



Waymo Safety Impact Data Hub

Introduction —

[The Waymo Safety Impact Data Hub](#) highlights Waymo's Rider-Only (RO) safety performance by comparing its crash rates to human benchmarks from the areas where Waymo operates. The Data Hub also includes several download files to help researchers reproduce the results or conduct their own research. The Data Hub is envisioned as a living dataset, which will be updated over time with both new data as it is released and new methodologies

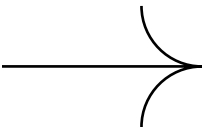
based on new research. This document is a detailed description of new releases of the data. It contains a high level overview of each release, including download links to historical versions of the data. It also contains details that explain and contextualize changes in the data or methods used on the Data Hub. Finally, there is a set of data dictionaries for the downloaded data that describe the content and format of the download data.

Release Notes —

Publish Date	Data End Date (Year Month)	NHTSA SGO Data Publish Date	Waymo Data Hub Files	Release Description
January 6, 2025	2024 September	November 15, 2024	CSV1 - Miles per Geo CSV2 - Crashes with SGO identifier and group membership CSV3 - Collision count, and comparisons to benchmark CSV4 - Geographic Adjustment	<ul style="list-style-type: none"> Updated data using NHTSA SGO through September 2024 (using SGO file released on November 15, 2024) We discovered an error in the CSV2 download released on October 24, 2024. The column “Is Any-Injury-Reported” mistakenly used the old SGO-based method from the September 5, 2024 release and not the updated methodology. As a result, 8 cases were reported as FALSE for the “Is Any-Injury-Reported” column that are now reported as TRUE. This error only applied to the CSV2 column values. The rates previously presented on the Data Hub site and in CSV3 correctly used the updated injury definition. Case 30270-8750 was included in the RO dataset. The “Driver / Operator Type” field for this SGO event states “Other, see Narrative”. The narrative states “At the time of the impact, the Waymo AV's Level 4 ADS was engaged in autonomous mode”, consistent with there being no driver behind the wheel. There are no changes to the methodology used to generate the files or results.
October 24, 2024	2024 July	September 15, 2024	CSV1 - Miles per Geo CSV2 - Crashes with SGO identifier and group membership CSV3 - Collision count, and comparisons to benchmark CSV4 - Geographic Adjustment	<ul style="list-style-type: none"> Updated data using NHTSA SGO through July 2024 (using SGO file released on September 16, 2024) Updated methodology to identify Any-Injury-Reported outcomes in the SGO data. Instead of using the “Highest Injury Severity Alleged” SGO field directly, cases were manually reviewed and reported injuries of unknown severity in the case materials (such as the narrative) were included. See section below for more detailed explanation and analysis of the impact of this change. Added a dynamic adjustment to the human benchmarks used in the Data Hub based on new research that accounts for differences in geographic distribution of driving between the Waymo RO fleet and the human population. See details below. Added a new data download file, CSV4, that contains the spatial distribution of mileage for each geo. This information enables the replication of the dynamic benchmark adjustment reported on the data hub. See the detailed release notes and data dictionary for details. Added an additional column to the “CSV2 - Crashes with SGO identifier and group membership” data file, “Location”, that describes the city in which the crash occurred (e.g., PHOENIX, SAN_FRANCISCO). The ‘Incident Date’ in CSV2 was updated for three events (30270-8157, 30270-7938, 30270-6542), which had mistakenly listed the ‘Notice Received’ date instead of the ‘Incident Date’ in the previous release.

Release Notes —

Publish Date	Data End Date (Year Month)	NHTSA SGO Data Publish Date	Waymo Data Hub Files	Release Description
September 5, 2024	2024 June	August 15, 2024	CSV1 - Miles per Geo CSV2 - Crashes with SGO identifier and group membership CSV3 - Collision count, and comparisons to benchmark	<ul style="list-style-type: none">• First release of Safety Impact Data Hub using NHTSA SGO through June 2024 published on August 15, 2024• Uses human benchmarks reported in Scanlon et al. (2024a) and outcomes from Kusano et al. (2024)



Detailed Release Notes—

October 24, 2024 Release

As the purpose of the Waymo Safety Impact Data Hub is transparent sharing of Waymo's safety performance, the Waymo Safety Research team welcomes feedback from all members of the broader community. We would like to thank Tim Lee and Edouard Hesse for engaging with the data and contacting us, which led to some of the changes in this release. This analysis will continue to be revisited as new data is made available and methodology is developed.

Methodology Change for Any-Injury-Reported Identification in SGO Data

In the initial version of the data hub, the *Any-Injury-Reported* outcome was determined using the value in the "Highest Injury Severity Alleged" field of the SGO data. Any crash that had "Minor", "Moderate", "Serious", or "Fatality" coded in the "Highest Injury Severity Alleged" field was considered an *Any-Injury-Reported* outcome. This classification excluded crashes where the "Highest Injury Severity Alleged" was reported as "Unknown". After reviewing this classification, we identified several "Unknown" injury severity cases where a person alleged or reported an injury but the severity of the injuries was unknown. Although there are nuances with how "Unknown" injuries like these are reported in the human crash benchmarks, in order to follow the suggestions in the RAVE Checklist in Scanlon et al. (2024b) we will now include crashes with "Unknown" injury severity that have alleged injuries of unknown severity in the *Any-Injury-Reported* outcome. This determination will be made using manual review of the information in the SGO narrative.

This methodology change resulted in a small material change in the *Any-Injury-Reported* status of crashes. In total, 4 cases in the previous releases that were not considered *Any-Injury-Reported* are now being included (30270-6149, 30270-7316, 30270-7637, 30270-7945). In addition, one SGO crash (30270-7157) previously had no injuries reported, but a revised report that was submitted in August 2024 (report version 3) indicated injuries of unspecified severity which we now consider an *Any-Injury-Reported* outcome. Of the new SGO crashes that occurred in July 2024 that were added to this data hub, two (2) SGO crashes (30270-8217, 30270-8243) had "Unknown"

maximum severity with reported injuries that are now included in the *Any-Injury-Reported* outcome.

In addition, after reviewing the SGO narratives we have decided to include case 30270-4768 as an *Any-Injury-Reported* crash. According to the SGO narrative, the Waymo vehicle "was in a collision with a minibike whose rider was no longer on it" because "a minibike rider in the group lost control and fell off the minibike a short distance ahead of the Waymo AV in the adjacent lane and the minibike, without a rider, tumbled into the Waymo AV's lane." The SGO "Highest Injury Severity Alleged" for this crash was "No Injuries Reported" because, as noted in the SGO narrative, the collision between the minibike without a rider and the Waymo AV did not cause the minor injuries to the minibike rider. Despite the nature of the crash sequence, we are including this case in the *Any-Injury-Reported* because an injury occurred in a collision sequence involving the Waymo AV, which matches the criteria for reporting in the human crash databases which make up the benchmark.

Spatial Distribution of Mileage Data for Download

Previously, only data was provided for download for the total Waymo driving mileage within the currently deployed geographic regions (CSV1). This update now provides an additional data download file, CSV4, which contains the spatial distribution of driving miles across the entire geographic regions where mileage has been accumulated. The data is provided by S2 cell (geometry level 13), which results in units of approximately 0.5 square miles (1.27 square kilometers). The purpose of providing this data is to enable external research that analyzes Waymo's driving distribution to account for differences in crash risk by where the Waymo fleet was driven.

Dynamic Human Benchmark Adjustment

Based on recently published research (Chen et al., 2024), we are introducing new human benchmarks that are more comparable to the geographic distribution of driving of the Waymo RO fleet. Waymo operates a ride-hailing service where the trips are largely driven by the demand of riders. Because of this demand, ride-hailing trips fall in more

densely populated areas compared to the entire human driving population. The human crash benchmarks reported by Scanlon et al. (2024a) and reported previously on this Data Hub used this entire human driving population. We have heard feedback from many members of the public and researchers wondering how the unique driving distribution of a ride-hailing service might affect crash rates. In a new study by Chen et al. (2024), the effect of spatial distribution on crash risk was analyzed. A methodology was then developed to adjust benchmarks to account for the unique driving distribution of the Waymo driving. The result of the reweighting method is human benchmarks that are more representative of the areas of the city Waymo drives in the most, which improves data alignment between the Waymo and human crash data. Achieving the best possible data alignment, given the limitations of the available data, are part of the newly published Retrospective Automated Vehicle Evaluation (RAVE) best practices (Scanlon et al., 2024b).

The table below shows the overall and dynamically adjusted human benchmarks using the mileage blended benchmarks for the three main outcomes shown on the Data Hub. The results show that the Waymo RO service drives more often in areas with higher crash rates than the overall human benchmark, resulting in the dynamic human benchmark being 18% to 29% higher than the overall benchmark.

Comparison of Overall and Dynamic Human Benchmark Crashes Incidents Per Million Miles (IPMM) (Data through July 2024, Mileage Blended Benchmark).

Outcome	Overall Human Benchmark (IPMM)	Dynamic Human Benchmark (IPMM)	Ratio Dynamic to Overall Benchmark
Airbag Deployment	1.45	1.71	1.18
Any-Injury-Reported	2.97	3.84	1.29
Police-Reported	4.72	5.77	1.22

Because we believe the dynamic benchmark improves the alignment between the human and AV data, we will now use the dynamic human benchmarks in the results shown on the main Data Hub website for the Police-Reported, Any-Injury-Reported, and Airbag Deployment outcomes. We will continue to report the comparisons using the overall population human benchmarks previously used in the download section of the data (CSV3). The comparisons using the dynamic benchmark will be labeled with “dynamic” in the “Benchmark Comparison” column of CSV3 (e.g., “Airbag Deployment (Dynamic)”) while the comparison using the overall human benchmark will have no additional label (e.g., “Airbag Deployment”).

References

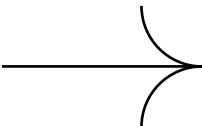
Chen, Y., Scanlon, J. M., Kusano, K. D., McMurry, T. L., Victor, T. (2024). Dynamic Benchmarks: Spatial and Temporal Alignment for ADS Performance Evaluation. arXiv preprint arXiv:2410.08903.

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.

Kusano, K. D., Scanlon, J. M., Chen, Y. H., McMurry, T. L., Chen, R., Gode, T., & Victor, T. (2023). Comparison of Waymo Rider-only crash data to human benchmarks at 7.1 million miles. Traffic Injury Prevention (In Press).

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 (In Press).

Scanlon, J. M., Teoh, E. R., Kidd, D. G., Kusano, K. D., Bärgrman, J., Chi-Johnston, G., ... & Victor, T. (2024b). RAVE Checklist: Recommendations for Overcoming Challenges in Retrospective Safety Studies of Automated Driving Systems. arXiv preprint arXiv:2408.07758.



Data Dictionary—

This section contains a description of the data elements in each file.

FILE

CSV1 - Miles per Geo

Data Element Name	Definition	Data Type	Allowable Values	Release Created	Release Updated
Ops Depot	Geographic location. Note this is the same field as “Location” in other CSVs	String	<ul style="list-style-type: none">• LOS_ANGELES• PHOENIX• SAN_FRANCISCO• AUSTIN	September 7, 2024	N/A
Waymo RO Miles (Millions)	Waymo Rider-Only (RO) miles traveled in millions of miles.	Numeric	Positive number	September 7, 2024	N/A

FILE

CSV2 - Crashes with SGO identifier and group membership

Data Element Name	Definition	Data Type	Allowable Values	Release Created	Release Updated
SGO Report ID	NHTSA Standing General Order Report Identifier	String	A two-part integer identified separated by a dash, where the first part is the company code and the second part is the event identifier (e.g., 30270-8266)	September 7, 2024	N/A
Year Month	Calendar year and month that crash occurred.	Integer	YYYYMM where YYYY is the century and year and MM is the calendar month.	September 7, 2024	N/A
Is NHTSA Reportable In-Transport	Indication whether this SGO crash was considered in the Any Property Damage or Injury outcome, which includes a requirement to not only be SGO-reported, but also for the Waymo vehicle to be in-transport (i.e., not parked) and impacted in the crashes. See Kusano et al. (2024) for details.	Boolean	TRUE or FALSE as a string	September 7, 2024	N/A
Is NHTSA Reportable In-Transport Delta-V Less than 1 MPH	Indicates whether this SGO crash had a delta-V less than 1 MPH, along with the in-transport and impacted requirements. This result is used to calculate the proportion of crashes with a delta-V less than 1 MPH on the Data Hub.	Boolean	TRUE or FALSE as a string	September 7, 2024	N/A
Is Police-Reported	Indication whether this SGO crash was considered in the Police-Reported outcome. See Kusano et al. (2024) for details.	Boolean	TRUE or FALSE as a string	September 7, 2024	N/A

Is Any-Injury-Reported	Indication whether this SGO crash was considered in the Any-Injury-Reported outcome. See Kusano et al. (2024) and these release notes for details.	Boolean	TRUE or FALSE as a string	September 7, 2024	N/A
Is Airbag Deployment	Indication whether this SGO crash was considered in the Airbag Deployment outcome. See Kusano et al. (2024) and these release notes for details.	Boolean	TRUE or FALSE as a string	September 7, 2024	N/A
Incident Date	The date the crash occurred.	Date	MM/DD/YYYY where MM is the month, DD is the day, and YYYY is the year.	September 7, 2024	N/A
Location Address / Description	The description of the location of the crash.	String	Examples are descriptions of cross streets, addresses, or other locations.	September 7, 2024	N/A
Zip Code	The zip code of the location of the crash.	Integer	5-digit zip code	September 7, 2024	N/A

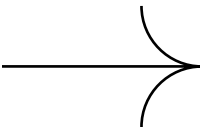
FILE

CSV3 - Collision count, and comparisons to benchmark

Data Element Name	Definition	Data Type	Allowable Values	Release Created	Release Updated
Outcome	Type of crash outcome	String	<ul style="list-style-type: none"> Airbag Deployment Any Injury Police Report SGO In Transport and Impacted SGO In Transport and Impacted Excluding Low Delta-V 	September 7, 2024	N/A
Benchmark Comparison	The benchmark that is used in the comparison	String	<ul style="list-style-type: none"> Airbag Deployment Airbag Deployment (Dynamic) Any-injury-reported Any-injury-reported (Dynamic) Any-injury-reported (observed) Any-injury-reported (Observed, Dynamic) Police-reported Police-reported (Dynamic) Any Property Damage - Blincoe Adjusted Any Property Damage - Blincoe Adjusted (Dynamic) 	September 7, 2024	October 24, 2024

Location	Geographic location. In addition to the geographic locations used in CSV1, this CSV3 Location includes some combinations of geographies for comparisons that are shown on the Data Hub.	String	<ul style="list-style-type: none"> • All Locations (mileage blended) • LOS_ANGELES • PHOENIX • Phoenix and San Francisco • SAN_FRANCISCO • AUSTIN 	September 7, 2024	October 24, 2024
Waymo Count	The number of Waymo RO crashes for this crash outcome.	Integer	Positive integer including zero	September 7, 2024	N/A
Delta V Less Than 1 Mph Percent	The percentage of Waymo SGO crashes in this outcome category with a delta-V less than 1 mph.	Numeric	Decimal between 0 and 1	September 7, 2024	N/A
Waymo IPMM	Waymo RO Incidents Per Million Miles (IPMM) in this outcome.	Numeric	Positive number	September 7, 2024	N/A
Waymo IPMM CI Lower	Waymo RO IPMM lower 95% confidence interval calculated using a Poisson Exact method. See Kusano et al. (2024) for calculation details.	Numeric	Positive number	September 7, 2024	N/A
Waymo IPMM CI Upper	Waymo RO IPMM upper 95% confidence interval calculated using a Poisson Exact method. See Kusano et al. (2024) for calculation details.	Numeric	Positive number	September 7, 2024	N/A
Benchmark IPMM	Human benchmark Incidents Per Million Miles (IPMM) in this outcome.	Numeric	Positive number	September 7, 2024	N/A
Benchmark IPMM CI Lower	Human Benchmark IPMM lower 95% confidence interval calculated using a Poisson Exact method. See Kusano et al. (2024) for calculation details.	Numeric	Positive number	September 7, 2024	October 24, 2024 (added missing CI in name)
Benchmark IPMM CI Upper	Human Benchmark IPMM upper 95% confidence interval calculated using a Poisson Exact method. See Kusano et al. (2024) for calculation details.	Numeric	Positive number	September 7, 2024	N/A
Predicted Benchmark Count	The average crash count predicted at the benchmark crash rate over the miles driven by Waymo.	Numeric	Positive number	September 7, 2024	N/A
Waymo Count Reduction	The Predicted Benchmark Count minus the Waymo Count, where a positive number indicates the Waymo RO service had few crashes than the average benchmark over the same distance.	Numeric	Real number	September 7, 2024	N/A

Waymo Percent Reduction	The percent reduction in crash rate from the human benchmark to the Waymo RO service. A negative number indicates the Waymo RO services had a lower crash rate than the benchmark. Note the percent reduction is the rate ratio minus 1.	Numeric	-1 to positive infinity	September 7, 2024	N/A
Waymo Percent Reduction CI Lower	The Waymo RO percent reduction lower 95% confidence interval calculated using a rate ratio method described by Nelson (1970). See Kusano et al. (2024) for calculation details.	Numeric	-1 to positive infinity	September 7, 2024	N/A
Waymo Percent Reduction CI Upper	The Waymo RO percent reduction upper 95% confidence interval calculated using a rate ratio method described by Nelson (1970). See Kusano et al. (2024) for calculation details.	Numeric	-1 to positive infinity	September 7, 2024	N/A
Waymo Rate Ratio	The ratio of the Waymo and benchmark crash rates.	Numeric	0 to positive infinity	September 7, 2024	N/A
Waymo Rate Ratio CI Lower	The rate ratio lower 95% confidence interval using a method described by Nelson (1970). See Kusano et al. (2024) for calculation details.	Numeric	0 to positive infinity	September 7, 2024	N/A
Waymo Rate Ratio CI Upper	The rate ratio upper 95% confidence interval using a method described by Nelson (1970). See Kusano et al. (2024) for calculation details.	Numeric	0 to positive infinity	September 7, 2024	N/A



Data Element Name	Definition	Data Type	Allowable Values	Release Created	Release Updated
Location	Geographic location	String	<ul style="list-style-type: none"> • LOS_ANGELES • PHOENIX • SAN_FRANCISCO 	October 24, 2024	N/A
S2 Cell	Geographic location specified by the S2 Cell system at zoom level 13.	String	A valid S2 cell, according to the S2 specification .	October 24, 2024	N/A
Outcome	Crash outcome group	String	<ul style="list-style-type: none"> • Airbag - A crash with an airbag deployment • Observed_any_injury - A crash with police-reported injury of any severity. • Blincoe_any_injury - An estimate of crashes with any injury reported using an underreporting correction from Blincoe et al. (2023) • Police_reported - police-reported crashes. • Blincoe - an estimate of crashes with any property damage or injury reported using an underreporting correction from Blincoe et al. (2023) 	October 24, 2024	N/A
Benchmark Crash Count	Count of crashes in the human benchmark crash data source corresponding to this Outcome.	Numeric	Positive real number	October 24, 2024	N/A
HPMS Yearly Vehicle Miles Traveled	The annual human Vehicle Miles Traveled (VMT) in the S2 cell as reported in the Highway Performance Monitoring System (HPMS). This value is replicated across Outcome levels (i.e., each S2 Cell should have one VMT value).	Numeric	Positive real number	October 24, 2024	N/A
Waymo RO Miles	The number of miles traveled by the Waymo RO service in the S2 cell during the reporting period. This value is replicated across Outcome levels (i.e., each S2 Cell should have one RO Miles value).	Numeric	Positive real number	October 24, 2024	N/A

References

Kusano, K. D., Scanlon, J. M., Chen, Y. H., McMurry, T. L., Chen, R., Gode, T., & Victor, T. (2023). Comparison of Waymo Rider-only crash data to human benchmarks at 7.1 million miles. *Traffic Injury Prevention* (In Press).

Nelson, W. (1970). Confidence intervals for the ratio of two Poisson means and Poisson predictor intervals. *IEEE Transactions on Reliability*, 19(2), 42-49.