national police-reported crash rates. This study had several methodological issues, including incorrect crash rate computations and misaligned reporting thresholds. Correcting the crash rate computation, mismatched reporting thresholds, and including a vehicle type effect alignment resulted in a Waymo police-reported crash rate 2-times lower than the national average, even without adjusting for human underreporting of police reportable crashes or geographic effects (Kusano et al., 2024). Kusano et al. (2024) published results on Waymo ADS performance over 7.1 million miles, using state police-reported crash databases as benchmarks and applying corrections for vehicle type, road type, and underreporting. They reported 55% and 80% reductions in police-reported (without underreporting correction) and any-injury-reported (with underreporting correction) crash rates, respectively. Abdel-Aty et al. (2024) employed a matched case-control analysis but faced limitations due to misaligned reporting thresholds and lack of exposure data. Di Lillo et al. (2024a) introduced a novel approach using third-party liability insurance claims as a comparative safety measure,

analyzing Waymo's performance over 3.8 million autonomous miles. This method provided unique insights into crash events where responsibility had been determined by claims adjusters.

These studies highlight the importance of addressing methodological challenges such as aligning reporting thresholds, accounting for geographical variations, and considering underreporting of crashes. Blincoe et al. (2023) and SFDPH (2017) have documented the underreporting issue in national and San Francisco contexts, respectively.

On top of the challenges above, the operational design domain (ODD) of ADS technologies is continually expanding, influencing risk evolution. Chen et al. (2024) demonstrated how the unique distribution of where and when ADS operates affects baseline crash risk. Factors such as weather (Theofilatos & Yannis, 2014; Edwards, 1998), day-of-week (Doherty et al., 1998), and seasonality (Pawlovich et al., 1953) also influence crash risk, necessitating dynamic benchmarking approaches.

The insurance industry regularly reassesses ADS deployment risk for pricing and underwriting purposes. This has led to the development of insurance-relevant methods to quantify autonomous vehicle risk and safety (Di Lillo et al., 2024a; Dyro et al., 2024). Regulatory frameworks, such as A.R.S. Motor vehicle liability policy requirements § 28-4009 and INS California Insurance Code § 11580.1b, provide the legal context for these assessments. These regulations underscore the importance of developing robust methods for evaluating ADS safety that align with existing insurance and liability frameworks.

The societal implications of ADS safety assessments extend far beyond technical considerations. Milakis et al. (2017) highlight how the outcomes of such assessments can significantly influence policy decisions, public acceptance, and the broader integration of ADS technology into society. Their review emphasizes the need for comprehensive safety evaluations that can inform policymakers and shape public perception, directly aligning with the goals of our study.

Building upon this rich body of research, particularly extending the work of Di Lillo et al. (2024a), the current study presents a groundbreaking analysis of the Waymo Driver's performance across an unprecedented 25.3 million fully autonomous miles. This substantial increase in mileage provides a more robust dataset for evaluation, allowing for more definitive conclusions about ADS safety at scale. Key innovations include the introduction of a novel latest-generation human-driven vehicles (HDV) benchmark, continued use of third-party liability insurance claims as a safety metric, and the application of a robust statistical framework considering rare events (Garwood, 1936).

The present study directly considers the RAVE Checklist (Scanlon et al., 2024b) in its design, addressing key methodological challenges identified in previous studies. By demonstrating significant reductions in both property damage and bodily injury claims compared to both the overall driving population and the latest-generation HDV benchmark, our research provides crucial data to inform ongoing discussions about ADS integration into existing transportation systems, as explored by Fagnant & Kockelman (2015) and Litman (2023).

This study not only complements key works in the field but also addresses their limitations by providing a more comprehensive, up-to-date, and methodologically robust assessment of ADS safety performance. By analyzing an unprecedented volume of autonomous miles, introducing a novel benchmark, and utilizing third-party liability insurance claims, we aim to set a new standard for ADS safety evaluation. Our findings have the potential to significantly impact policy decisions, insurance practices, and public acceptance of ADS technology, contributing to the broader societal dialogue on the future of autonomous transportation.

Insurance Claims Data-Based Performance Assessments

The insurance industry has a long history of claims-based performance assessments, dating back to the mid-20th-century when motor insurance became widely mandatory in the US (Helm, 1968). An auto liability claim is defined as a request for compensation when, based on the available data, someone is alleged to be responsible for the damage to property or injury to another person, typically following a collision. Claims records became an important source of collision, vehicle, and driver data, eliciting a myriad of risk differentiation research. For example, early risk research venues assessed claims frequencies by vehicle make and model, suggested a relationship between driver age and claim frequency, and evaluated whether claims risk was associated with vehicle size (Joksch, 1981; Krishnan & Carnahan, 1985). In more recent years, claims data have also been used to assess the claims performance between electric vehicles and non-electric vehicles (HLDI 2020), vehicles equipped and not equipped with Advanced Driver Assistance Systems (ADAS) such as Automated Emergency Braking (AEB) (HLDI, 2023; Di Lillo et al. 2024b), as well as between human drivers and ADS deployments (Di Lillo et al., 2024a).

Insurance claims data helps overcome some key challenges when performing ADS safety impact evaluation. This data is particularly useful for defining equitable reporting thresholds (i.e., what constitutes a crash event) between an ADS and benchmark, and addressing the underreporting of collisions (i.e., is every relevant event being captured?) (Di Lillo et al., 2024a; Scanlon et al., 2024a; Scanlon et al., 2024b). The rate at which insurance claims are generated provides a consistent reporting threshold that exists for the majority of the US, where 87% of drivers are insured (IRC, 2021), although among human drivers there remains some potential for underreporting (Blincoe et al., 2023). Other data source options commonly utilized for crash risk analysis include naturalistic driving data (Blanco et al., 2016) and police-reported data (Scanlon et al., 2024; Kusano et al., 2024), which are both out-of-scope for the current analysis.

Introducing the use of claims data in ADS safety impact research, Di Lillo et al. (2024a) recently found that the Waymo One service significantly reduced the rate of third-party property damage and bodily injury claims with respect to the current road users across the San Francisco and Phoenix Metropolitan Area. There have since been notable increases in fully autonomous mileage (i.e., without a human driver seated behind the wheel), and expansions of the commercial service into Los Angeles and Austin. The current study expands on this prior work, now incorporating the most recent Waymo data and updating benchmarks to include Waymo's new operating regions.

When using claims data, careful consideration should be given to the insurance coverages they apply to, and to the differences between commercial and private passenger use of the vehicles to ensure an apples-to-apples comparison. Consideration also needs to be given to the liability frameworks being used for claims adjudication to ensure consistency of comparison. In the case of Waymo, cases have been handled to date through insurance claims adjudication aligned with the auto negligence laws applicable in each jurisdiction. This is no different than when two human drivers get into a collision and the claims adjudication that ensues. Claims adjudication involving ADS is for the most part simpler and more consistent given the increased availability of objective data related to the event giving rise to the claim, which is superior to evidence consisting of after-the-fact crash reconstructions based on vehicle damage and subjective statements from witnesses and crash participants – who might have an interest in avoiding fault determinations.

Safety Performance of Human-Driven Vehicles

Past ADS safety research has called for a wider range of safety targets, such as HDV benchmarks controlling for safety-relevant driver demographics, or benchmarks capturing better-than-average human driving (Goodall 2021, Fraade-Blanar et. al 2018, Nees 2019). As noted previously, safety impact analyses have largely

focused on the population-level collision performance of all active vehicles as a benchmark. Such population-level assessment provides the *safety impact* of ADS on the current collision record—that is, how ADS performs in comparison to- or impacts the current status quo. To complement existing work, alternative benchmarks representing elevated levels of collision performance can provide more nuanced insights into how ADS compares to safer cohorts of human-driven vehicles. However, such studies have been sparse, and no such study has been conducted on ADS systems without a human driver behind the wheel. Accordingly, it is informative to begin to define safer HDV benchmarks (a safer subset) in addition to the collision performance of the current on-road vehicle fleet (all HDV vehicles), because they may more closely represent a future version of the current overall driving population.

The current study introduces a novel benchmark titled "latest-generation HDVs", which is a custom filtering of historical claims data to focus on drivers of the latest vehicle model years. This benchmark is primarily motivated by the evolution of safety technology available in personal vehicles, such as the increasing inclusion of ADAS technologies. For example, over 50% of 2019 model year vehicles and 80% of 2021 model year vehicles are equipped with Automatic Emergency Breaking (AEB) in contrast to only 8.7% of 2016 model year vehicles (PARTS, 2024). Other ADAS technologies include Forward Collision Warning, Lane Keeping Assistance, and Adaptive Cruise Control, which saw 80.4%, 69.3%, and 59.5% penetration in 2021 model year vehicles, respectively, as opposed to 15.3%, 9.5%, 8.6% penetration in 2016 model year vehicles. The growing evolution in assisted human driving performance represents a level of safety being actively introduced to the fleet alongside ADS and can be indicative of emerging trends in population-level driving performance as newer vehicle models increasingly penetrate existing road traffic (see Figure 1). Note, therefore, that the performance of newer HDVs is a *combination* of human driver and vehicle assistance performance. An interesting research question is whether ADS deployments outperform drivers of the *latest-generation HDV* models. This addresses the question whether human drivers, *assisted* by evolving vehicle technology, are a better performing alternative to ADS in improving the overall safety of road users.

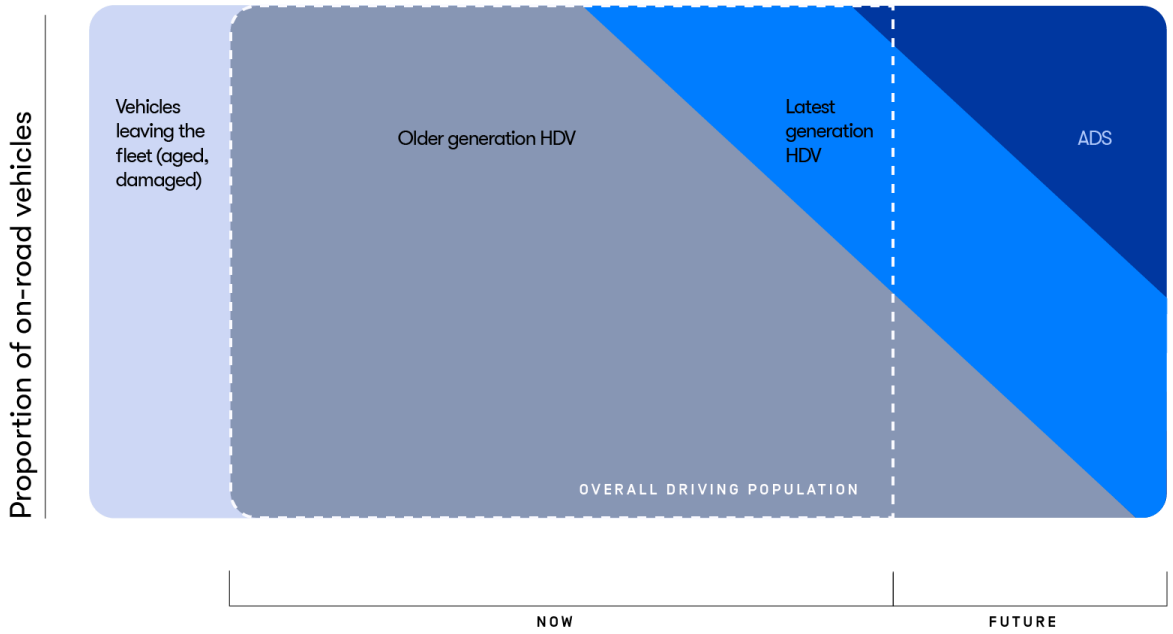

Figure 1. Conceptual illustration of the *overall driving population* benchmark, the *latest-generation HDV* benchmark, and ADS population proportion now and in the future. A *latest-generation HDV* benchmark conceptually represents a future version of the current *overall driving population*.

Prior to any benchmark comparison, this study tests the hypothesis that drivers of the *latest-generation HDVs* (defined here as model years 2018 to 2021) will have better third-party claims performance with respect to the

average human driver across all vehicle models currently on the road. The increasing inclusion of crash avoidance and ADAS features, such as automated emergency braking (AEB), has already been demonstrated to decrease crash rates in new vehicle models (Cicchino, 2016; Cicchino, 2024), and accounting for vehicle model year in insurance pricing is a well-established practice across the motor insurance industry (Lemaire 2013). In addition, other driver-related factors can also lead to improvements in the safety performance of drivers of newer vehicle models. For one, drivers of newer vehicle models tend to be older while teenage drivers—the most at-risk age group on the road—are more likely to drive older vehicles (AAA, 2024). One expectation is that drivers of *latest-generation HDVs* should have more experience and exhibit safer driving behavior in comparison to the population average (Kelly & Nielson, 2006). Overall, across technological and behavioral factors, we expect that drivers of *latest-generation HDVs* represent a higher-level of safety performance and, therefore, the need of a comparison with a safer cohort of HDVs.

Study Objectives and Research Questions

As noted previously, this current work is a continuation and expansion of Di Lillo et al. (2024a) to benchmark the Waymo ADS against a robust and responsibility-calibrated human-driven vehicle benchmark. In this current study, we reflect the expanded geographical and mileage expansion of the Waymo One service in both the HDV and ADS datasets. The liability claims data used in this study includes claims with accident dates reported up to July 31st, 2024 and are evaluated as of November 30th, 2024. We also update the zip-code calibration of the HDV benchmark to include claims from all new zip codes serviced by the Waymo One service. Finally, we update the benchmark data to use claims from 2017 to 2022 (as opposed to the prior study's 2016 to 2021), which includes more recent data and adjusts the HDV claims count to start from 2018 to align with Waymo's own start year.

Regular and consistent comparisons between ADS deployments and current vehicle fleets provide increasing confidence in the safety impact estimates as mileage increases and ensures that the public and other stakeholders are kept informed on the performance of ADS deployments as operations expand. This work takes advantage of the insurance industry's continuous maintenance of claims data and metadata to consistently benchmark human driving risk in new operating regions as ADS deployments scale, to regularly update comparisons while including more recent human driving incident data, and readily segment the human driving population into categories for more targeted comparative assessments due to the inherent structure of claims data. Three research questions are explored in this manuscript. First, do drivers of *latest-generation HDVs* have a better third-party claims performance than the *overall driving population*? Second, how does the Waymo ADS compare to these two benchmarks? Third, how does the Waymo ADS's auto liability claims performance at 25.3 million miles compare to the previous 3.8 million miles analysis?

Methodology

Benchmark Data

The benchmark was generated by analyzing claims arising from third-party personal auto liability insurance policies, which drivers are required to carry by law in California, Arizona, and Texas (California INS § 11580.1b; A.R.S. § 28-4009; Tex. Admin. Code. § 218.16). These claims were divided into Property Damage Liability and Bodily Injury Liability coverages. Property Damage Liability insures against damages that at-fault drivers cause to other people's vehicles and property in crashes. Bodily Injury Liability insures drivers against medical, hospital, and other expenses for occupants of involved vehicles or other road users, such as vulnerable road users. All third-party liability claims are considered in the benchmark. Since non-contact events are exceedingly rare in personal third-party auto liability claims within the HDV context (e.g. one party is financially

liable for damage against another party when no contact occurred), this benchmark is assumed to effectively represent contact events only.

The HDV benchmarks are based on fully anonymized property damage liability (PD) and bodily injury liability (BI) claims data from Accident Years 2017 to 2022 provided by Swiss Re, from over 500,000 claims and over 200 billion miles of exposure. The 2017-2022 period was chosen based on the data availability of recent HDV claims data, as longer periods are required for claims to be filed, adjudicated, and settled (Taylor, 2023). The benchmark was calibrated using both mileage (driving exposure) and residence zip-code (geographic region). Specifically, only the claims associated with vehicles registered to addresses (i.e., where the insured resides) within Waymo's operating zip codes in San Francisco, the Phoenix metropolitan region, Los Angeles, and Austin were included. The benchmark for the totality of these claims is referred to as the *overall driving population* benchmark (claims from all vehicles during the 2017-2022 period). For the *latest-generation HDV* benchmark which captures the increasing fitment of the latest ADAS technologies, we selected only insurance policies on vehicles garaged within Waymo's operating zip codes with a model year between and including 2018 and 2022, which represented approximately 17% of policies, and yielded over 80,000 claims and over 40 billion miles of exposure.

To produce these benchmarks, we converted the exposure measure from vehicle years, the exposure basis used in insurance metrics, to mileage by calculating the average annual vehicle-miles-traveled (VMT) per vehicle within each of Waymo's operating regions, due to variation in driving patterns across different US regions. To increase robustness in estimation, we averaged both across the state and across the city to generate two estimations for each region. Each estimate requires two values: annual mileage and the number of vehicles in the targeted region. For mileage, we use aggregate VMT statistics provided by the Federal Highway Administration (FHWA) which reports the total monthly VMT by state (FHWA, 2022) and total annual VMT by urbanized area (FHWA, 2023a). For per-vehicle statistics, we use the FHWA's annual vehicle registration statistics (FHWA, 2023b) and US Census demographic data (U.S. Census Bureau). For the analysis presented below, we chose to use estimates based on state VMT data because they yielded a lower (more conservative) baseline frequency, given that our estimates for the state average annual miles per vehicle were higher than those of the urbanized areas. For the *latest-generation HDV* benchmark, we note that drivers of newer vehicle models may on average drive longer distances, as found in a survey of approximately 15,000 American drivers (AAA, 2023). While the study represents only a sample of American drivers and precise values should not necessarily be representative of the population, the results of this study show an approximately 14% increase in annual mileage for drivers of vehicles less than four years old in comparison to all drivers. This means the *latest-generation* benchmark should be *lower* if accounting for mileage variations, and the relativity between the overall and the *latest-generation HDV* benchmarks should be *greater*.

Four claim frequencies were independently calculated for vehicles within Waymo's operating zip codes in San Francisco, Phoenix, Los Angeles, and Austin. Thereafter, these frequencies were proportionally weighted according to Waymo's mileage distribution across these four regions. This resulted in the definition of a benchmark driving population which had driven the same driving distribution across geographic regions as the Waymo Driver.

Waymo ADS Data

Waymo provided both the fully anonymized claims' data and zip-code level vehicle miles traveled across all 25.3 million miles to Swiss Re. The Waymo data included in this study are from miles driven and claims that occurred during the period from January 1, 2018, through July 31, 2024. Waymo claims data was collected from Waymo’s claims administrator. The data was then reviewed by Swiss Re. Claims not applicable to a private passenger risk exposure, mainly claims not related to a contact event, were removed from the analysis.

For the purposes of this study, we are using claims that are resolved with a liability payment as of the evaluation date of November 30th, 2024 or are still open and likely to resolve with an indemnity payment as a proxy for partial or full responsibility. For open claims, we are using case reserves as an indication of the adjustor's assessment of the likelihood of the claim to be closed with payment. It is possible for some of the open claims to close without any liability payment in the future. Following the methodology outlined in Di Lillo et al., 2024a, claim count development is considered by reviewing known events. Future development (additional claim counts) is still possible from unreported and/or underreported claims. For Waymo, claims emergence months after the collision date is less likely due to the ability to detect event occurrence in a timely manner as compared to HDV.

Waymo includes an additional medical payments coverage (Med-Pay) for all passengers, which reimburses medical bills up to a specified limit regardless of who was responsible for the collision. This coverage is not mandatory and therefore not always available in a private passenger insurance product. To the extent that any medical claim arose from an event where Waymo had corresponding liability claims for either Bodily Injury or property damage from any party, that Med-Pay claim is also counted in the Bodily Injury claim count for the purposes of this analysis. This ensures that claim counts are not artificially decreased due to the existence of the Med-Pay coverage and allows for a fairer comparison to the HDV benchmarks where Med-Pay coverage may not be available. In general, when doing comparisons of claims data, an understanding and review of the applicable coverages is important to ensure an apples-to-apples comparison. All cases were handled through insurance claims adjudication aligned with the auto negligence laws applicable in each jurisdiction, the same as standard practices for HDV-to-HDV collisions claims adjudication.

For significance testing, a statistically significant difference between claims frequencies was concluded ascertained if the 95% confidence intervals of Waymo's collision frequencies and their corresponding benchmarks did not overlap. Namely, that there was a difference even when accounting for statistical uncertainty due to the limited mileage volume. For the HDV benchmarks, due to the large sample, a normal approximation confidence interval was used, which considered the mileage distribution between San Francisco and Phoenix when computing the standard error. For Waymo's operations, due to the smaller sample and exposure size, confidence intervals were calculated using the Poisson Exact Method (Garwood, 1936).

Results

Relationship between Vehicle Model Year and Liability Claims Performance in the Current Vehicle Fleet

To understand and illustrate the relationship between vehicle model year and liability claims performance, the average liability claims performance for each vehicle model year between 2010 and 2022 was analyzed (Figure 2). On average, bodily injury claims frequency decreased by 2.9% year-over-year from 2010 to 2022. Property damage claims decreased by 3.3% year-over-year over this same period. This demonstrates a directional change in BI and PDL frequencies by model year. While a univariate analysis would not capture influences from factors other than model year, this shows directional signs of reduced crash frequency in drivers of newer vehicle models.

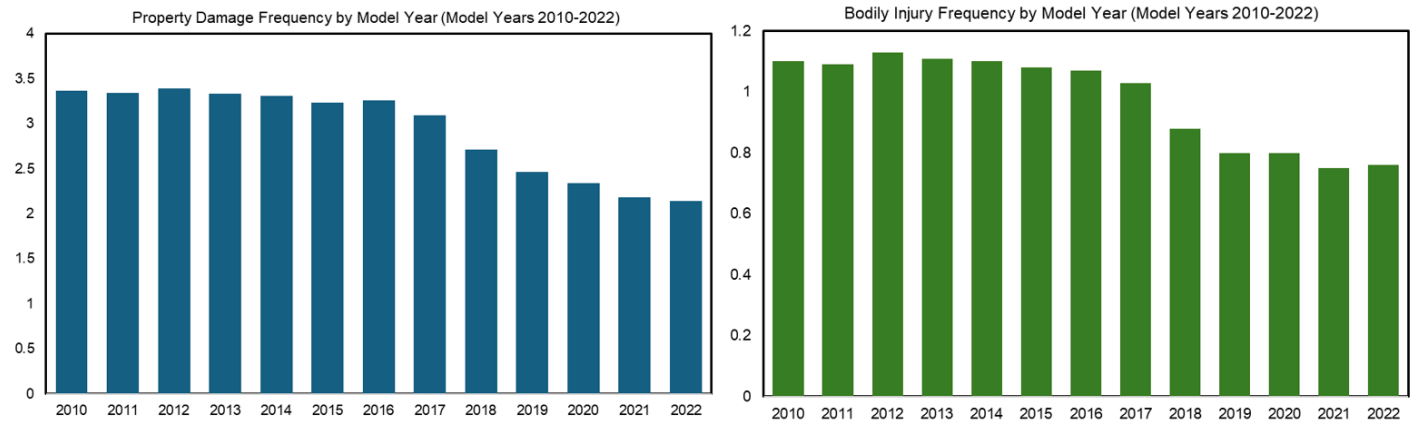

Figure 2. Year-by-year plot depicting relationship between human driving claims frequency and model year category. Claims data from 2017 to 2022 are used in this analysis.

Benchmarking the Waymo ADS

Figure 3 shows the results of the comparison of liability insurance claims for property damage (left) and bodily injury (right) of Swiss Re's HDV baselines and Waymo ADS data.

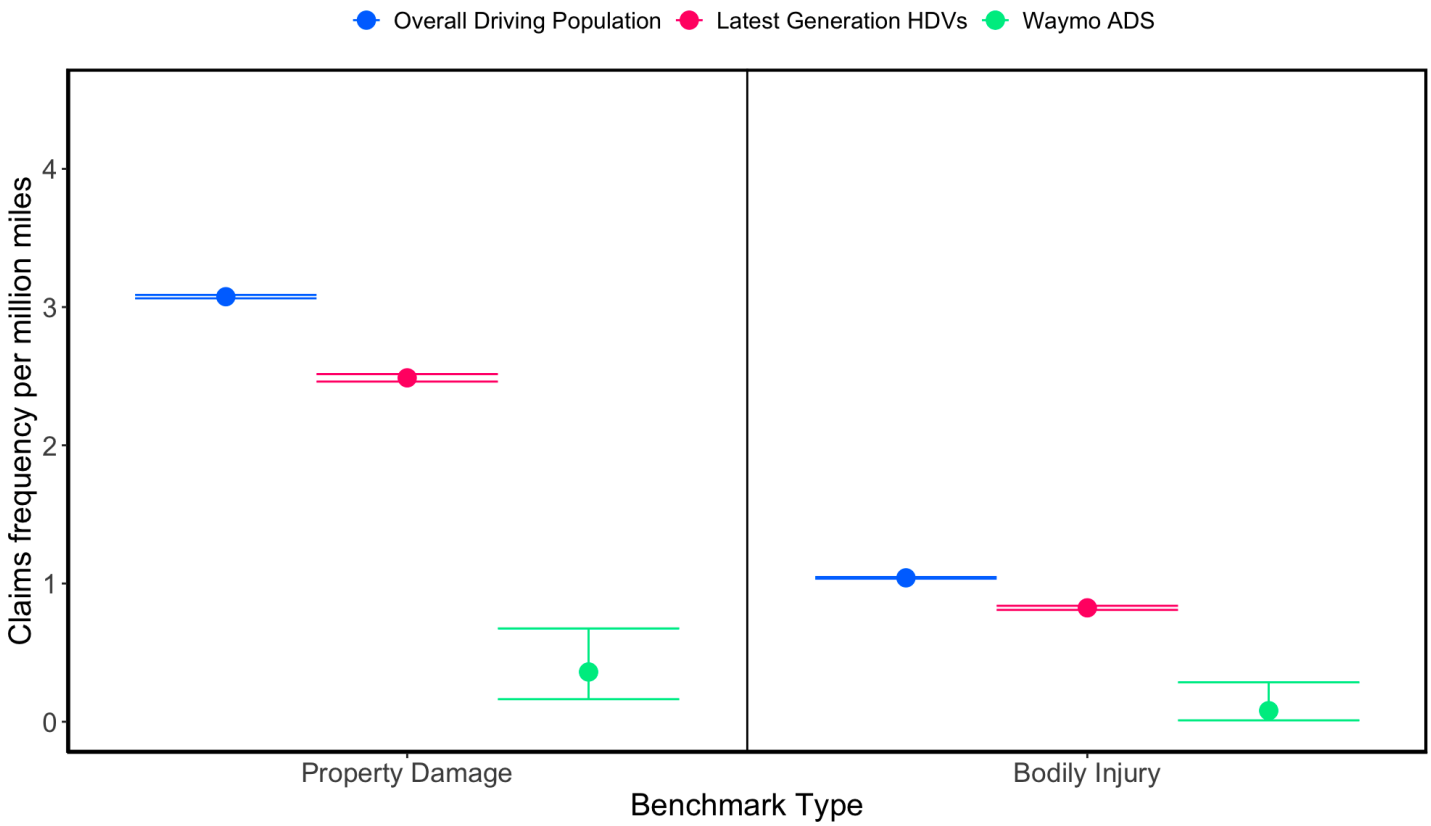

Figure 3. Comparison of Swiss Re *overall driving population* and *latest-generation HDV* baselines with Waymo ADS liability insurance claims for property damage (left) and bodily injury (right).

When examining claims filed under third party liability policies, the Waymo ADS yielded nine property damage and two bodily injury claims, arising from 10 unique collisions, over 25.3 million miles. Two of the property damage claims and both bodily injury claims are still open (unresolved) and could close without any liability

payment. In comparison, the *overall driving population* benchmark was expected to have 78 property damage and 26 bodily injury claims for an equivalent amount of driving exposure in the deployed driving regions. The *latest-generation HDV* benchmark would be expected to result in 63 property damage and 21 bodily injury claims.

When compared to the *overall driving population* benchmark, the Waymo ADS saw an 88% reduction in third-party property damage claims (ADS_{PDL}: 0.36 [0.163, 0.675] vs $Overall_{PDL}$: 3.08 [3.063, 3.088] claims per million miles), and a 92% reduction in third-party bodily injury claims (ADS_{BI}: 0.08 [0.010, 0.285] vs $Overall_{BI}$: 1.04 [1.035, 1.047] claims per million miles), respectively.

When compared to the *latest-generation HDV* benchmark, the Waymo ADS recorded an 86% reduction in third-party property damage claims (ADS_{PDL}: 0.36 [0.163, 0.675] vs $Latest\text{-}Generation_{PDL}$: 2.49 [2.461, 2.515] claims per million miles) and a 90% reduction in third-party bodily injury claims (ADS_{BI}: 0.08 [0.010, 0.285] vs $Latest\text{-}Generation_{BI}$: 0.82 [0.809, 0.839] claims per million miles) were observed.

There was a 19% reduction in property damage claims between the *overall driving population* benchmark and the *latest-generation HDV* benchmark ($Overall_{BL}$: 3.08 [3.063, 3.088] vs $Latest\text{-}Generation_{PDL}$: 2.49 [2.461, 2.515] claims per million miles). Likewise, there was a 21% reduction in bodily injury claims between the *overall driving population* benchmark and the *latest-generation HDV* benchmark ($Overall_{BI}$: 1.04 [1.035, 1.047] vs $Latest\text{-}Generation_{BI}$: 0.82 [0.809, 0.839] claims per million miles).

Comparing Waymo ADS Performance at 3.8 and 25.3 Million Miles

In Figure 4, we show side-by-side how the Waymo ADS performance in relation to the HDV benchmark has changed since our last study at 3.8 million miles (Di Lillo et al., 2024a).

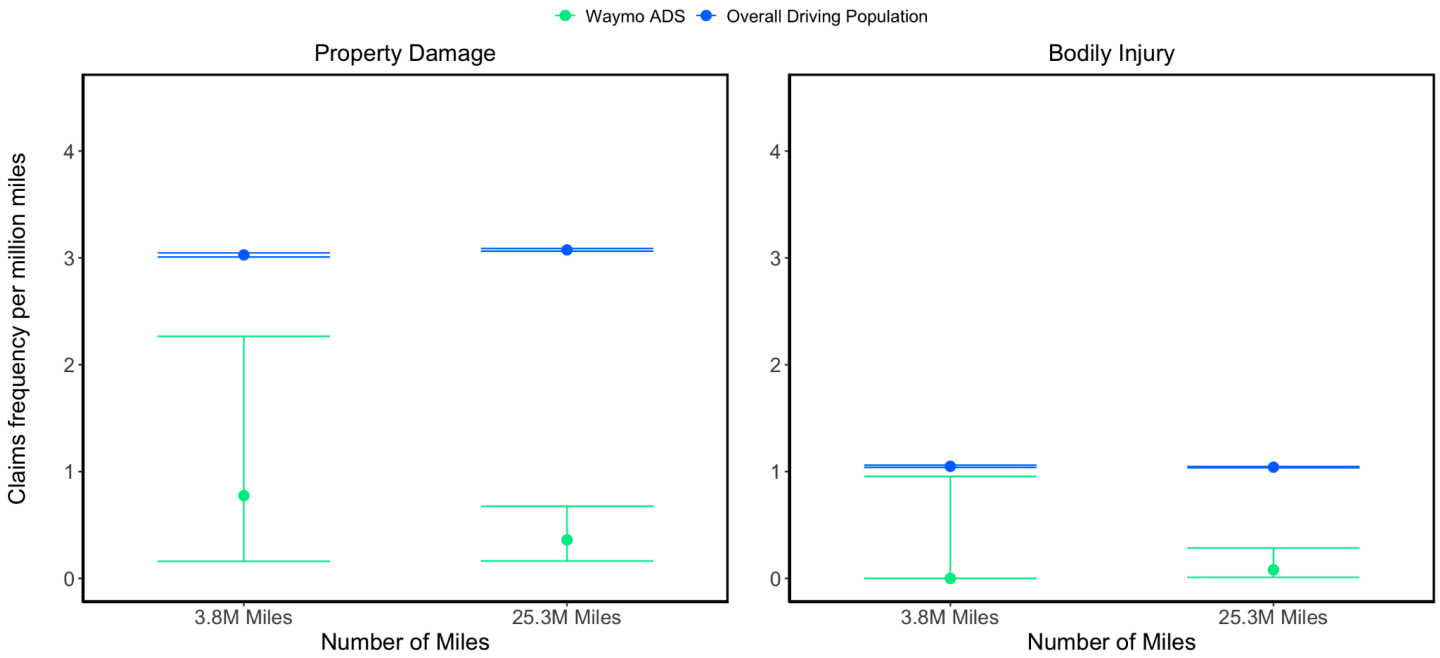

Figure 4. Comparison of Waymo ADS liability insurance claims and the corresponding *overall driving population* benchmarks for property damage (left) and bodily injury (right) at both 3.8 and 25.3 million miles. Both benchmarks are based on property damage liability (PD) and bodily injury liability (BI) claims data from 2017 to 2022, which represents an update from Di Lillo et al.'s prior 2024 study (Di Lillo et al. (2024a)) which used data from 2016 to 2021.

Over the last 21.5 million miles accumulated since the last study, Waymo's operations have generated an additional six property damage claims (0.28 PDL claims per million miles) and two bodily injury claims (0.09 BI claims per million miles). Although we see no statistical difference between Waymo performance at 3.8 and 25.3 million miles due to overlapping confidence intervals in both property damage ($ADS_{3.8M}$: 0.78 [0.160, 2.266] vs $ADS_{25.3M}$: 0.36 [0.163, 0.675]) and bodily injury ($ADS_{3.8M}$: 0.00 [0.000, 0.954] vs $ADS_{25.3M}$: 0.08 [0.010, 0.285]), we note that 95% confidence intervals have narrowed substantially during this time. As Waymo operations continue to scale, we expect that statistical confidence in this result will also increase.

We also note that the *overall driving population* baseline has stayed relatively constant, with no significant difference between 3.8 million and 21.5 million miles for both property damage ($Overall_{3.8M}$: 3.26 [3.24, 3.27]*vs* $Overall_{25.3M}$: 3.08 [3.063, 3.088]) and bodily injury ($Overall_{3.8M}$: 1.11 [1.10, 1.12] vs $Overall_{25.3M}$: 1.04 [1.035, 1.047]) benchmarks. This is due to minimal change in the proportion of fully autonomous mileage between driving regions, despite a substantial increase in mileage. At 3.8 million miles, the proportion of fully autonomous mileage in San Francisco and Phoenix was 73.3% and 26.7%, respectively, with no mileage in Los Angeles or Austin. In comparison, at 25.3 million miles, the proportion of mileage in San Francisco, Phoenix, Los Angeles and Austin was 28.2%, 67.4%, 4.3%, 0.1%, respectively. In addition, the benchmark data now uses claims from 2017-2022 to include more recent data, which is an update from the prior study's use of claims from 2016-2021. As a result, we also see a 6% decrease (3.26 to 3.08) in PD claims, and a 6% decrease (1.11 to 1.04) in BI claims frequencies between the updated 3.8M benchmarks in Figure 4 and those provided in Di Lillo et al. (2024a).

Discussion

The Waymo ADS outperformed the *overall driving population* benchmark with respect to property damage and bodily injury third party claims. This represents a 9-times reduction in property damage (3.08 to 0.36) and 12-times reduction in bodily injury (1.04 to 0.08) claims to the overall on-road vehicles in the areas that the Waymo ADS currently operates (the status quo). Because these data represent third-party claims, they should be considered a reflection of the safety performance in events where Waymo is alleged to have some responsibility for damages to a third-party. These data represent a subset of the total collision involvement of the Waymo ADS, which includes collisions that did not result in third party auto liability claims towards Waymo. The research questions in the current study should be considered unique from safety performance assessments of the total crash performance regardless of the Waymo ADS's role during the crash sequence (e.g. Teoh and Kidd, 2016; Scanlon et al., 2024; Kusano et al., 2024; Chen and Shladover, 2024).

Assessments of total crash involvement (irrespective of responsibility) and third-party claims (e.g. an evaluation of when Waymo is alleged to have some responsibility for damages to a third-party) are distinctly different analyses. To put this in perspective, the two bodily injury third-party claims reported here over 25.3 million miles represent 7% of the 28 total any-injury-reported collisions Waymo reported to NHTSA under the SGO, or 0.8% of the total 241 collisions reported to NHTSA under the SGO (Waymo, 2024). Further, these two bodily injury claims represent 2% of the 97 any-injury-reported collisions that an average human driver would be involved in over this 25.3M mile period (Waymo, 2024). As the property damage thresholds are different in the SGO and insurance data, they are not suitable to comparisons.

A key innovation of the current study was the introduction of the *latest-generation HDV* benchmark. The *latest-generation HDV* outperformed the *overall driving population* benchmark with a property damage and bodily injury third-party claim rate that was 19% and 21% lower, respectively. Given the correlations between model year, driver age, and vehicle active safety features, this result could be a reflection of more experienced drivers operating better-performing vehicles. Because the *latest-generation HDV* benchmark represents the increasing penetration rates of the latest passenger vehicle technology active on roads, it provides informative

feedback on ADS performance with respect to the latest technology with other variables such as driver age acting as a confounder. The Waymo ADS had a property claim and bodily injury third-party claim rates that were 7.0-times and 10.4-times lower, respectively than the *latest-generation HDV* benchmark.

After expanding into two new locations and increasing mileage by over 500% from 3.8 to 25.3 million miles, the Waymo ADS performance remained statistically significant with respect to the established benchmarks. Directly comparing the Waymo ADS's third-party claims rates between the 3.8 million and 25.3 million miles did not result in statistical differences between the Waymo performance over time, and the claims rates remained very low with 9 property damage and 2 bodily injury claims totaling over 25.3 million miles. Relative performance over time should be monitored, and, as more mileage is accumulated, performance with more strict geographic controls will be statistically testable. The expectation is that, as Waymo's operations continue to scale, statistical certainty around Waymo's third-party claims frequencies will continue to improve.

Although crash frequency is an important indicator of safety, severity should also be considered. Direct comparison of injury severity between human and ADV-related claims is challenging, however, because claims data typically relies on dollar values. The appendix provides further details on the two bodily injury claims involving contact events, both of which are still open.

The Effects of Mileage Exposure and Future Considerations

For both benchmarks, estimated and proxy values were used in parts of the calculations. Namely, the number of miles driven was estimated from the number of insurance policy years. There may be differences in the amount of mileage driven by policy holders by geographic region that are not accounted for in this analysis. Our mileage estimation practices are developed to prevent against an overestimation of human driving claims per million miles; that is, to prevent depicting human driving performance as poorer than it is. Such precautions are made by choosing to use mileage estimates which would yield a lower baseline frequency (i.e., choosing state-based aggregate data over urbanized-region-based aggregate data). For this estimation uncertainty to result in a "false positive," that is, a significant demonstration of positive safety impact which otherwise would be found inconclusive, our human driving mileage estimates would have to have instead been underestimated (which would lead to an overestimated baseline). Here, the mileage would have to have been underestimated by 49% for a false positive to occur in a comparison with any HDV benchmark. However, it is important to note that drivers of newer vehicle models may on average drive longer distances (AAA, 2023), and therefore the *latest-generation HDV* benchmark may have a better relative performance to the *overall driving population* benchmark than depicted in this study. Future related work can further improve this level of estimation uncertainty by incorporating insurance telematics data that precisely track vehicle mileage, or vehicle mileage data based on reported odometer readings.

The garaging zip code of the insured vehicle was used as a proxy for the city (Phoenix, San Francisco, Los Angeles, Austin) in which the vehicle drives. Waymo also almost exclusively operated on surface streets (non access-controlled freeways) with a unique distribution of driving that is representative of a ride-hailing fleet. In contrast, the benchmark represents the privately insured driver population that resides in these geographic regions. The associated benchmark mileage has more freeway driving than the Waymo ADS. There are several considerations when examining these results with respect to this limitation. First, freeway driving has a lower crash rate (Scanlon et al., 2024a). Including freeway driving makes this benchmark crash rate artificially lower, so, by including freeways in this study's benchmark, the benchmark crash rate underestimates the true driving risk of where the Waymo ADS operates. Second, driving outside of these denser urban areas that the Waymo ADS operates would likely represent a reduction in overall relative crash risk. For example, commuters from the city would likely experience a reduced crash risk as they travel to less densely populated areas (Chen et al., 2024). Previous studies have shown that most injury collisions occur within a small radius from

residency, and that American drivers rarely travel far from their place of residence, with approximately 80% of one-way household trips being less than 10 miles (DOE, 2022). Third, the benchmark drivers garaged in the Waymo deployment area are not operating with the same distribution of mileage within the geographical limits as the Waymo ADS. Chen et al. (2024) explored the effect of Waymo's driving distribution on benchmark crash risk and found that - should the benchmark driving distribution match Waymo's in San Francisco, Phoenix, and Los Angeles - the benchmark police-reported crash rates would have been between 14% and 38% higher. Due to all three of these limitations being expected to artificially suppress the benchmark crash rate (underestimation), the benchmarking results in this study are considered to be conservative. Surely, there is an opportunity in future work to leverage new data, such as insurance telematics, to more precisely define and leverage the benchmark driving exposure data to better account for this potential confounder.

Other Factors Influencing ADS and HDV Claims Counts

We may see some differences in the number and nature of third-party liability claims filed against Waymo that are not related to performance, i.e. claims filed that would likely not have been filed against a human driver, with recent studies finding potential increased perceived liability of fully autonomous vehicles after collisions in comparison to a human driver in the same position (De Freitas et al., 2022). There may also be additional claims arising from the commercial ride-hailing nature of Waymo's operations such as claims filed by the passengers, which are comparably rarer in the personal motor liability context due to the possibility of passengers being insured under the same policy as their driver (e.g., as a family member) (California Department of Insurance, 2023) or personal relationships with the driver affecting the decision to file a claim. This discrepancy would be expected to be mitigated by using a commercial liability benchmark capturing claims frequencies for human driver ride-hailing. This is a potential future development of our work. Generally speaking, if a commercial ride-hailing driving benchmark were used, it is expected to yield a higher frequency than the personal liability claims benchmark presented in this study, also due to a range of increased collision risk factors such as interacting with smartphones and driving fatigue (Mao et al., 2021). Because of the differences in personal liability and commercial liability claims generation, the current benchmark based on personal liability claims is considered to underrepresent a commercial liability benchmark (a personal liability benchmark is believed to represent a lower collision frequency than a commercial liability benchmark) and is therefore representing a conservative assessment of the Waymo ADS safety performance.

Additional Future Considerations

Claims data continue to open interesting paths for research in the larger context of safety impact assessment. In addition to pursuing the above-mentioned directions, such as the use of telematics data or odometer readings to reduce the need for estimations and proxies, further directions could first lean into claims data's incorporation of liability—how many collisions lead to liability claims in ADS-involved contact events? In HDV-only events? Are there differences between the types of collisions that lead to liability claims in ADS and HDV-only driving?

Second, in the future it may be possible to perform claims adjudication with safety envelope models or human behavioral reference models, wherein driving performance is compared against some type of model behavior. This approach aims to establish a normative reference for acceptable or reasonable driving behavior and can potentially provide a foundation for defining and assessing claims. *Safety envelope models* compare against a spatio-temporal envelope using a set of kinematic assumptions. Examples include Responsibility-Sensitive Safety (RSS) (Shalev-Shwartz et al., 2017), the Safety Force Field (Nister et al., 2019), models based on reachable sets (Althoff & Dola, 2014; Pek et al., 2020), and the Instantaneous Safety Metric (ISM; Weng et al., 2020). *Human behavioral reference models* may represent a proper human behavioral response to an impending hazard. Examples include the Careful and Competent Driver Model (CCDM) (Japan Automobile

Manufacturers Association, Inc., 2022), the Fuzzy Safety Model (FSM) (Mattas et al., 2020; 2022), and the NIEON model (Engström et al., 2024). These can be used to represent a driver that “can anticipate the risk of a collision and apply proportionate braking” (*UN Regulation No. 157 Amend.4*, 2023).

Third, future comparisons could further leverage claims data to define additional human driving subsets. In our current study, claims data provide insights into drivers of newer vehicles, who tend to be older and more experienced, and benefit from the inclusion of ADAS and other safety-relevant technologies. Further studies could delve deeper into the latter, i.e. investigating ADS performance in comparison to specifically ADAS-equipped vehicles, which is increasingly becoming the standard for vehicles on the road. The structure of claims data could further be leveraged to design a *distribution* of human driving performance, allowing researchers to explore further quantitative claims about ADS performance in comparison to better-performing human drivers.

Conclusions

This study presents a comprehensive analysis of the Waymo Automated Driving System's (ADS) safety performance, leveraging an expanded dataset that reflects the geographical and operational growth of the Waymo ADS service. By utilizing insurance claims data, we provide a novel, responsibility-calibrated benchmark for comparing ADS performance to *latest-generation Human-Driven Vehicles (HDV)* in addition to comparison with *overall driving population*, addressing key limitations in previous assessment methodologies.

When examining claims filed under third party liability policies, the Waymo ADS yielded nine property damage and two bodily injury claims, arising from 10 unique collisions, over 25.3 million miles. Two of the property damage claims and both bodily injury claims are still open (unresolved) and could close without any liability payment. Our findings demonstrate that the Waymo ADS significantly outperformed both the *overall driving population* and the *latest-generation HDV* benchmarks in terms of third-party liability claims. Specifically, we observed an 88% reduction in property damage claims (0.36 vs 3.08 claims per million miles) and an 92% reduction in bodily injury claims (0.08 vs 1.04 claims per million miles) compared to the *overall driving population*. The *overall driving population* benchmark was expected to have 78 property damage and 26 bodily injury claims. Even when benchmarked against drivers of newer vehicle models (2018+), the *latest-generation HDV* benchmark, which represents a comparison with vehicle model years with lower claims frequency (see Figure 2), the Waymo ADS maintained substantial improvements of 86% and 90% in property damage and bodily injury claims, respectively. The *latest-generation HDV* benchmark would be expected to result in 63 property damage and 21 bodily injury claims.

There was a low frequency of Waymo ADS third-party claims in both the 3.8- and 25.3-million-mile operational periods. The low claim counts, despite the increased mileage, underscores the ADS’s consistent safety performance as it scaled across diverse operational domains. Notably, the 95% confidence intervals narrowed substantially during this time.

The introduction of the *latest-generation HDV* benchmark represents a methodological advancement in ADS performance evaluation. By comparing against drivers of newer vehicles, we establish a more forward-looking and challenging benchmark that accounts for recent improvements in vehicle safety technologies and potentially more experienced drivers.

Appendix

Claim #1
For BI claim #1, the Waymo AV was traveling on a road with one lane in each direction, approached an intersection and prepared to make an unprotected left turn when a vehicle on the opposite side of the intersection slowed down to prepare for an unprotected left. As the Waymo AV was beginning its unprotected left turn, another vehicle became visible as it passed on the right side of the line of traffic behind the left turning vehicle, partially in the westbound bicycle lane and partially in a paved area between the bicycle lane and the curb. The vehicle in the bicycle lane then entered the intersection and the front passenger side corner of that vehicle made contact with the passenger side of the Waymo AV.

Claim #2:
For BI claim #2, the Waymo AV was traveling in the left lane and stopped at a red light at an intersection alongside a passenger vehicle in the right lane. After the light turned green, both the Waymo AV and the adjacent passenger car proceeded into the intersection. While in the intersection, a passenger car ran the red light and the front left corner and left side of this vehicle made contact with the front right of the Waymo AV and the front of the adjacent passenger car. After impact, the vehicle that ran the red light struck pedestrians that had been standing on the sidewalk on the northwest corner of the intersection.

References

AAA Foundation for Traffic Safety. (2024). American Driving Survey: 2023. *Driver Behavior & Performance.*

A.R.S. Motor vehicle liability policy requirements § 28-4009.

Abdel-Aty, M., Ding, S. A matched case-control analysis of autonomous vs human-driven vehicle accidents. *Nature Communications* 15, 4931 (2024).

Althoff, M., & Dola, J. M. (2014). Online Verification of Automated Road Vehicles Using Reachability Analysis. *IEEE Transactions on Robotics*, 30(4).

Bin-Nun, A.Y., Derler, P., Mehdipour, N. & Tebbens, R. D. (2022). How should autonomous vehicles drive? Policy, methodological, and social considerations for designing a driver. *Humanit Soc Sci Commun* 9, 299.

Blanco, M., Atwood, J., Russell, S. M., Trimble, T. E., McClafferty, J. A., & Perez, M. A. (2016). Automated vehicle crash rate comparison using naturalistic data. Virginia Tech Transportation Institute.

Blincoe, L., Miller, T., Wang, J.-S., Swedler, D., Coughlin, T., Lawrence, B., Guo, F., Klauer, S., & Dingus, T. (2023, February). The economic and societal impact of motor vehicle crashes, 2019 (Revised) (Report No. DOT HS 813 403). National Highway Traffic Safety Administration.

California Department of Insurance (2023). *Automobile Insurance Text Version*. https://www.insurance.ca.gov/01-consumers/105-type/95-guides/01-auto/auto101.cfm

Chen, J. J., & Shladover, S. E. (2024). Initial Indications of Safety of Driverless Automated Driving Systems. *arXiv preprint arXiv:2403.14648*.

Chen, Y., Scanlon, J. M., Kusano, K. D., McMurry, T., Victor, T. (2024). Dynamic Benchmarks: Spatial and Temporal Alignment for ADS Performance Evaluation. *ArXiv* preprint arXiv:2410.08903.

Cicchino, J. B. (2017). Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. *Accident Analysis & Prevention*, 99, 142-152.

Cook, C. F. (1970). Trend and loss development factors. Proceedings at CAS (Vol. 57, No. 107).

Di Lillo, L., Gode, T., Zhou, X., Atzei, M., Chen, R., & Victor, T. (2024a). Comparative safety performance of autonomous-and human drivers: A real-world case study of the Waymo Driver. *Heliyon*, *10*(14).

Di Lillo, L., Triscari, A., Zhou, X., Dyro, R., Li, R., & Pavone, M. (2024b). Performance assessment of ADAS in a representative subset of critical traffic situations. *ArXiv* preprint arXiv:2409.16942.

DOE (US Department of Energy). (2022). More than Half of all Daily Trips Were Less than Three Miles in 2021. FOTW #1230, March 21, 2022: More than Half of all Daily Trips Were Less than Three Miles in 2021 | Department of Energy

Doherty, S. T., Andrey, J. C., & MacGregor, C. (1998). The situational risks of young drivers: The influence of passengers, time of day and day of week on accident rates. *Accident Analysis & Prevention*, *30*(1), 45-52.

Dyro, R., Foutter, M., Li, R., Di Lillo, L., Schmerling, E., Zhou, X., & Pavone, M. (2024). Realistic Extreme Behavior Generation for Improved AV Testing. *arXiv preprint arXiv:2409.10669*.

Edwards, J. B. (1998). The relationship between road accident severity and recorded weather. *Journal of Safety Research*, *29*(4), 249-262.

Engstrom, J., Liu, S.-Y., Dinparastdjadid ,, A., & Simoiu, C. (2024). Modeling road user response timing in naturalistic traffic conflicts: A surprise-based framework. *Accident Analysis & Prevention*, 198.

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

Favarò F.M., Nader N., Eurich S.O., Tripp M., & Varadaraju N. (2017). Examining accident reports involving autonomous vehicles in California. PLoS One, 12(9):e0184952.

Feng, S., Feng, Y., Yu, C., Zhang, Y., & Liu, H. X. (2020). Testing Scenario Library Generation for Connected and Automated Vehicles: An Adaptive Framework. *IEEE Transactions on Intelligent Transportation Systems*, 23(2), 1213-1222.

FHWA (Federal Highway Administration). (2023a). Highway Statistics Series. Table HM-71 - Highway Statistics 2021 - Policy | Federal Highway Administration (dot.gov)

FHWA (Federal Highway Administration). (2023b). Highway Statistics Series. Table MV-1 - Highway Statistics 2021 - Policy | Federal Highway Administration (dot.gov)

FHWA (Federal Highway Administration). (2022). Traffic Volume Trends. Office of Highway Policy Information - Policy | Federal Highway Administration (dot.gov)

Garwood, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika*, *28*(3/4), 437-442.

Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A Survey of Deep Learning Techniques for Autonomous Driving. Journal of Field Robotics, 37(3), 362-386.

Helm, R. E. (1968). Motor vehicle liability insurance: A brief history. *John's L. Rev.*, *43*, 25.

HLDI (Highway Loss Data Institute). (2020). Insurance losses of electric vehicles and their conventional counterparts while adjusting for mileage. *Bulletin, 37*(25).

HLDI (Highway Loss Data Institute). (2023). Compendium of HLDI collision avoidance research. *Bulletin, 40*(4).

INS California Insurance Code § 11580.1b.

IRC (Insurance Research Council). (2021). Uninsured Motorists, 2021 Edition.

Japan Automobile Manufacturers Association, Inc. (2022, December). Automated Driving Safety Evaluation Framework Ver 3.0. Retrieved October 18, 2024, from https://www.jama.or.jp/english/reports/docs/Automated_Driving_Safety_Evaluation_Framework_Ver3.0.pdf

Huang, X., Wang, P., Cheng, X., Zhou, D., Geng, Q., & Yang, R. (2020). The ApolloScape Open Dataset for Autonomous Driving and its Application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10), 2702-2719.

Joksch, H. C. (1981). A relation between collision claim frequency and driver age distribution in HLDI data. *Accident Analysis & Prevention*, *13*(2), 147-151.

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*, Volume 94, 182-193.

Kelly, M., & Nielson, N. (2006). Age as a variable in insurance pricing and risk classification. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *31*, 212-232.

Kiran, B.R., Sobh, I., Talpaert, V., Mannion, P., Al Sallab, A.A., & Yogamani, S. (2022). Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(6), 4909-4926.

Krishnan, K. S., & Carnahan, J. V. (1985). Analysis of the effect of car size on accident injury probability using automobile insurance data. *Accident Analysis & Prevention*, *17*(2), 171-177.

Kusano, K. D., Scanlon, J. M., Chen, Y. H., McMurry, T. L., Chen, R., Gode, T., & Victor, T. (2024). Comparison of Waymo rider-only crash data to human benchmarks at 7.1 million miles. *Traffic Injury Prevention*, 1-12.

Lemaire, J. (2013). *Automobile insurance: actuarial models* (Vol. 4). Springer Science & Business Media.

Litman, T. (2023). Autonomous vehicle implementation predictions: Implications for transport planning. Victoria Transport Policy Institute. https://www.vtpi.org/avip.pdf

Litman, T. (2023). Autonomous vehicle implementation predictions: Implications for transport planning. *Victoria Transport Policy Institute*. https://www.vtpi.org/avip.pdf

Mao, H., Deng, X., Jiang, H., Shi, L., Li, H., Tuo, L., Shi, D., & Guo, F. (2021). Driving safety assessment for ride-hailing drivers. *Accident Analysis & Prevention*, *149*, 105574.

Mattas, K., Albano, G., Dona, R., Galassi, M. C., Suarez-Bertoa, R., Vass, S., & Ciuffo, B. (2022). Driver models for the definition of safety requirements of automated vehicles in international regulations. Application to motorway driving conditions. *Accident Analysis & Prevention*, 174.

Mattas, K., Makridis, M., Botzoris, G., Kriston, A., Minarini, F., Papadopoulos, B., Re, F., Rognelund, G., & Ciuffo, B. (2020, October 5). Fuzzy Surrogate Safety Metrics for real-time assessment of rear-end collision risk. A study based on empirical observations. ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0001457520316146?via%3Dihub

Milakis, D., Snelder, M., Arem, B. van, Wee, B. van, & Homem de Almeida Correia, G. (2017). Development and transport implications of automated vehicles in the Netherlands: scenarios for 2030 and 2050. *European Journal of Transport and Infrastructure Research*, *17*(1).

NHTSA (National Highway Traffic Safety Administration). (2024). Early estimates of motor vehicle traffic fatalities and fatality rate by sub-categories in 2023. *Crash Stats Brief Statistical Summary. Report No. DOT HS*, *813*, 581.

Nister, D., Lee, H.-L., Ng, J., & Wang, Y. (2019, March 1). The Safety Force Field. Nvidia. Retrieved October 7, 2024, from https://www.nvidia.com/content/dam/en-zz/Solutions/self-driving-cars/safety-force-field/the-safety-force-field.pdf

PARTS (Partnership for Analytics Research in Traffic Safety). (2024). PARTS: Market Penetration of Advanced Driver Assistance Systems (ADAS).

Pawlovich, M. D., Li, W., Carriquiry, A., & Welch, T. (2006). Iowa's experience with road diet measures: use of Bayesian approach to assess impacts on crash frequencies and crash rates. *Transportation research record*, *1953*(1), 163-171.

Pek, C., Manzinger, S., Kochii, M., & Althoff, M. (2020). Using online verification to prevent autonomous vehicles from causing accidents. *Nature Machine Intelligence*, 2, 518-528.

Scanlon, J. M., Kusano, K. D., Fraade-Blanar, L. A., McMurry, T. L., Chen, Y. H., & Victor, T. (2024a). Benchmarks for retrospective automated driving system crash rate analysis using police-reported crash data. Traffic Injury Prevention, 25(sup1), S51–S65. https://doi.org/10.1080/15389588.2024.2380522

Scanlon J. M., Teoh E. R., Kidd D. G., Kusano K. D., Bärgman J., Chi-Johnston G., Di Lillo L., Favaro F., Flannagan C., Liers H., Lin B., Lindman, M., McLaughlin, S., Perez, M., Victor, T. (2024b) "RAVE Checklist: Recommendations for Overcoming Challenges in Retrospective Safety Studies of Automated Driving Systems." *Traffic Injury Prevention (Accepted Manuscript).* .https://doi.org/10.1080/15389588.2024.2435620

Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2017). On a Formal Model of Safe and Scalable Self-driving Cars. *ArXiv, abs/1708.06374*.

Taylor, G. (2012). *Loss reserving: an actuarial perspective* (Vol. 21). Springer Science & Business Media.

Teoh ER, Kidd DG. 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.

Tex. Admin. Code. § 218.16

Theofilatos, A., & Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. *Accident Analysis & Prevention*, *72*, 244-256.

U.S. Census Bureau. ACS Demographic and Housing Estimates.

UN Regulation No. 157 Amend.4. (2023, March 7). UNECE. Retrieved September 5, 2024, from https://unece.org/transport/documents/2023/03/standards/un-regulation-no-157-amend4

Yurtsever, E., Lambert, J., Carballo, A., & Takeda, K. (2020). A Survey of Autonomous Driving: Common Practices and Emerging Technologies. IEEE Access, 8, 58443-58469.

Waymo (2024). Safety Impact. https://waymo.com/safety/impact/

Weng, B., Rao, S. J., Deosthale, E., Schnelle, S., & Barickman, F. (2020). Model Predictive Instantaneous Safety Metric for Evaluation of Automated Driving Systems. 2020 *IEEE Intelligent Vehicles Symposium*.