Establishing a Crash Rate Benchmark Using Large-Scale Naturalistic Human Ridehail Data

Prepared by

Carol Flannagan*, Andrew Leslie*, Raymond Kiefer†, Scott Bogard*, Geoff Chi-Johnston‡, Laura Freeman‡, Rayman Huang‡, David Walsh‡, Antony Joseph‡

* University of Michigan Transportation Research Institute (UMTRI)
† General Motors LLC
‡ Cruise LLC

June 2023

Table of Contents

0. Executive Summary

This paper presents a study of human driving performance for the targeted geographic area of San Francisco. The goal of the study was to generate a crash rate estimate that could be used as a human benchmark representing the crash rate for ridehail drivers driving in a lower-speed and dense urban driving environment.

The study represents an unprecedented large-scale naturalistic ridehail data collection effort over a 2-year period from 2016 to 2018: a collaboration that took place between General Motors (GM), Cruise LLC, the University of Michigan Transportation Research Institute (UMTRI), and the Virginia Tech Transportation Institute (VTTI).

The specific Operational Design Domain (ODD) designated in this study is defined as a geofenced area covering the entirety of the city of San Francisco excluding select high speed roads (e.g., posted speeds of 35 mph or less).

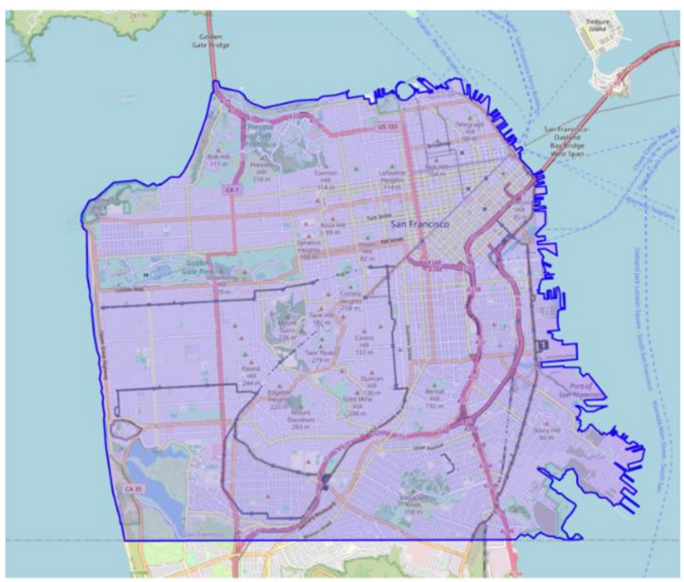

Figure 1 *– Depiction of the Geofence Area - the city of San Francisco*

The study gathered data from drivers renting vehicles from Maven,[1] primarily for ridehail driving (estimated to be 80% of the total driving activity observed). Accordingly, the safety benchmarking approach described here incorporates the unique characteristics of ridehail drivers (e.g., predominantly younger male drivers with long driving hours), ridehail driving patterns (e.g., trips with origins and destinations, and circuitous routes), and crashes associated with lower-speed urban ridehail driving.

[1] A former subsidiary of GM

The study leveraged two concurrent naturalistic driving studies that used different methods to derive a statistically robust measurement of the human ridehail crash rate in the designated ODD. This method combined the precision of instrumented vehicle data collected by VTTI with the scale of telematic data collected by GM, enhanced with associated insurance claims and analyzed by UMTRI. Notably, mileage collected by a shared UMTRI-VTTI fleet allowed for a detailed comparison of the respective crash-detection mechanisms, which enabled production of a single combined estimate of the human ridehail crash rate in the designated ODD.

Here is a summary of the data output from each fleet:

	UMTRI Fleet	**VTTI Fleet**	**Shared Fleet**[2]
Fleet Size	1,149 vehicles	89 vehicles	59 vehicles
ODD Miles Generated	5,424,077 miles	415,901 miles	228,213 miles
Crash Detection Method	1. Data surfaced through Event Data Record (EDR) system 2. Insurance claims	1. Manual reporting with kinetic data confirmation 2. Human review of events detected by specialized on-board instruments	Combined UMTRI and VTTI methods
ODD Crashes	196.19 crashes[3]	21 crashes	15 crashes • 3 detected by UMTRI only • 7 detected by VTTI only • 5 detected by both
ODD Crash Mileage	27,647 miles per crash	19,805 miles per crash	
ODD Crash per Million Miles	36.2 crashes per million miles	50.5 crashes per million miles	

Table 1 *– Summary of mileage, methodology, and crash count data across the UMTRI, VTTI, and UMTRI-VTTI shared fleets*

[2] The data presented in this column represents a subset in the UMTRI Fleet and VTTI Fleet columns.
[3] Derivation of the fractional output listed here is described in detail in 2.3.3.1 UMTRI Crash Measurement Methodology and 3.2.1 UMTRI Crash Count.

The total mileage of UMTRI Fleet and VTTI Fleet with deduplication of shared fleet miles produced a total of **5,611,765 ODD miles** observed in this study.

The observed UMTRI and VTTI ODD driving mileage and crash counts were analyzed using a Bayesian Fusion statistical model[4] to correct for crashes missed by either data collection approach.

The output generated a single estimate of the human ridehail crash rate:

1 crash in 15,414.4 ODD driving miles, or **64.9 crashes per million ODD miles**.

1. Introduction

1.1 Goals of the Study

A significant challenge in understanding human driving performance within an ODD is that different driving environments (e.g., limited access highways vs urban streets) result in different crash rates. Thus, it is important to compare crash rates from driving in similar environments (road type, time of day, etc.). While publicly available national crash datasets have detail on the types of locations of crashes, datasets on vehicle miles traveled do not. Moreover, national crash datasets are limited to police-reported crashes, which include only the more damaging or injurious crashes. Given these data issues, publicly available national datasets cannot produce human crash rate estimates that are appropriate to the urban ridehail driving environment.

This paper presents a study of human driving performance by ridehail drivers operating in San Francisco. The goal of the study was to generate a crash rate estimate that could be used as a human benchmark representing the crash rate for ridehail drivers driving in a low-speed and dense urban driving environment. Moreover, this environment was specifically limited to driving in the initial target San Francisco-based ODD of Cruise vehicles to further refine the relevance of the estimate.

1.2 Urban Ridehail Driving

The study represents an unprecedented large-scale naturalistic ridehail data collection effort over a 2-year period from 2016 to 2018: a collaboration that took place between General Motors (GM), Cruise LLC, the University of Michigan Transportation Research Institute (UMTRI), and the Virginia Tech Transportation Institute (VTTI).

---

[4] To correct for crashes missed by either UMTRI or VTTI, a statistical model was developed by Cruise to combine the individual crash detection methods into a single estimate of the human ridehail crash rate. This model assumes that crashes follow a Poisson distribution, and divides the observed mileage and crash counts into unshared UMTRI crashes, shared crashes, and unshared VTTI crashes. See more in 3.3.2 Fused Crash Rate.

The study gathered data from drivers who engaged primarily in ridehail (estimated to be 80% of total driving activity observed). The majority of drivers rented their vehicles from Maven,[5] with the remainder using their own vehicles for ridehail. Accordingly, the safety benchmarking approach described here incorporates the unique characteristics of ridehail drivers (e.g., predominantly younger male drivers with long driving hours), ridehail driving patterns (e.g., trips with origins and destinations, and circuitous routes), and crashes associated with lower-speed urban ridehail driving.

1.2.1 Ridehail Driver Demographics

Ridehail drivers are more likely to be male and/or young compared to the average driver in the United States. Although driver demographics of the San Francisco UMTRI fleet was not known, aggregate data was available for the larger set of 8,583 Maven drivers across 9 urban areas in the United States. Table 2 shows the distribution of driver ages for the broad set of Maven drivers, as well as the targeted VTTI San Francisco fleet.

Age Range	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70+**
Percent of Maven Drivers	32%	31%	21%	12%	4%	0.5%
Percent of VTTI-SF Drivers	28%	26%	23%	17%	5%	1%

***Table 2** – Distribution of driver age for an aggregate Maven driver pool (8,583 drivers over 9 urban regions) and the observed VTTI San Francisco Fleet*

For both the Maven and VTTI fleets, the driver populations skewed towards younger drivers, with 63% of Maven drivers and 54% of VTTI drivers under 40 years old. Additionally, the VTTI-SF fleet consisted of 72% male drivers, which is likely to be representative of the broader UMTRI fleet.

Thus, the human ridehail crash rate benchmark established in this study measures the driving performance of a specific demographic of ridehail drivers in San Francisco.

1.2.2 Ridehail Driving Patterns

When engaged in ridehail driving, human driving patterns (i.e., routes, road types) are often determined by their customer’s desired pick-up and drop-off locations. Ridehail routes often involve U-turns and other unusual maneuvers to find a safe drop-off point or pick someone up in an opposing direction. These routes also tend to include city streets and pass high-traffic areas to pick-up or drop-off customers at popular ridehail locations, such as restaurants and downtown areas.

The speed distribution for the UMTRI ridehail fleet within the geofenced map area is presented below in Figure 2.

---

[5] A former subsidiary of GM

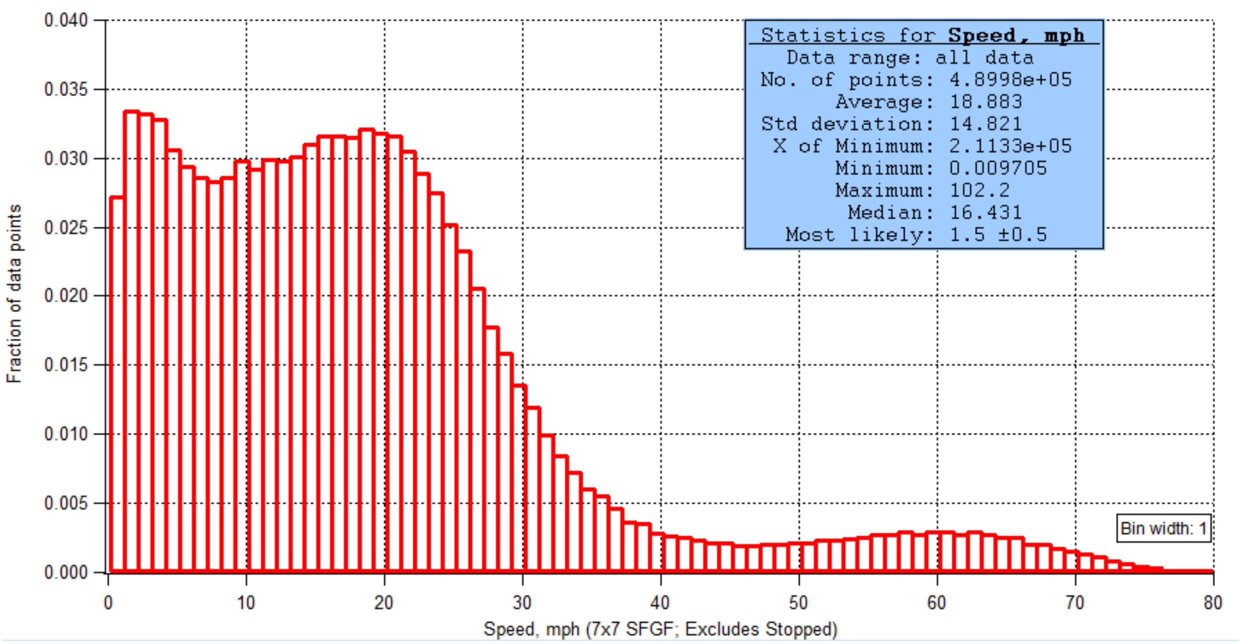

***Figure 2** – Travel speeds for the Maven fleet within the Geofence Area (excluding stopped time)*

This figure shows that human driver travel speeds are quite low, with a median speed of 16.4 mph. Note that these travel speeds are further reduced when considering the additional ODD road type constraints described later in this paper.

Table 3 also provides insight into unique ridehail driving patterns in the form of skewed distribution of road types towards surface streets. Only 22.3% of the study fleet's driving miles were on highways (i.e., Motorway or Trunk).

OpenStreetMap Road Type	Percent of Fleet Driving
Motorway	17.8%
Trunk	4.5%
Primary	22.4%
Secondary	16.9%
Tertiary	12.0%
Residential	19.3%
Ramp or Link	7.1%

***Table 3** – Distribution of Maven fleet driving by OpenStreetMap road type*

Finally, ridehail drivers tend to have very high weekly mileages compared to the general population of drivers. The average weekly mileage in the UMTRI fleet was 689 miles, with Maven's estimate of approximately 80% of miles involving ridehail activities.

2. Human Ridehail Safety Benchmarking Methodology

2.1 Overview

The two datasets were used in combination to develop an estimate of the crash rate for human ridehail driving. Importantly, the data samples for the two studies were partially overlapping, enabling direct comparison of the capabilities and sensitivities of the two data collection approaches for detecting crashes of various severities. To support estimating human ridehail crash rates, both approaches collected data related to crash counts and miles traveled. A detailed comparison of the respective crash data detection mechanisms enabled the production of a single estimate of the human ridehail crash rate using a Bayesian Fusion statistical model. The results of the study indicate that the combination of UMTRI and VTTI datasets substantially mitigated the risk of missing human ridehail crashes in its derivation of the estimate.

2.2 Operational Design Domain

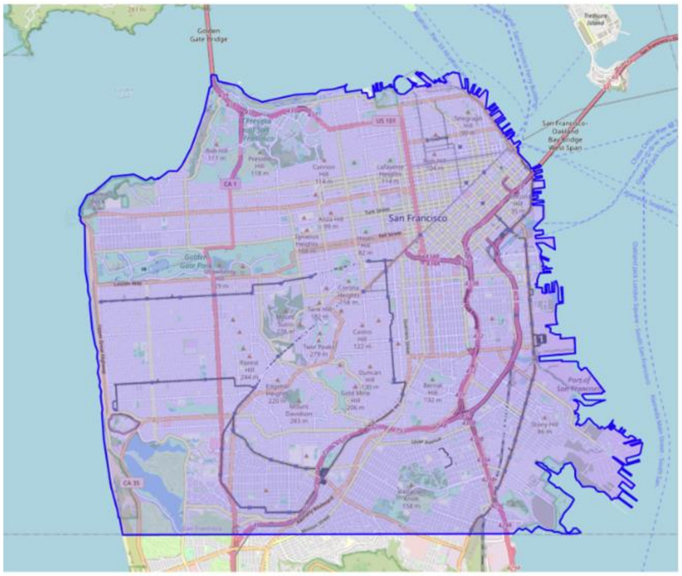

Figure 3 *- Depiction of the Geofence Area - the city of San Francisco*

The specific Operational Design Domain (ODD) designated in this study is defined as roads within the Geofenced Area of San Francisco illustrated above, with posted speed limits of 35 mph or less; and also excludes select high-traffic roads, as specified in 2.3.2 Mileage Measurement Methodology.

The data collected in this study was filtered to identify mileage and crash counts that took place in the specified ODD. While the exact methodology for establishing ODD driving differed between UMTRI and VTTI due to the types of data available to the respective studies, the core approach was to count miles driven on ODD-eligible roads, which included both trips that were entirely within the ODD and trips with only a portion of the miles in the ODD.

2.3 Comparison of UMTRI and VTTI Methods of Data Collection

2.3.1 Fleet and Data Collection Instrumentation

2.3.1.1 UMTRI Fleet and Instruments

The UMTRI San Francisco-based ridehail fleet consisted of 1,149 GM Model Year 2015-2017 vehicles[6] owned by the Maven subsidiary of GM. These Maven vehicles were rented to users, primarily (but not entirely) for ridehail activities. A total of 4,464 drivers were observed. The UMTRI data collection effort extended from November 2016 to December 2018.

All data addressing miles driven and crashes observed for the UMTRI fleet were collected using GM's telematic-based OnStar data collection capabilities. This included Event Data Recorder (EDR) data for crashes with speed greater than or equal to 5 mph. This telematic data was supplemented with insurance claims provided by Maven for the study fleet.

During the trips, UMTRI study vehicles reported GPS (latitude, longitude, altitude, heading, date, and time), speed, steering wheel angle, and longitudinal acceleration at 1 Hz. This data was important for defining ODD driving, measuring distance traveled, and identifying the routes traveled.

A set of OnStar "Smart Driver" events – a feature available to all OnStar subscribers – that were collected in the UMTRI fleet was defined by certain kinematic triggers intended to indicate driving situations predicting crash risk or associated with safety-critical situations. For each triggered event location, the speed, acceleration, seat belt status, and odometer data was reported. The analysis used the odometer measure as a ground-truth measure for ODD distance traveled, which was critical for enhancing the accuracy of mileage estimation.

[6] This fleet of Model Year 2015-2017 vehicles did not have the range of advanced driver assistance systems (ADAS) that are offered on newer vehicles. As a result, these estimated crash rates do not account for the crash avoidance benefits of those safety systems.

2.3.1.2 VTTI Fleet and Instruments

The VTTI San Francisco-based ridehail fleet consisted of 61 Maven GM Model Year 2015-2016 vehicles. In addition, the VTTI fleet included 28 privately-owned Model Year 2008-2018 vehicles across a range of Make-Model combinations. As with the Maven vehicles, these vehicles were also used primarily for ridehail activities. VTTI fleet data collection ran from November 2016 to July 2018, ending five months prior to the end of the UMTRI fleet data collection.

VTTI gathered data via an on-board Data Acquisition System (DAS) installed on the vehicle. Additionally, drivers were required to provide informed consent to participate in data collection; and, consequently, driving data from unconsenting drivers was not included in this study.

The VTTI DAS collected data continuously and automatically from key-on to key-off, and included a cellular modem that was used for periodic DAS health checks.[7]

The data gathered included GPS (latitude, longitude, altitude, heading, date, time) at 1 Hz; acceleration (x, y, z axes) and gyro (x, y, z axes) sensor data at 10 Hz; and range and relative speed to detected forward targets via a forward-looking MobileEye camera at 11.5 Hz. Additionally, on-board 1080 HD cameras were used to gather video streams of the driver's face at 15 Hz, as well as in views surrounding the vehicle (forward, left, right, rear left, rear right). Finally, where possible, the vehicle's Controller Area Network (CAN) data was used for gathering accelerator position, brake pedal activation, and/or speed data.

2.3.1.3 Shared Fleet

A fleet subset consisting of 59 vehicles, called the UMTRI-VTTI Shared Fleet, was observed by both the UMTRI and VTTI data collection methods. This enabled comparison of the results of both mileage-counting systems and crash-identification approaches. This fleet was also critical to the Bayesian Fusion estimation method used to establish the singular human ridehail crash rate benchmark.

2.3.2 Mileage Measurement Methodology

2.3.2.1 UMTRI Mileage Measurement Methodology

UMTRI data collection included both 1 Hz speed data and periodic odometer data. For the overall mileage observed over the entire course of the study, the key-on and key-off odometer readings were used. However, for mileages in specific subsets of the data, it was necessary to calculate intermediate mileages for partial trips, i.e., if the trip did not wholly take place in the ODD. This study considered the odometer readings as the gold standard for mileage, and used 1 Hz speed data only where necessary to calculate partial mileage.

---

[7] In rare cases, severe crash impacts may prevent DAS from saving crash data after the moment immediately preceding a crash.

The UMTRI data collection used OpenStreetMap (OSM) as the base map of roads. This map was supplemented with a public dataset of posted speed limits in an effort to fill known gaps in the OSM dataset.

The UMTRI dataset filtered the Road Type component of the ODD using the following criteria:

Included	**Excluded**
Roads with posted speed limit of 35 mph or less	Roads with posted speed limit of more than 35 mph
Roads with OSM road type: ● Primary ● Secondary ● Tertiary ● Residential	Roads with OSM road type: ● Motorway ● Motorway Link ● Trunk ● Trunk Link
	Specific roads labeled as Function Class 2 roads in other mapping systems

Table 4 *– UMTRI ODD filter of OpenStreetMap map database road types*

Travel on roads in the Geofence Area satisfying these criteria were defined as ODD driving in the UMTRI dataset.

Three measures of distance within the Geofence Area were used to determine mileage:

1. **Odometer Estimate:** Trips exclusively within the Geofence Area had start and end odometer values. Trips with transitions in or out of the Geofence Area used an odometer gain value to correct distance from integrated speed on portions of the trip not covered by actual odometer values.
2. **Mapped Estimate**: Sum of estimated distances per trip from integrated speed (speed multiplied by time) for driving within the Geofence Area that was map-matched using OSM.
3. **Mapped Estimate Filtered for Road Type**: Sum of all estimated distances per trip for driving on public roads using the road type and posted speed filters.

A correction factor of 0.903 was derived by using 1.67 million miles of travel across 247,294 trips that were entirely within the Geofence Area. This correction estimate was multiplied with the difference between the **Odometer Estimate** and **Mapped Estimate** to account for missed distance due to integrated speed and mapping errors, and then added to the **Mapped Estimate Filtered for Road Type**, to produce the final calculated mileage within ODD.

The final calculation for mileage estimate was generated as the following:

(**Odometer Estimate** - **Mapped Estimate**)(0.903 correction factor) + **Mapped Estimate Filtered for Road Type** = Final UMTRI Mileage

2.3.2.2 VTTI Mileage Measurement Methodology

VTTI fleet driving mileages were estimated using CAN speed data when available, and otherwise used 1 Hz GPS data. The available speed measure was integrated to estimate distance traveled. Mileage was not recorded when the DAS was not recording (e.g., first minute of trip, during data downloads, during repairs); nor was it counted for any driving mileage associated with non-consenting drivers.

VTTI used the Navteq / Here map database to produce mileage according to the same ODD constraints. In the Navteq / Here map database, the road type components were filtered using the following criteria:

Included	**Excluded**
Functional Class 1 and 2 roads	Functional Class 3, 4, or 5 roads (i.e., connections between higher functional class, arterials, neighborhoods, and smaller roads) Highways and freeways
Travel speed of 40 mph or less	Travel speed greater than 40 mph
	Controlled-access roadway or on- / off- ramp

Table 5 – *VTTI ODD filter of Navteq / Here map database road types*

Three measures of distance within the Geofence Area were used to determine mileage:

1. **Speed Estimate**: 1 Hz vehicle speed integrated (via CAN or GPS)
2. **Mapped Estimate**: Integrated speed distance for travel that was map-matched driving on public and private areas
3. **Mapped Estimate Filtered for Road Type**: Integrated speed distance for travel that was map-matched driving filtered by road type.

To account for non-mapped driving, a correction factor was applied to the proportion of unmapped driving within the ODD. The correction was based on the proportion of all driving to all driving that was mapped (i.e., ratio of **Speed Estimate** to **Mapped Estimate**).

The VTTI mileage was derived by multiplying the estimated ODD mileage by the correction factor to account for mapping errors. The final calculation for mileage estimate was generated as the following:

(**Speed Estimate** / **Mapped Estimate**)(**Mapped Estimate Filtered for Road Type**)
= Final VTTI Mileage

2.3.2.3 Shared Fleet Mileage Comparison

Within the shared fleet, the UMTRI fleet reported 228,213 miles and the VTTI fleet reported 189,415 miles within the ODD. In other words, the VTTI methodology appeared to underreport the miles reported by UMTRI methodology, reporting only 83% of the UMTRI mileage. Accordingly, all VTTI miles were corrected using this estimate based off of the UMTRI odometry reading, as described in detail in 3.1 Human Ridehail Mileage in ODD.

Further investigation indicated this underestimation was likely due to the following aspects of the VTTI mileage estimation approach:

1. The integrated-speed method generally results in lower estimates than an odometer-based approach
2. Unconsented miles from VTTI test participants were removed from consideration
3. The DAS does not always collect data (e.g., at start or end of trip) resulting in mileage dropout.

Of these sources of underestimation, unconsented miles in the shared fleet, which were estimated to be approximately 4% of the total mileage, were expected to account for only a small portion of the discrepancy. Most of the difference in total miles is attributed to DAS dropout and underestimation of the integrated-speed method compared to odometer measurements.

2.3.3 Crash Count Methodology

2.3.3.1 UMTRI Crash Count Methodology

UMTRI fleet crashes were captured via three streams:

1. Automatic Advanced Collision Notification (AACN) event data recorder (EDR) events
2. “Low-level” Event Data Recorder (LLEDR) events not meeting AACN criteria
3. Insurance Claims

AACN and LLEDR events were captured automatically by the vehicle to create crash event records that could be accessed telematically, whereas insurance claim data included crashes of all severity levels including those which were not captured by AACN and LLEDR events.

OnStar is automatically notified of AACN crashes, which are severe crashes meeting specific criteria such as directionally-specific delta-V[8] thresholds and airbag deployment. The resulting AACN event reports:

1. Crash impact severity as measured by delta-V
2. Crash direction as measured by the Principle Direction of Force, or PDOF
3. Crash GPS location
4. Crash date and time

[8] Delta-Velocity (more commonly referred to as delta-V) corresponds to the change in velocity that the vehicle sensing system experienced during the crash event.

Delta-V is a particularly important measure of crash severity and can be directly linked to injury risk. AACN crashes also triggered an EDR event report with some exceptions due to data collection limitations.

In this report, crashes triggering either AACN or LLEDR events are collectively referred to as Crash Data Record (CDR) events. The CDR data was captured telematically, and the trigger threshold was lowered to automatically capture crash events exceeding 5 mph delta-V in any crash direction. These crash events triggered an EDR event report, which included delta-V and PDOF measures.

CDR events were tracked using counters onboard each vehicle. Tasks within the OnStar module monitored these counters every minute, and upon a counter change, pushed the counter data to the servers in the GM OnStar back office. Additionally, these counters were sent to the back office servers on a weekly basis regardless of counter changes as a further check on any counter status changes. Since many of these production vehicles in the study were already in the field at the start of data collection, the early counter history established a baseline from which new crash events during the study data collection could be identified.

In addition to the automatically-recorded CDR crashes, UMTRI was provided with redacted excerpts from insurance claims filed for the study vehicles. These claims, which sometimes included empty data fields, provided data on the basic crash type (e.g., struck object crash), the date of the crash, and an event narrative. For lower severity crashes, the event narrative was the only potential source of crash time and location used for determining whether the crash met ODD requirements. The event narrative data was used to categorize the crashes according to the accident type coding used in the NHTSA Fatality Analysis Reporting System (FARS)[9]. Insurance claims were often made for relatively minor crash events, and could be initiated by a customer or manager. Since these vehicles were not owned by the drivers, it is likely that a claim was made for crashes leading to any damage.

Since the three crash data sources (AACN, LLEDR, and insurance claim) could identify the same crashes, further efforts were required to identify the locations and times of crashes to avoid double-counting crashes, and also to ensure that only crashes within the ODD were counted.

AACN records included an automated call to OnStar which included precise GPS location and time.

LLEDR events did not include an automated call with crash location, and in some cases, LLEDR records did not enable pinpointing the crash location due to imprecise reporting of LLEDR time. These cases required exact crash time and location to be determined via crash-related event

[9] National Center for Statistics and Analysis. (2022, March). Fatality Analysis Reporting System analytical user's manual, 1975 - 2020 (Report No. HOT HS 813 254). National Highway Traffic Safety Administration.

signals in the kinematic data. In instances where this data was absent, there was uncertainty whether the crash took place in the designated ODD. The specifics of how low confidence LLEDR crash counts were calculated are specified in 3.2.1 UMTRI Crash Count.

Any insurance claims that were located could be linked to CDR events by identifying crash records that occurred close in time. For the remaining CDR events, any insurance narrative associated with the same driver was further examined for evidence of similarity of the crash. After completing this process, 18 of 23 AACN events (78%) and 60 of 123 LLEDR events (49%) in the San Francisco region were associated with an insurance claim.

For crashes detected only through insurance claims (and not associated with CDR events), the reported crash date, crash narrative, and city were examined to give an initial estimate of crash time and location. Delta-V, which was not included in the insurance claim data, could be inferred for these crashes to be less than 5 mph since no LLEDR was recorded by the production-based crash sensors. GPS location and vehicle kinematic data could reveal common crash signatures, such as excessive deceleration, trips ending on a road shoulder, or long periods of being stopped in an intersection or lane, to identify potential crash behavior. A human inspection of insurance data could produce a confidence crash location rating (e.g., no confidence, low confidence, high confidence) along with other details. Of the insurance-only crashes, only 9% could not be located with the available crash narrative.

2.3.3.2 VTTI Crash Count Methodology

VTTI identified crashes through two key mechanisms. First, crashes were manually reported directly to VTTI, and then confirmed in the kinematic data. Second, potential crashes were identified by examining cases where one or more trigger threshold metrics were exceeded, which triggered further human video review to verify the crash and assign a VTTI-defined crash severity level. These metrics were based on longitudinal deceleration / acceleration, yaw, lateral acceleration, or ABS/ESC/TC[10] activation data. Crashes meeting VTTI-defined “most severe” (Level 1), “police-reportable” (Level 2), or “minor” (Level 3) crash levels were used here for crash counting purposes. Crashes less severe than Level 3, e.g., contact with soft poles, curb strikes, and potholes, were not counted.

3. Human Ridehail Benchmark

3.1 Human Ridehail Mileage in ODD

During UMTRI fleet data collection from November 2016 to December 2018, over 30 million miles of driving were observed in San Francisco. In this study, the data was restricted to driving within the Geofence Area and applying the Road Type ODD filter, which reduced the mileage to 5,424,077 miles. This was the final UMTRI fleet ODD mileage exposure value. This value

[10] ABS: Anti-lock Braking System; ESC: Electronic Stability Control; TC: Traction Control

consisted of 228,213 miles of shared (UMTRI-VTTI) fleet driving and 5,195,864 miles of UMTRI non-shared fleet driving unique to the UMTRI dataset.

During VTTI fleet data collection from November 2016 to August 2018, over 2 million miles of driving were observed in San Francisco. In this study, the data was restricted to driving within the Geofence Area and applying the Road Type ODD filter, which reduced the mileage to 345,191 miles. The VTTI fleet ODD miles consisted of 189,413 miles of shared fleet driving and 155,778 miles of VTTI non-shared fleet driving unique to the VTTI dataset.

As indicated in 2.3.2.3 Shared Fleet Mileage Comparison, shared fleet mileages reported by VTTI were about 83% of the miles observed by UMTRI (i.e., 189,413 vs 228,213). Since odometer-based mileages used by UMTRI were considered the “gold standard” for measuring mileage accumulation in this study, a correction factor of 1.20484831 was established as a multiplier on VTTI miles in both the UMTRI-VTTI shared and VTTI unshared fleets. After applying this mileage correction, VTTI fleet ODD mileage exposure increased to 415,901 miles, which consisted of 228,213 miles of shared fleet driving and 187,688 miles of non-shared VTTI fleet driving.

The final mileage of the unshared UMTRI fleet (5,195,864 miles), the unshared VTTI fleet (187,688 miles), and the shared UMTRI-VTTI fleet (228,213 miles) came out to a collective total of 5,611,765 ODD miles.

3.2 Human Ridehail Crash Count in ODD

3.2.1 UMTRI Crash Count

Detailed crash counts for the UMTRI fleet are shown in Table 6. The methodology for deriving the final UMTRI mileage (ODD Imputation of Geofence Area with Road Type ODD Filter) is laid out in 2.3.2.1 UMTRI Mileage Measurement Methodology.

Driving Area	Total	High Location-Confidence Crash Data Recorder	Low Location-Confidence Crash Data Recorder	High Location-Confidence Insurance Claim Only	Low Location-Confidence Insurance Claim Only
San Francisco	489	146	13	300	30
Geofence Area	215	58		157	
Geofence Area with Road Type	171	51		120	

Driving Area	Total	High Location-Confidence Crash Data Recorder	Low Location-Confidence Crash Data Recorder	High Location-Confidence Insurance Claim Only	Low Location-Confidence Insurance Claim Only
ODD Filter					
ODD Imputation of Geofence Area with Road Type ODD Filter	**196.19**	**51**	**5.42**	**120**	**19.77**

Table 6 *– Breakdown of UMTRI crash counts across Driving Areas*

Crashes with high-confidence locations could be clearly classified as taking place inside or outside of the designated ODD. Crashes with low-confidence locations, i.e., those belonging to the **Low Location-Confidence Crash Data Recorder** or **Low Location-Confidence Insurance Claim Only** categories, went through a two-stage imputation process for generating an estimate of the number of crashes meeting the ODD definition. Details of these calculations are shown in Table 7.

Imputation Step	Sub-step	CDR Crash Count	Insurance Claim Only Count	Multiplier	Imputed CDR Crash Count	Imputed Insurance Claim Only Crash Count
Geofence Area	In San Francisco	0	24	0.834	0.00	20.02
	Not in San Francisco	2	3	0.099	0.20	0.30
	Unknown City	11	3	0.528	5.80	1.58
	Total				6.00	21.90
Road Type ODD Filter	Assumed to be in Geofence Area	6.00	21.90	0.903	**5.42**	**19.77**

Table 7 *– UMTRI imputation of crash estimates for low confidence crashes*

In the first stage, the likelihood that the crashes were located within the Geofence Area was evaluated based on the crash location reported in the insurance claim. Based on the high location-confidence crashes (combining CDR and insurance-only crashes), it was estimated that 83.4% of crashes marked as occurring in “San Francisco” actually occurred within the Geofence Area; 9.9% of crashes marked as occurring in another city actually occurred within the Geofence Area, and 52.8% of crashes with an unknown crash city occurred within the Geofence Area. These three estimates were used as multipliers for the observed low-confidence crashes with matching characteristics.

The second stage of this imputation process estimated the proportion of these low location-confidence crashes taking place in the Geofence Area that would pass the Road Type ODD filter. Based on an identified set of trips entirely within the Geofence Area, it was estimated that approximately 90.3% of driving within the Geofence Area was on a qualifying road type.[11]

Hence, the UMTRI fleet observed a total of 56.42 CDR crashes and 139.77 insurance-only crashes for a total of 196.19 estimated crashes within the designated ODD. Since integer valued crashes were not required for the Bayesian Fusion analysis, this fractional estimate of the UMTRI fleet crashes was used in the analysis.

3.2.2 VTTI Crash Count

Unlike the UMTRI fleet, the VTTI on-board DAS-based approach allowed researchers to consistently establish high location-confidence crash locations. A total of 21 ODD crashes were observed for the VTTI fleet, which consisted of 12 crashes in the UMTRI-VTTI shared fleet,[12] and 9 crashes in the VTTI non-shared fleet.

3.2.3 Shared Crash Count

The combination of the crashes detected in the shared fleet, with both VTTI and UMTRI contributing unique crashes, indicated that 15 crashes occurred in the shared fleet.

Table 8 shows the number of crashes in the shared fleet that were detected by the UMTRI fleet, the VTTI fleet, and the shared UMTRI-VTTI fleet.

	Detected by UMTRI Fleet	**Not Detected by UMTRI Fleet**	**Total**
Detected by VTTI Fleet	5	7	12

[11] This estimate excludes trips passing through the Geofence Area boundaries.
[12] Note that this is not the total number of crashes detected in the UMTRI-VTTI shared fleet, but rather the number of VTTI-detected crashes in the UMTRI-VTTI shared fleet.

	Detected by UMTRI Fleet	Not Detected by UMTRI Fleet	Total
Not Detected by VTTI Fleet	3		3
Total	8	7	15

***Table 8** – Comparison of crashes detected and not detected by UMTRI and VTTI fleets*

Overall, 7 (46.7%) crashes were detected only by VTTI, 3 (20.0%) were detected only by UMTRI, and 5 (33.3%) crashes were detected by both UMTRI and VTTI, totalling 15 crashes. A more detailed evaluation indicated that the 3 UMTRI-only crashes were all from insurance claims, and 2 of those appear to have involved sideswipes, suggesting that lateral impact crashes might be less likely to be detected by VTTI's kinematically-oriented triggers. On the other hand, VTTI's approach was expected to identify more lower-severity crashes, which appeared to be the case among the 7 detected only by VTTI.

The number of crashes not detected by either the UMTRI or VTTI fleet is unknown. However, the methodology of using three crash detection methods (CDR, insurance claim, and instrumented on-board DAS with human review) attempts to minimize the number of undetected crashes within the designated ODD. For benchmarking purposes, we assume that all relevant crashes were detected.

3.3 Human Ridehail Crash Rate in ODD

3.3.1 Individual Crash Rates

In the UMTRI fleet, 196.19 crashes were estimated to have occurred during the 5,424,077 miles of driving in the ODD. This implies an UMTRI fleet ODD crash rate of 27,647 miles per crash, or 36.2 crashes per million miles.

In the VTTI fleet, 21 crashes were estimated to have occurred during the 415,901 miles of driving in the ODD.[13] This implies a VTTI ODD crash rate of 19,805 miles per crash, or 50.5 crashes per million miles.

Within the shared fleet, the 15 crashes and 228,213 miles produced a higher estimated crash rate of 15,214 miles per crash, or 65.7 crashes per million miles. This reflected the more comprehensive capture of crashes possible when using both the VTTI and UMTRI detection methods.

[13] Note that throughout this section the corrected VTTI mileages are used, as described in 3.1 Human Ridehail Mileage in ODD.

3.3.2 Fused Crash Rate

As indicated above, relative to the UMTRI fleet, the VTTI fleet provided substantially less data volume (i.e., mileage and crash counts). However, the more detailed VTTI fleet data afforded the critical opportunity to identify unique crashes (particularly those of lower severity) that were not identified in the UMTRI fleet via insurance claim or CDR data. Similarly, the UMTRI fleet identified crashes not identified by VTTI (e.g., where the VTTI trigger approach did not capture certain crashes).

Thus, a combination of crashes identified by UMTRI and VTTI mitigated the potential risk of missing crashes that took place in the designated ODD. The comparison of crashes captured by the shared UMTRI-VTTI fleet played a foundational role in the statistical approach used below.

To correct for crashes missed by either UMTRI or VTTI, a statistical model called the “Bayesian Fusion Model” was developed by Cruise to combine the individual crash detection methods into a single estimate of the human ridehail crash rate. This model assumes that crashes follow a Poisson distribution, and divides the observed mileage and crash counts into three components: unshared UMTRI crashes, shared crashes, and unshared VTTI crashes.

These were modeled as three Poisson distributions which, when summed, give the estimated crash behavior of the full fleet, as shown below.

$$\begin{aligned}
\text{Unshared UMTRI Crashes} &\sim Poisson(m_{umtri}\lambda_{umtri}) \\
\text{Shared Crashes} &\sim Poisson(m_{shared}\lambda_{shared}) \\
\text{Unshared VTTI Crashes} &\sim Poisson(m_{vtti}\lambda_{vtti})
\end{aligned}$$

In this equation, m_w indicates the miles observed in a fleet w ($umtri$ = UMTRI unshared, $shared$ = UMTRI-VTTI shared, $vtti$ = VTTI unshared) and λ_w is the corresponding crash rate for that same fleet. This equation was further simplified by recognizing that the λ_w parameters are related through crash detection behavior.

Assuming the combination of the UMTRI and VTTI datasets identify all detectable crashes, the shared fleet has the true human ridehail crash rate, denoted as λ, while the other two non-shared fleets have lower observed crash rates depending on their relative detection probabilities. This can be reflected by parameterizing the detection probabilities as $\boldsymbol{p} = (p_{umtri}, p_{both}, p_{vtti})$, with $\sum p_i = 1$, where p_{both} is the probability that a crash was identified by both UMTRI and VTTI, and p_{umtri} and p_{vtti} are the probabilities that crashes were identified only by UMTRI or by VTTI, respectively. Under this parametrization, the λ_w can be defined in terms of λ and $(p_{umtri}, p_{both}, p_{vtti})$ as follows:

$$\begin{aligned}
\lambda_{umtri} &= \lambda(p_{umtri} + p_{both}) \\
\lambda_{shared} &= \lambda(p_{umtri} + p_{both} + p_{vtti}) = \lambda \\
\lambda_{vtti} &= \lambda(p_{vtti} + p_{both})
\end{aligned}$$

Since the number of crashes detected in the shared fleet is known, the rate was further differentiated into three components to permit modeling of the detection probabilities for the three fleets separately. This is done as shown below:

$$\lambda_{shared\_umtri} = \lambda p_{umtri}$$
$$\lambda_{shared\_both} = \lambda p_{both}$$
$$\lambda_{shared\_vtti} = \lambda p_{vtti}$$

The final Bayesian Fusion model is a mixture of the five Poisson distributions shown below:

$$\text{Unshared UMTRI Crashes} \sim Poisson(m_{umtri}\lambda(p_{umtri} + p_{both}))$$
$$\text{Shared UMTRI Only Crashes} \sim Poisson(m_{shared}\lambda p_{umtri})$$
$$\text{Shared UMTRI-VTTI Crashes} \sim Poisson(m_{shared}\lambda p_{both})$$
$$\text{Shared VTTI Only Crashes} \sim Poisson(m_{shared}\lambda p_{vtti})$$
$$\text{Unshared VTTI Crashes} \sim Poisson(m_{vttii}\lambda(p_{vtti} + p_{both}))$$

This parametrization of the model permits estimation of the overall human ridehail crash rate, λ, while still using the full combined dataset, rather than averaging the specific crash rates associated with each of the UMTRI- and VTTI-specific data collection methods. Note that the shared fleet serves as a lynchpin when using this approach, as the shared fleet provides full information about the detection probabilities ($\boldsymbol{p}$).

Based on the mileage and crash counts reported above for the UMTRI and VTTI fleets, the input data for the Bayesian Fusion model is shown in Table 9.

<table>
<tr><th>Fleet</th><th>Miles</th><th>Crashes</th></tr>
<tr><td>Unshared UMTRI Fleet</td><td>5,195,864</td><td>188.19</td></tr>
<tr><td>Shared Fleet - UMTRI Only Detection</td><td rowspan="3">228,213</td><td>3</td></tr>
<tr><td>Shared Fleet - UMTRI and VTTI Detection</td><td>5</td></tr>
<tr><td>Shared Fleet - VTTI Only Detection</td><td>7</td></tr>
<tr><td>Unshared VTTI Fleet</td><td>187,688</td><td>9</td></tr>
</table>

***Table 9** – Input data for the Bayesian Fusion model*

A visualization of the sizes and roles of the three fleets is shown in Figure 4.

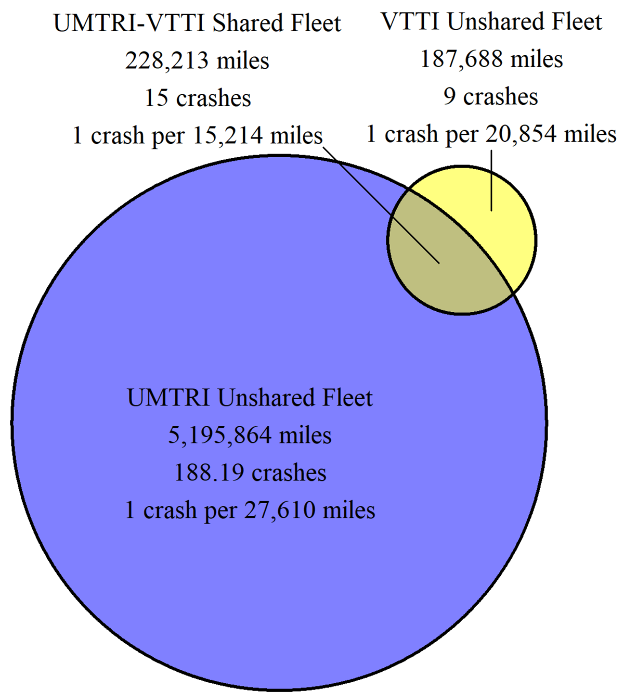

Figure 4 *– Visual representation of the UMTRI, VTTI, and UMTRI-VTTI shared fleets*

From this specification, the Bayesian Fusion model was fit, and a point estimate and 95% credible interval for λ (i.e., crash rate) was estimated.[14] This produced an estimated human ridehail crash rate of 1 crash in 15,414.4 ODD driving miles, with a corresponding 95% credible interval ranging between 11,682.1 and 20,477.8 miles.[15] This corresponds to an estimated 64.9 crashes per million miles of driving, with a corresponding 95% credible interval ranging between 48.8 and 85.6 crashes per million miles of driving.

3.4 Using the Benchmark

As noted in the introduction, the challenge for benchmarking driving performance within an ODD is that different driving environments (e.g., traffic level, time of day, road type, urban / rural, etc.) vary in the level of risk presented to drivers. This means that crash rates can vary even within a well-defined region and understanding driving exposure is a critical feature of any crash rate comparison.

The benchmark crash rate developed in this study represents the human ridehail driver crash rate given the travel demands of customers in San Francisco (and the preferences of the human

[14] To account for the fractional (0.19) crash in the UMTRI data, the Bayesian Fusion model was run twice, once rounding down to 196 and once rounding up to 197. The final posterior distribution was a weighted sampling of these distributions with a sampling ratio of 4 to 1 favoring the 196 count.

[15] Note that this fused rate is similar, but slightly higher, than the raw fused data set (which averages to 15,214 miles per crash). However, this method provides a much tighter confidence bound, and captures a much more diverse set of driving from a greater number of miles.

ridehail drivers observed). Given that there may be differences in the travel patterns of AVs and human drivers, it is appropriate to be aware of those differences and how they affect the environment-driven challenges presented when comparing AV and human crash rates.

4. Conclusion

The human ridehail crash rate data collection efforts described in this document are unprecedented. It consisted of 5.6 million miles of human ridehail driving data, as well as crash circumstance detail available across the UMTRI and VTTI datasets (e.g., automatic crash data recording, insurance claims, on-board data acquisition systems with multi-channel video). While developing the analysis method was challenging, the approach resulted in a first-of-its-kind targeted human-driver benchmark to match a specific ODD and ridehail driving purpose.

5. Acknowledgements

The authors would like to express sincere appreciation to Jon Hankey and Shane McLaughlin, who led the extensive naturalistic human ridehail instrumented vehicle data collection and analysis efforts conducted by VTTI. These efforts, as well as countless hours of discussion and collaboration with the authors, played a key foundational role in establishing the human ridehail crash rate benchmark reported in this paper.