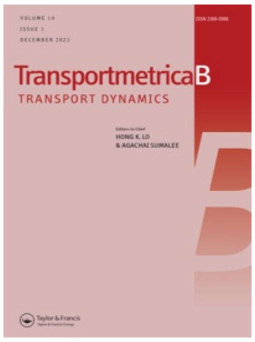

Transportmetrica B: Transport Dynamics

ISSN: 2168-0566 (Print) 2168-0582 (Online) Journal homepage: www.tandfonline.com/journals/ttrb20

Assessing factors influencing the occurrence of traffic conflicts: a vehicle-by-vehicle approach

Franco Basso, Yanara Muñoz, Raúl Pezoa & Mauricio Varas

To cite this article: Franco Basso, Yanara Muñoz, Raúl Pezoa & Mauricio Varas (2024) Assessing factors influencing the occurrence of traffic conflicts: a vehicle-by-vehicle approach, Transportmetrica B: Transport Dynamics, 12:1, 2332716, DOI: 10.1080/21680566.2024.2332716

To link to this article: https://doi.org/10.1080/21680566.2024.2332716

TRANSPORTMETRICA B: TRANSPORT DYNAMICS
2024, VOL. 12, NO. 1, 2332716
https://doi.org/10.1080/21680566.2024.2332716

Assessing factors influencing the occurrence of traffic conflicts: a vehicle-by-vehicle approach

Franco Basso[a,b], Yanara Muñoz[c], Raúl Pezoa[c] and Mauricio Varas[d]

[a]School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile; [b]Instituto Sistemas Complejos de Ingeniería, Santiago, Chile; [c]Escuela de Ingeniería Industrial, Universidad Diego Portales, Santiago, Chile; [d]Centro de Investigación en Sustentabilidad y Gestión Estratégica de Recursos, Facultad de Ingeniería, Universidad del Desarrollo, Santiago, Chile

ABSTRACT

This research analyzes the main factors that lead to the occurrence of traffic conflicts on the urban highway Autopista Central in Santiago, Chile. We propose a vehicle-by-vehicle approach in which we estimate the probability that each vehicle passing through each gate of Autopista Central has a conflict. This approach allows us to study variables at an individual level that have not been analyzed before using full real-world data, such as driver (age and sex) and vehicle information (type and age). We propose a Pseudo Time-to-Collision (PTTC) as a surrogate safety measure, defining different critical PTTC thresholds for each gate. Subsequently, a logistic regression model is built to understand the input variables' influence on the probability of conflict occurrence. Our results show that men and young drivers are more likely to have a conflict. In addition, the age of the car is negatively correlated with the occurrence of conflicts.

ARTICLE HISTORY
Received 29 September 2023
Accepted 14 March 2024

KEYWORDS
Conflicts analysis; vehicle-by-Vehicle; pseudo time-to-collision; logistic regression

1. Introduction

According to the World Health Organization, traffic accidents are a priority public health problem and correspond to the leading cause of death among young people. According to the World Health Organization, approximately 1.3 million people die yearly from traffic accidents, while 20 and 50 million people are injured. In addition, these traffic accidents involve a high monetary cost, accounting for about 3% of the GDP of most countries in the world. This global situation is also reflected in Chile. For example, according to data gathered by the National Traffic Safety Commission, during the year 2021, 80,751 road accidents were registered, resulting in 1,688 people killed by this cause and 51,928 people injured (Comisión Nacional de Seguridad de Tránsito 2022a), reaching costs of approximately 2% of the national GDP (Comisión Nacional de Seguridad de Tránsito 2021). The most common factor in these accidents is driver recklessness, which includes driving without paying attention to traffic conditions and driving without keeping a reasonable distance, among others (Comisión Nacional de Seguridad de Tránsito 2022b).

Literature shows several efforts conducted to study crash causes in order to improve traffic safety and decrease traffic accidents using real-world data (Basso et al. 2018; Hossain et al. 2019; Stylinaou, Dimitriou, and Abdel-Aty 2019; Basso, Basso, and Pezoa 2020; Basso et al. 2021; Wang et al. 2023). Crash data, however, usually suffers from several flaws that might limit the analysis, such as underreporting, misreporting, small sample size, and unobserved heterogeneity (Orsini et al. 2021; Zheng, Sayed, and

CONTACT Franco Basso francobasso@gmail.com School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile; Instituto Sistemas Complejos de Ingeniería, Chile

Mannering 2021; Essa and Sayed 2018; Hu et al. 2022). Due to these complications, some alternative methods to evaluate road safety have begun to be studied. One of these alternatives is to analyze traffic conflicts, which are related to traffic accidents (Orsini et al. 2021; Guo, Sayed, and Essa 2020). Road conflict analysis contributions have grown sharply in recent years (Orsini et al. 2021). The main advantage of conflicts is that they are more common than traffic accidents, so it does not take as much time to collect a sufficient amount of data. Nevertheless, a conflict analysis requires access to traffic data on a disaggregated level to compute the surrogate safety measures. Previous papers usually do not use this data granularity when analyzing the conflict's influential factors; instead, they aggregate the data usually in five-minute intervals.

In this context, a traffic conflict is an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements do not change (Güttinger 1984). This research analyzes traffic conflicts occurring on Autopista Central, an urban highway in Santiago, Chile. The traffic data is gathered from toll gates equipped with Automatic Vehicle Identification (AVI), whose collection devices are located on the gates over the highway lanes. This system features a low failure rate and allows us to collect vehicle-by-vehicle information from all users traveling on the highway. However, this data collection system cannot provide continuous vehicle trajectories, preventing us from constructing well-known traffic safety measures like the Time-to-Collision (TTC). This paper proposes a new safety measure, the Pseudo Time-to-Collision (PTTC), which can be computed using AVI gates information and allows us to define when a traffic conflict may occur. Moreover, AVI information allows us to obtain information about the vehicle and its owner. With this information, we analyze how vehicles and drivers' characteristics, such as sex and age, influence the risk of a collision and, thus, traffic conflicts.

The contribution of this work is twofold. First, we propose a novel vehicle-by-vehicle approach to studying the influential factors in road conflicts' occurrence. The proposed approach estimates the probability that each vehicle passing through each gate has a conflict, avoiding data aggregation as most of previous papers do. Our methodology allows us to include in the analysis variables that have not been studied before using full real-world data, such as driver and vehicle-related characteristics. Second, unlike other studies, the PTTC thresholds that define a conflict are determined differently for each highway segment in order to maximize the area under the Receiver Operating Characteristic (ROC) curve considering conflicts as a predictor for crashes.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on traffic conflicts. Next, Section 3 describes the data. Then, Section 4 explains the proposed methodology. In Section 5, the study's results are shown and analyzed, and finally, in Section 6, we present the main conclusions of this work and we establish a future research agenda.

2. Literature review

This review discusses the articles associated with traffic conflict analysis and is divided into four subsections. Subsection 2.1 describes the surrogate safety measures presented in the literature; Subsection 2.2 describes the different data collection methods used to study traffic conflicts; Subsection 2.3 presents articles analyzing factors influencing traffic conflicts; finally, Subsection 2.4 concludes the literature review, emphasizing the contributions of this research.

2.1. Surrogate safety measures

Some authors have proposed traffic conflicts as an alternative method for assessing road safety since accidents are rare events. The underlying hypothesis of traffic conflict analysis is that traffic conflicts are pre-accident events, so there is a relationship between them (Yuan et al. 2022b).

Different metrics, called surrogate safety measures (SSM), have been proposed in the literature to estimate the existence of a traffic conflict. Generally, conflict occurs when the value of any SSM exceeds a predetermined critical threshold. According to Wang et al. (2021), there are three main subcategories of SSMs based on different criteria: (i) time, (ii) deceleration, and (iii) energy.

Table 1. Most common surrogate safety measures.

SSM	Definition	Source	Formula
Time-to-Collision (TTC)	Time remaining before a collision if the users involved continue with their speeds and trajectories.	Hayward (1971)	$TTC = \frac{(x_{i-1} - x_i) - L_{i-1}}{v_i - v_{i-1}};\ \forall (v_i - v_{i-1}) > 0$
Post-Encroachment Time (PET)	Difference between the time when the first vehicle ends encroachment over the area of conflict and the second vehicle enters the area of conflict.	Allen, Shin, and Cooper (1978)	$PET = t_2 - t_1$
Deceleration Rate to Avoid a Crash (DRAC)	Rate at which a following vehicle must decelerate to avoid a collision with the leading vehicle	Cooper and Ferguson (1976)	$DRAC_i = \frac{(v_i - v_{i-1})^2}{2[(x_{i-1} - x_i) - L_{i-1}]};\ \forall (v_i - v_{i-1}) > 0$
Delta-V	Change of a velocity vector experienced by a road user during a crash	Shelby (2011)	$\Delta v_{i-1} = \frac{m_i}{m_{i-1} + m_i}(v_i - v_{i-1})$ $\Delta v_i = \frac{m_{i-1}}{m_{i-1} + m_i}(v_{i-1} - v_i)$

The most commonly used SSMs to measure temporal proximity to a collision is TTC and the post-encroachment time (PET) (Zheng, Sayed, and Mannering 2021). On the other hand, the most common deceleration-based indicator is the deceleration rate to avoid a crash (DRAC). Finally, there are several energy-based SSMs, with the Delta-V indicator being one of the first contributions to the literature (Wang et al. 2021). The aforementioned measures are described in Table 1. In this table, x_{i-1} and x_i are the positions of the leading and following vehicle, respectively; v_{i-1} and v_i are the speeds of the leading and following vehicle, respectively; L_{i-1} is the length of the leading vehicle; t_1 is the time when the first vehicle ends encroachment over the area of conflict; while t_2 is the time when the second vehicle enters the area of conflict; m_{i-1} y m_i are the masses of the leading and following vehicle, respectively.

In the literature, the most common SSM to identify conflicts is the TTC (Zheng et al. 2018; Hu et al. 2022; Yuan et al. 2022a; Kamel, Sayed, and Fu 2023). For this measure, different thresholds have been proposed to define a conflict. Examples include: 1.5 s (Essa and Sayed 2018; Formosa et al. 2020), 2.5 s (Guo, Sayed, and Essa 2020), 3 s (Yuan et al. 2022b; Arun et al. 2021) and 4 s (Hu et al. 2022; Huang et al. 2022). For the choice of these values, these studies rely on previous work or expert judgment, without developing a methodology for their choice.

In contrast, few contributions have developed a methodology to define the threshold when considering the TTC measure. Two notable examples are Orsini et al. (2021) and Yuan et al. (2022a), which we described next. Orsini et al. (2021) uses a threshold of 0.78 s, which is chosen by a probabilistic connection of conflicts to accidents, modeled with extreme value theory. On the other hand, Yuan et al. (2022a) performs sensitivity analysis for different TTC thresholds (1.5-2-2.5 s) using an XGBoost model with RENN, where the model achieves the maximum prediction performance when considering a threshold of 1.5 s.

2.2. Data collection methods used to study traffic conflicts

The literature reports the use of multiple data sources to study traffic conflicts. For instance, video cameras have been used to assess safety at signalized intersections (Guo, Sayed, and Essa 2020; Essa and Sayed 2018; Zheng and Sayed 2020; Arun et al. 2021). Additionally, drones and fixed detectors have also been used for collecting vehicle trajectory and traffic condition data, respectively (Hu et al. 2022; Yuan et al. 2022a; Yuan et al. 2022b). Likewise, Orsini et al. (2021) uses radar sensors on multiple highway sections for data collection.

2.3. Factors influencing road conflicts

The influence of traffic factors on conflicts has been the subject of several studies. Some relevant efforts on this topic are described next.

Hu et al. (2022) use the TTC indicator to identify conflicts and deploy machine learning models for prediction. The results show that the traffic condition factors are statistically significant for the machine learning models. Particularly, lane type is an important factor for road safety. In addition, the volume and standard deviation of speed have a positive impact on the occurrence of conflicts. On the other hand, Yuan et al. (2022b) also use the TTC indicator to define a traffic conflict and apply the Shapley Additive Explanations (SHAP) technique to interpret the results of four machine learning approaches. The authors conclude that the variation in speed, truck ratio, and traffic volume is positively associated with the occurrence of conflicts.

For the case of signalized intersections, Guo, Sayed, and Essa (2020) use the TTC, MTTC, and DRAC indicators to measure traffic conflicts. In their work, the authors develop safety performance functions that compare four Bayesian Tobit models. The results show that higher conflict rates are associated with shock wave characteristics, higher traffic volume, and maximum queue length. Essa and Sayed (2018) show that most rear-end conflicts occur at the beginning of the red light and the beginning of.

the green time, where the flow that was stopped begins to move forward gradually while other vehicles arrive at the end of the queue at higher speeds.

Aiming to provide route safety information to travelers, Huang et al. (2022) propose an approach that runs on a simulation platform. In their work, the authors use TTC and PET indicators to define a traffic conflict and then test three machine learning models to evaluate their performance. They conclude that dynamic factors, such as speed, acceleration, and traffic volume, contribute more to conflicts than stationary factors, such as road characteristics.

Finally, Formosa et al. (2020) develops deep neural network models to predict conflicts. In addition, the authors relate various safety measures to traffic variables. According to their results, the time to collision varies with speed, weather, and traffic density, suggesting that a single TTC threshold may not be appropriate for detecting traffic conflicts at different speeds.

2.4. Research gaps

Traffic conflict analysis is a topic that has been gaining traction in the literature in recent years, which can be explained due to the increasing amount of disaggregated vehicle-by-vehicle traffic data available for research.

In summary, conflict studies typically collect data using video cameras, fixed detectors on highway sections, or sensors in instrumented vehicles. However, all these methods present some drawbacks. In the case of video cameras, they collect data for a limited amount of time, and their analysis is usually limited to a specific location in a city. On the other hand, fixed detectors do not usually consider vehicle or driver information. In the case of instrumented vehicles, they have a low penetration rate. Moreover, the data obtained may not represent the usual behavior of drivers since drivers who know that they are in an experiment tend to change their behavior (Wouters and Bos 2000).

Besides, this review concludes that the most widely used measure of conflict by other authors is the TTC indicator. However, the threshold for defining a conflict has not been widely studied. Furthermore, the variables used in previous studies to explain conflicts generally only consider traffic and vehicle data and exclude drivers' personal factors.

> This paper attempts to fill these two gaps by studying vehicle-to-vehicle traffic conflicts using data obtained from AVI gates located on a whole urban highway. Using such data does not allow us to compute the TTC since we do not have continuous trajectories. Thus, to tackle this drawback, we propose a new safety measure called the PTTC that matches AVI data characteristics. Moreover, we analyze different thresholds to define a conflict, considering its relation to traffic accidents. Finally, drivers' personal factors are included in the analysis.

3. Data description

In this research, we use disaggregated data, including vehicle-by-vehicle information of all users traveling on Autopista Central, an urban highway located in Santiago, Chile. This highway has a length of 60.5 km and crosses the city from north to south through the Ruta 5 and General Velásquez routes, as shown in Figure 1.

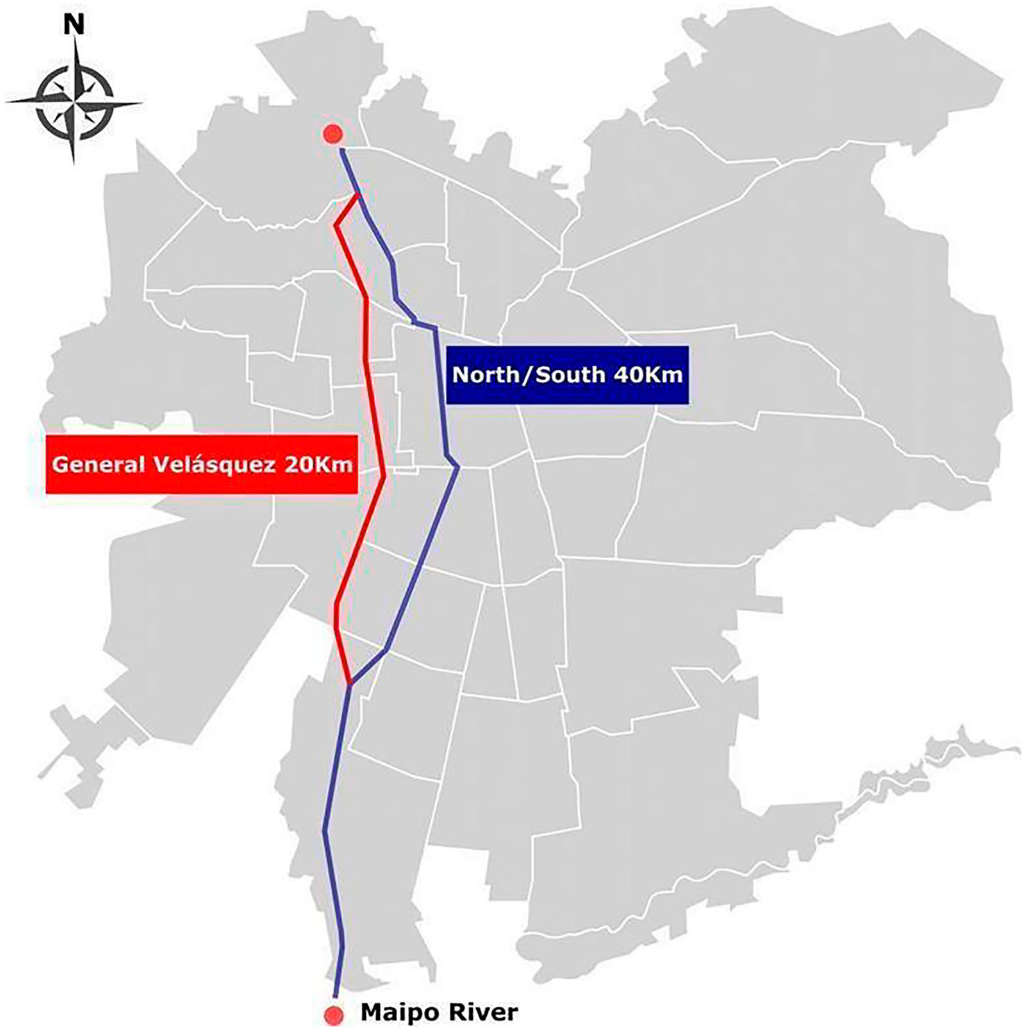

Figure 1. Autopista Central, Santiago, Chile. Source: Basso et al. (2018).

The vehicle's information comes from AVI technology, a system located in all the highway free flow gates that allow automatic fare charging by identifying the TAG device. It is relevant to point out that, by law, this device must be installed on the front windshield of each vehicle traveling on the highway. So, we have complete information about all the users of the highway.

3.1. Autopista central data

Autopista Central provides two different databases for the studied period, which ranges from January to June 2021. The first contains information on each vehicle that used the highway, while the second includes the traffic accidents that occurred during the period studied. Both databases are described below.

3.1.1. AVI database

There are 246,259,352 observations recorded in this database, with each entry corresponding to a vehicle passing through one of the 31 AVI gates along the highway. However, 16,871,432 observations (6.9% of the total) lack information in the license plate field. This absence could be attributed to issues with the AVI gate technology or, in some cases, because the vehicle itself did not have a visible license

Table 2. List of variables contained in the AVI Database.

Variable	Description
Date	Time the vehicle passed through the gate
License plate	Vehicle identification
Gate	Number of the gate the vehicle passed through
Category	Category of vehicle (light, heavy, or motorcycle)
Lane	Lane in which the vehicle passed through the gate (right, middle, or left)
Speed	The instantaneous speed at which the vehicle passed through the gate (km/h)

Table 3. Monthly descriptive statistics of the AVI database.

Month	Vehicle flow	Average Speed (km/h)	Right Lane Flow	Middle Lane Flow	Left Lane Flow
January	39,832,732	84.467	10,650,609	14,942,710	14,239,413
February	38,735,370	85.936	10,269,882	14,442,230	14,023,258
March	39,914,992	83.366	10,711,657	14,803,755	14,399,580
April	32,774,974	88.019	8,830,851	12,438,672	11,505,551
May	38,543,343	83.434	10,370,498	14,494,167	13,678,678
June	39,586,509	83.768	10,699,948	15,044,282	13,842,279
Average	**38,231,320**	**84.832**	**10,255,574**	**14,360,969**	**13,614,793**

Table 4. Accident database variables.

Variable	Description
Date	Day and time at which the accident occurred
Segment	Highway stretch at which the accident occurred
Carriageway	The direction where the accident occurred (eastbound or westbound)
Kilometer	Place where the accident occurred

plate. Consequently, this limitation prevents the retrieval of driver-related information for those specific instances. Following the exclusion of these entries, the database is left with a total of 229,387,920 observations. Table 2 lists and describes the variables included in the database.

Table 3 shows a descriptive analysis of the period studied. The total number of vehicles passing through an AVI gate is relatively constant, almost 39 million, except for April, when the number of passing vehicles decreases. This is explained by the Covid-19 pandemic, which forced the declaration of a lockdown in the metropolitan area in April 2021. As a result, the average speed increases in April, reaching 88 km/h. In addition, note that the right lane has a significantly lower number of vehicles than the middle and left lanes. One possible explanation is that highway users do not prefer the right lane because of the congestion caused by entering and exiting vehicles.

3.1.2. Accidents database

This database has 1,036 observations corresponding to all Autopista Central crashes during the studied period. The variables of the accidents database are shown in Table 4.

Figure 2 shows the monthly traffic accidents for the 1,036 accidents that occurred during the study period. February and March had the lowest accidents, with 155 observations each. In addition, in May it is reached the maximum number of accidents compared to the other months, with 222 observations.

On the other hand, Figure 3 shows the spatial distribution of accidents on Autopista Central. The highest risk areas are located in the central part of Route 5 (gates 12, 13, 14, and 31) and at the two intersections of Route 5 and General Velásquez (northern zone gates: 17 and 18; southern zone gates: 5, 6, and 32).

3.2. Drivers database

The third database used contains information about all drivers circulating on Autopista Central. The creation of this database involved querying the Civil Registry website to gather details such as the age of the car, the gender, and the age of the owner. This information was obtained for each of the 1,357,754 license plates documented in the flow database in January 2021. However, 96,227 license

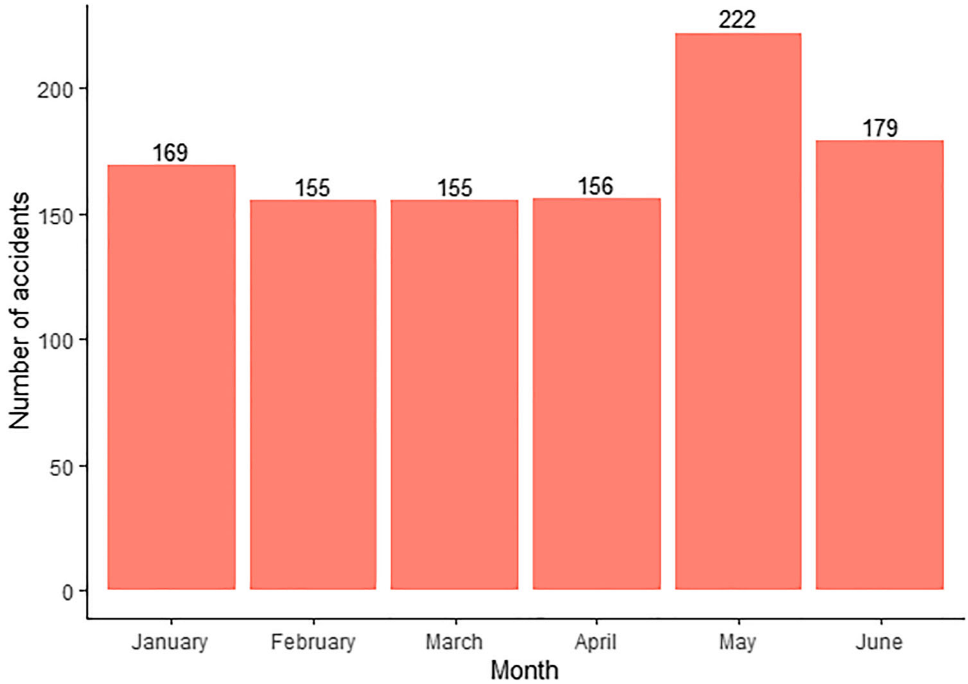

Figure 2. Number of accidents by month.

Table 5. Gender of the drivers circulating during January 2021.

Sex	Number of drivers	Percentage
Female	402,208	38%
Male	655,855	62%

plates (7.1% of the total) lacked available information, and an additional 203,464 license plates (15.0%) were registered as company vehicles. We restrict our analysis to the 1,058,063 licenses plate belonging to natural persons, and then, similarly to Basso et al. (2022), we assume that who drives is the vehicle's owner. This strong but reasonable assumption allows us to obtain unavailable information. Table 5 shows the number of drivers by gender who traveled on the highway during the study period and the corresponding percentage distribution. On the other hand, Figure 4 shows vehicle owners' age distribution.

Finally, it is important to point out that driver's information is not needed for determining the PTTC threshold. Thus, we utilize the full dataset available from January to June 2021. This is because our methodology involves adjusting the threshold to enable conflicts to serve as predictors of accidents, which are relatively rare events. Consequently, a larger volume of data is necessary to ensure a sufficient number of accidents for analysis, with the minimum threshold established at 15 accidents per AVI gate. However, we consider January 2021 only for calibrating the logistic regression since we do not have more drivers' information. Nevertheless, this is not problematic because the vehicle-by-vehicle nature of our approach implies that the logistic regression is calibrated with an extensive data set with a sample size of 21,509,091 (see Subsection 5.1). Incorporating more months should not change the results.[1]

4. Methodology

In this section, we present the methodology used in this paper. For the sake of the exposition, this section is structured as follows. In Subsection 4.1, we present the PTTC safety measure, whereas in

Figure 3. Spatial distribution of accidents.

Subsection 4.2, we propose an algorithm for determining, for each gate, a PTTC threshold that defines the occurrence of a conflict. Finally, in Subsection 4.3, we show how we construct the input variables used in the logistic regression model.

4.1. PTTC definition

The AVI data collection system inhibits the capture of continuous vehicle trajectories as it only tracks vehicles at toll gates. This limitation hampers the construction of established traffic safety measures reliant on trajectories, such as TTC. As a result, in this paper, we slightly modified the TTC definition to match our data characteristics, as shown in Figure 5. Specifically, for each pair of consecutive vehicles

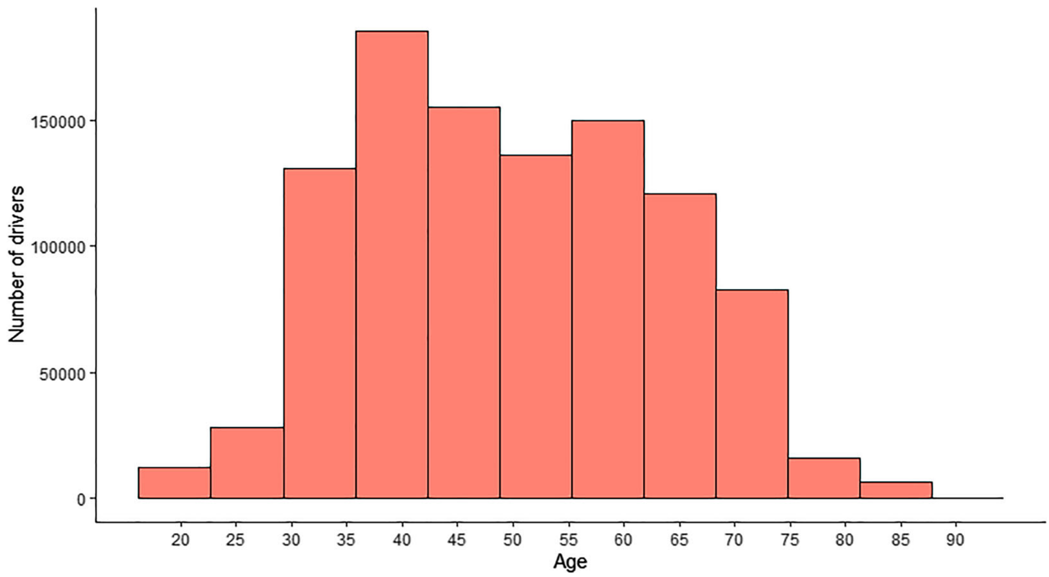

Figure 4. Vehicle owners' age distribution during January 2021

traveling on the same gate and lane let t be the time that elapses between their passages, and by s_{i-1} and s_i be the speeds of the leading and following vehicle at the AVI gate, respectively. Then, considering L_{i-1} as the length of the leading vehicle, the PTTC formula is given by Equation 1.

$$PTTC = \begin{cases} \frac{s_{i-1}t - L_{i-1}}{s_i - s_{i-1}}, & \text{if } s_i > s_{i-1} \\ \infty, & \text{otherwise} \end{cases} \tag{1}$$

The main difference between TTC and PTTC lies in the temporality of the data. In the TTC, both vehicles are observed at the same instant. In contrast, in the PTTC, the vehicles are observed at different moments, corresponding to the instant when each one crosses the AVI gate. Note that the PTTC corresponds to the TTC computed when the following vehicle passes through the AVI gate, assuming the leading vehicle keeps its speed constant. This last assumption is reasonable, particularly for urban highways, considering that the AVI free-flow gates do not interfere with the normal circulation of vehicles. It is worth noting that since the variables we use for the computation consider the speeds of two consecutive vehicles in the same lane, the PTTC does not account for lane-changing scenarios, focusing only on rear-end conflicts.

4.2. Defining critical threshold for each AVI gate

Since conflicts can be considered quasi-accidents, the primary motivation for studying them relies on the fact that reducing the factors that favor the occurrence of conflicts could improve road safety. On this, conflicts should also differ because the factors influencing a crash's occurrence differ depending on the spatial locations studied (Mitra 2009). To tackle this issue, in this paper, we consider different definitions of conflicts depending on the AVI gate considered, which contrast with most of the previous contributions.

Several authors have argued that keeping the TTC threshold constant when defining conflicts might be problematic for assessing crash risk (Das and Maurya 2019; Nadimi, NaserAlavi, and Asadamraji 2022). This is because the concept of risk could differ depending on several factors, such as the type of vehicle involved, the type of road studied, and the participants' driving skills. For example, more experienced drivers usually have more aggressive behavior, leading to an average lower TTC

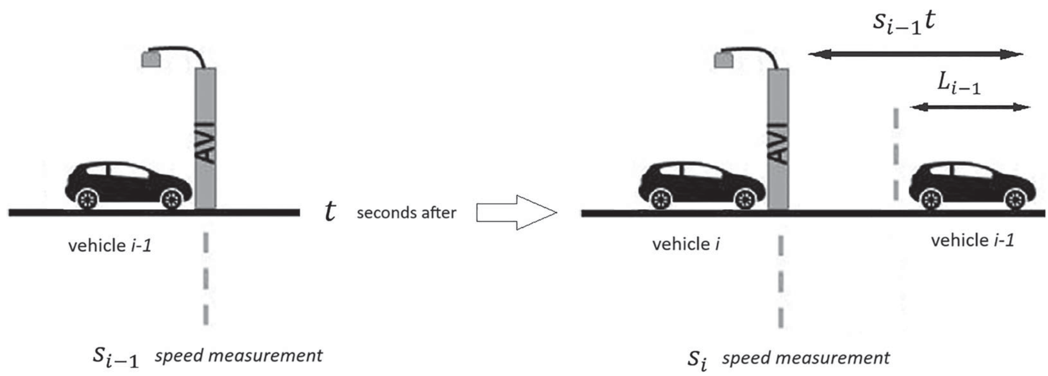

Figure 5. PTTC computation

value (Li, Jiang, and Lu 2011) without necessarily implying a higher crash rate (Yuan et al. 2021). In our case, we consider different TTC thresholds according to road geometry, i.e. different TTC thresholds at each AVI gate. The assumption behind this decision is that the link between calculated conflicts and crashes (or observed conflicts) differs depending on the road geometry (Pinnow et al. 2021).

This study uses the PTTC indicator to define a traffic conflict. However, there is no consensus on the suitable threshold value for defining a conflict using the TTC. Therefore, different values have been used in the literature to define a conflict. To cope with this issue, in this research, we develop a novel methodology for systematically determining the critical thresholds that maximize the predictive accuracy of conflicts (as a predictor for crashes). For this purpose, in our paper, we use the ROC curve, a graphical plot that is used to evaluate the performance of a binary classification model. It is particularly useful in determining the trade-offs across various thresholds between true positive rate (sensitivity), which we take as the x-axis, and false positive rate (1-specificity), which we take as the y-axis.

The sensitivity and specificity performance metrics are computed using Equations (2) and (3), respectively. In this case, TP represents the true positive, FN the false negative, TN the true negative, and FP the false positive for the resulting confusion matrix.

$$sensitivity = \frac{TP}{(TP + FN)} \tag{2}$$

$$specificity = \frac{TN}{(TN + FP)} \tag{3}$$

Consider Algorithm 1. This procedure search for the optimal value α^* that maximizes the area under the ROC_α curve for $\alpha = \{0.5, 1, 2, \ldots, 15\}$. The ROC_α curve is computed using the sensitivity and specificity measures for the confusion matrix given by the variables Y_t and $\hat{Y}^{\alpha}_{t,i}$ where the former is binary variable indicating whether or not there was an accident in period t, and the latter is a variable that predicts an accident if the number of conflicts Z^{α}_{t-1} is greater than the index i in period t.

4.3. Input variables and model

In the final base, each row corresponds to a driver passing through a gate during the studied period. For each pair of consecutive vehicles, the PTTC is calculated, and the dependent variable *Traffic Conflict* is constructed. This binary variable takes the value one if the driver has a traffic conflict (*i.e.*, $0 < PTTC_v \leq \alpha^*_p$ for vehicle v and the threshold α^*_p for gate p) and zero otherwise.

Algorithm 1.

for $\alpha = 0.5, 1, \ldots, 15$
Define n_α as the maximum number of conflicts using the threshold α considering all periods.
for $i = 1, \ldots, n_\alpha$
for $t = 1, \ldots, T$
Define Y_t as binary variable indicating whether or not there was an accident in period t.
Define Z_{t-1}^α as the number of conflicts occurred in period $t-1$ considering the threshold α.
if $Z_{t-1}^\alpha \geq i$ **then.**
Define and set $\hat{Y}_{t,i}^\alpha \leftarrow 1$ (An accident interval is predicted).
else
Define and set $\hat{Y}_{t,i}^\alpha \leftarrow 0$ (A non accident interval is predicted).
end if
end for
Compute $sens_i^\alpha = Sensitivity(Y_t, \hat{Y}_{t,i}^\alpha)$
Compute $spec_i^\alpha = Specificity(Y_t, \hat{Y}_{t,i}^\alpha)$
end for
Compute the area under the ROC_α curve using vectors $sens_i^\alpha$ y $spec_i^\alpha$.
end for
Return $\alpha^* \leftarrow argmaxROC_\alpha$

Table 6. Independent variables of the model.

Category	Variable	Type
Driver	Sex	Categorical (Male, Female)
	Driver Age	Numerical (integer)
Vehicle	Vehicle Age	Numerical (integer)
	Type	Categorical (Light, Heavy, Motorcycle)
	Lane	Categorical (Right, Middle, Left)
	Speed	Numerical (positive real)
Traffic	Flow	Numerical (integer)
	Average Speed	Numerical (positive real)
	Speed Standard Deviation	Numerical (positive real)
	Average Speed Difference	Numerical (positive real)
Geometry	Straightness	Numerical ([0,1])
Context	Night	Categorical (Night, Day)

Table 6 shows the independent variables. These variables are grouped into five categories: Driver, Vehicle, Traffic, Road Geometry, and Context. Variables Flow, Average Speed, and Speed Standard Deviation contain traffic characteristics one.

minute before the vehicle under study passes. The variable Average Speed Difference corresponds to the difference in speed between the lane in which the vehicle under investigation was traveling and the adjacent lane one minute before. If the studied vehicle crosses the gate in the middle lane, this variable considers the other two lanes. The Straightness variable is the quotient between the straight line distance and the network distance. These two distances are calculated taking into account the gate through which the vehicle under study passed and the previous one.

Finally, we run a logistic regression model considering *Traffic Conflict* as the response variables and the ones shown in Table 6 as independent variables. Overall, the modelling process we follow is depicted in Figure 6.

5. Results

This section presents the results of the study. In particular, Subsection 5.1 presents the thresholds chosen for each AVI gate and the descriptive statistics of traffic conflicts for the study period. Then, Subsection 5.2 presents and analyzes the results of the logistic regression model.

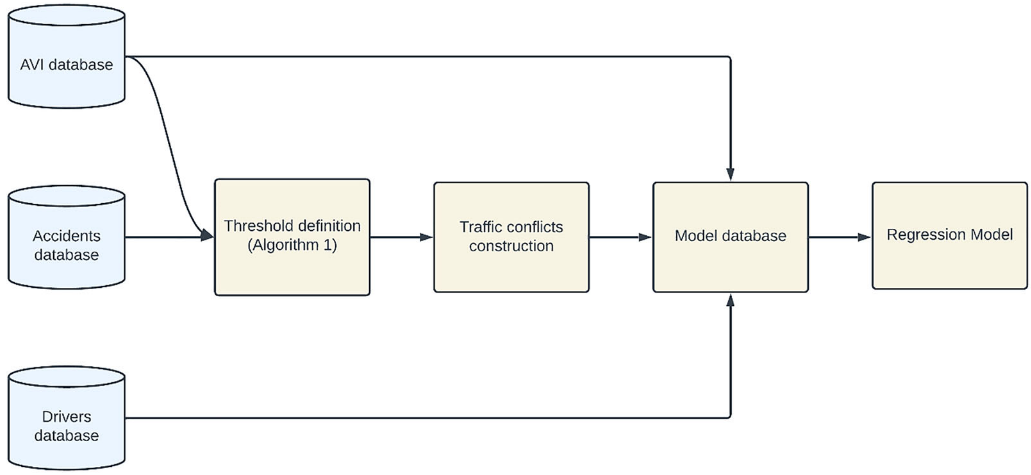

Figure 6. Modelling process.

5.1. PTTC Thresholds

Following the methodology described in 4.2, we propose the critical thresholds using the data from January 2021 to June 2021. Table 7 shows, for each AVI gate, the selected thresholds, the AUC, and the number of crashes. We take $L_{i-1} = 4$ meters for light vehicles, $L_{i-1} = 13$ meters for trucks and buses and, $L_{i-1} = 2.5$ for motorcycles.

Only those gates with at least 15 crashes were considered in this study. We think this number allows us to choose a threshold that properly represents the occurrence of conflicts and crashes. Therefore, following this rule, 26 gates are considered for the conflict study, discarding gates 1, 8, 17, 26, and 28. In addition, the data of vehicles belonging to companies are not considered since we aim to analyze the personal factors of the drivers, and this can only be achieved by studying vehicles belonging to private drivers. This leaves 21,509,091 observations (which correspond to the vehicles passing through one of the AVI gates).

Table 7 shows that the computed thresholds are much higher than those previously obtained in the relevant literature. This can be explained by considering that crashes are rare events; consequently, a high number of crash predictions must be made to achieve high sensitivity, as suggested in Equation 1. Furthermore, in our context, a crash prediction is linked to many conflicts (Line 7, Algorithm 1). Therefore, our approach finds it desirable to establish high thresholds to detect more conflicts, thus increasing sensitivity and the area under the ROC curve.

Also, because of this high threshold, raising the question of whether the crash occurs in the same AVI gate where the conflict is detected is natural. In fact, since the maximum TTC threshold selected is 15 s (AVI gates 2, 9, 12, 25, and 30) and considering a vehicle circulating at 80 km/h, the potential crash would occur 333 meters after the measurement. Moreover, as the studied road is an expressway (with a maximum legal speed of 100 km/h) and the AVI gates are separated on average by 1.94 km, we think that aligning the crash data occurrences at the same location as the AVI is reasonable.

Even though our primary goal is not to predict crashes but to determine thresholds for the PTTC, our AUC values (mean $= 0.6901$) align with several previous papers that test their model in full real-world unbalanced data, as we do. For example, Theofilatos, Chen, and Antoniou (2019) reports an AUC value of 0.641 using deep learning techniques for predicting crashes in a real-world setting. Orsini et al. (2021) obtains an average AUC of 0.68 after conducting 100 repetitions for training, test, and full-test datasets. Basso et al. (2021) reports an AUC of 0.73 for their best model considering DCGAN oversampling technique and a new image-inspired data architecture. Thus, we believe that the proposed PTTC is a suitable conflict indicator.

Table 7. Thresholds selected for each AVI gate.

Gate	Thresholds	AUC	Crashes
1	5.5	0.9927197	1
2	15	0.7272382	15
3	13	0.7262644	37
4	14	0.6275027	21
5	10.5	0.7978325	50
6	4.5	0.6622788	20
7	13.5	0.7078474	42
8	13.5	0.6368184	10
9	15	0.645574	46
10	14.5	0.7916816	61
11	11.5	0.6919964	20
12	15	0.6378802	49
13	11.5	0.6310587	54
14	4.5	0.7750877	28
15	14	0.6921721	67
16	13	0.7542039	96
17	9	0.6672352	13
18	8.5	0.657902	57
19	12	0.6731778	40
20	14	0.549195	15
21	10.5	0.6222679	26
22	9.5	0.6735767	22
23	13	0.7964983	17
24	9.5	0.6511217	21
25	15	0.7161509	61
26	12	0.7253671	11
28	7.5	0.9027991	4
29	14	0.6175327	41
30	15	0.7299808	28
31	10	0.584501	28
32	8	0.6920037	22

On the other hand, Table 8 shows the number of vehicles circulating at each gate, the number of conflicts that occurred, and their percentage of the vehicle flow. The highest number of conflicts occurred at gates 10, 12, and 13, located in the south-central zone of the studied highway. The highest percentage of conflicts is far more concentrated, compared to the flow, in gates 12 and 25, corresponding to the south-central zone of Route 5 and the north-central zone of General Velásquez, respectively.

Figure 7 shows a bubble chart of traffic conflicts on the studied highway. The area with the highest concentration of conflicts is located at gates 10, 12, 13, 14, 14, 25, and 30. These results support the relationship between crashes and traffic conflicts since the heat map shown in Figure 3 and the bubble chart in Figure 7 show similar patterns.

On the other hand, Figure 8 shows the distribution of conflicts by the time of day. Note that there are more conflicts during peak hours, i.e. between 8 and 9 am and between 5 and 7 pm.

5.2. Logistic regression

To understand the factors that influence the likelihood of a traffic conflict, we construct a logistic regression using the independent variables listed in Table 6. Table 9 shows the descriptive statistics of the explanatory variables of the model. Table 10 shows the results of this regression, including the estimated coefficients for each variable, the estimated standard error, the *p*-value associated with the statistical significance test, the odds ratios (OR), and their 95% confidence intervals (Guo et al. 2014; Bai et al. 2015). The signs of the estimated coefficients are used to analyze the signs of the effects of the variables, while the OR is used to determine its impact.

Table 8. Number of conflicts and number of vehicles per AVI gate.

Gate	Number of vehicles	Number of conflicts	Percentage
2	668,018	56,919	8.52%
3	780,182	57,519	7.37%
4	636,297	41,375	6.50%
5	1,093,417	64,374	5.89%
6	913,638	14,780	1.62%
7	726,386	54,039	7.44%
9	945,995	80,450	8.50%
10	1,057,741	98,153	9.28%
11	992,005	56,762	5.72%
12	1,177,133	121,120	10.20%
13	1,245,999	96,482	7.74%
14	1,114,227	40,477	3.63%
15	523,476	53,913	10.30%
16	659,832	51,868	7.86%
18	937,950	49,398	5.27%
19	478,830	44,485	9.30%
20	537,522	44,036	8.20%
21	710,924	34,529	4.86%
22	654,934	26,143	3.99%
23	726,630	54,528	7.50%
24	724,250	36,442	5.03%
25	743,066	81,332	10.95%
29	354,764	46,430	13.09%
30	982,736	81,358	8.28%
31	1,253,396	60,343	4.81%
32	869,743	46,479	5.34%

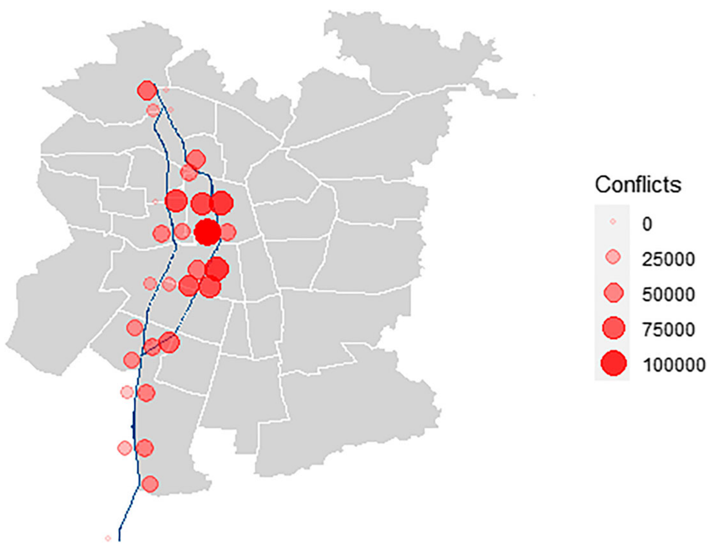

Figure 7. Conflict bubble chart in Autopista Central.

Regarding the variables related to the driver, it can be observed that if the driver is male, there is an increase of 8.0% in the probability of having a conflict than if the driver is female. On the other hand, older drivers are less likely to be involved in conflicts. These variables are not usually included in the analysis of traffic conflicts; however, they have been studied in previous research in other contexts. For example, in the context of lane changes, Basso et al. (2022) find that younger drivers change lanes more often than older drivers. In addition, Adanu et al. (2021) show that younger drivers are involved in twice as many crashes as older drivers.

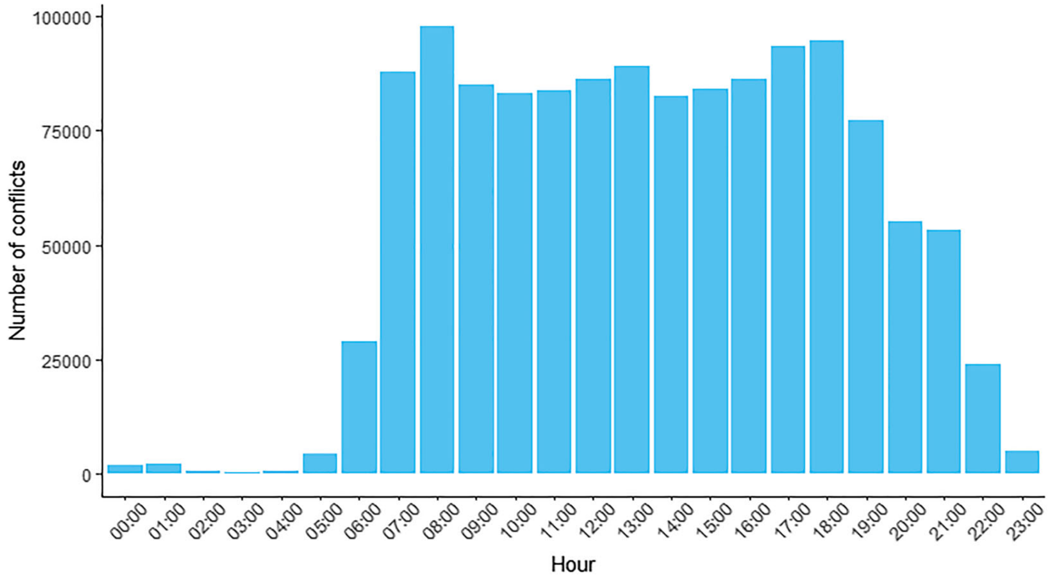

Figure 8. Distribution of conflicts per hour.

Table 9. Descriptive statistics of explanatory.

Variable	Min	Max	Mean	SD	Category Frequency
Sex	-	-	-	-	Male (13,627,004), female (7,882,087)
Driver Age	19.040	88.240	48.694	14.001	-
Vehicle Age	2.000	28.000	8.551	5.509	-
Type	-	-	-	-	Light (20,551,494), Heavy (872,849), Motorcycle (84,748)
Lane	-	-	-	-	Right (5,672,527), Middle (8,044,625), Left (7,791,939)
Speed	1.000	200.000	85.118	23.435	-
Flow	1.000	128.000	30.659	15.100	-
Average Speed	5.000	150.091	85.231	19.134	-
Speed Standard Deviation	0.688	51.576	13.327	4.494	-
Average Speed Difference	0.000	96.600	11.524	10.050	-
Straightness	0.901	1.000	0.979	0.028	-
Night	-	-	-	-	Day (18,562,807), Night (2,946,284)

Table 10. Logistic regression results.

Variable	Estimated Coefficient	Standard Error	*p*-value	Odds Ratio	95% Confidence Interval for OR
Intercept	$-6.598e+00$	3.309e-02	< 2e-16***	0.001	0.001-0001
Sex Male	7.737e-02	1.831e-03	< 2e-16***	1.080	1.077-1.084
Driver Age	$-3.198e-03$	6.337e-05	< 2e-16***	0.997	0.997-0.997
Vehicle Age	$-1.144e-02$	1.735e-04	< 2e-16***	0.989	0.988-0.989
Type Heavy	$-5.823e-01$	6.550e-03	< 2e-16***	0.559	0.551-0.566
Type Motorcycle	6.774e-01	9.737e-03	< 2e-16***	1.969	1.932-2.007
Lane Middle	$-2.654e-01$	2.164e-03	< 2e-16***	0.767	0.764-0.770
Lane Left	$-1.387e+00$	2.595e-03	< 2e-16***	0.250	0.249-0.251
Speed	4.646e-02	6.552e-05	< 2e-16***	1.048	1.047-1.048
Flow	2.677e-03	7.123e-05	< 2e-16***	1.003	1.003-1.003
Average Speed	$-5.226e-02$	8.146e-05	< 2e-16***	0.949	0.949-0.949
Speed Standard Deviation	$-6.287e-04$	2.304e-04	0.00636**	0.999	0.999-0.999
Average Speed Difference	1.349e-03	8.546e-05	< 2e-16***	1.001	1.001-1.002
Straightness	$5.175e+00$	3.308e-02	< 2e-16***	176.871	165.768-188.718
Night	$-3.949e-01$	3.059e-03	< 2e-16***	0.674	0.670-0.678

For the variables related to the vehicle, the results show that the age of the vehicle is negatively associated with conflict occurrence. Furthermore, regarding the type of vehicle, there is a decrease of 44.1% for heavy vehicles and an increase 96.9% for motorcycles in the probability of having a conflict compared to light vehicles. In addition, Table 9 shows a decrease in the probability of having a conflict of 75.0% for the left lane and 23.3% for the middle lanes compared to the right lane. The latter can be explained by the fact that in the right lane are the entrances and exits of the cars, which generate more interactions and, therefore, riskier situations. Finally, regarding the speed of the studied vehicle, it can be concluded that the higher the speed, the higher the probability of conflict.

The model also shows that the higher the flow one minute before the study vehicle passes, the higher the probability of conflict. This may occur because in situations of high flow, the interactions between vehicles increase, thus increasing the probability of a rear-end collision (Dimitriou, Stylianou, and Abdel-Aty 2018). The standard deviation of speed is negatively correlated with the conflict occurrence. These findings differ with those of Hu et al. (2022) and Yuan et al. (2022a). On the contrary, according to the model results, the probability of a conflict increases when other vehicles have low average speeds. This can be explained since a higher density of vehicles accompanies a lower speed, so there is more interaction between them. This is consistent with previous studies that have found the same results (Yuan et al. 2022b; Yang et al. 2018).

Finally, we can observe that the less curved the road a vehicle travels before approaching a gate, the more likely it is that a conflict will occur. In addition, the probability of conflicts is lower at night than during the day, which can be explained by the fact that at night there is more distance between vehicles.

6. Conclusions

Traffic accidents have increased in the last decades, resulting in material and human damage. In this context, several studies have been carried out on traffic accidents to understand their causes and implement safety measures to reduce them. However, there are some limitations in using accident data, mainly related to the low number of available cases. Therefore, analyzing traffic conflicts has become relevant as an alternative method to evaluate road safety.

This study analyzes the factors that influence the occurrence of traffic conflicts. For this purpose, vehicle-by-vehicle information is collected from drivers traveling on Autopista Central in Chile. Since our data does not include continuous vehicles' trajectories, we propose a Pseudo Time-to-Collision (PTTC) measure, which can be computed using AVI gates information and allows us to define when a traffic conflict may occur. Then, traffic variables are considered in addition to variables not previously included in the literature, such as driver and vehicle information. On the other hand, the PTTC indicator between two consecutive vehicles is calculated to identify a traffic conflict, with a critical threshold defined for each gate, differentiating the risk in different zones. Subsequently, a logistic regression model is calibrated to explain the factors that influence the occurrence of conflicts, where each row of the database created corresponds to a single driver passing through a gate.

The model results show that the probability of a conflict increases with the variables speed, flow, speed standard deviation, average speed difference, and straightness. On the other hand, the probability of having a conflict decreases with the driver's age, the vehicle's age, and the average speed of the vehicles that passed one minute before the vehicle under study. It is also observed that the likelihood of a conflict is higher during the day than at night. Moreover, men are more likely to have conflicts than women. Finally, the results indicate that light vehicles and vehicles driving in the right lane are more likely to have a conflict.

The results of this study can be used by highway decision-makers to develop strategies and propose safety measures to reduce traffic conflicts and accidents, thus improving road safety. For example, the results show that users with different characteristics behave differently. Therefore, one proposal is to transmit information through variable message signs with personalized messages that depend on the group of drivers circulating on the roads, taking into account their age, vehicle age, and traffic conditions, among other variables. In addition, considering that the right lane has a higher probability of conflict than the other lanes, another alternative is to increase the distances between entrances and exits in some sections where the distances are small, thus reducing vehicle congestion in these sections. On the other hand, the results of this research can help to understand the behavior of human drivers and improve the parameters in traffic simulators to have more realistic representations.

Several lines of future research follow from this study. First, further investigation of the relationship between conflicts and crashes is necessary, for example, separating by crash type, which we do not do in this research. Although we contribute by using crashes to determine different PTTC thresholds, predicting crashes using conflicts is a relatively unexplored area (Orsini et al. 2021) that might benefit from the inclusion of vehicle-by-vehicle and individual driver information, as we do in this research. Second, new sources of information, such as in-vehicle cameras or biometric information, can be included to explore further personal variables in order to explain the occurrence of conflicts. Third, more complex models could be used to capture other problem features. For example, using random parameters logistics regression could help reveal unobserved heterogeneity. However, to implement such a model, due to the vehicle-by-vehicle nature of our approach, it would be necessary to develop faster methods for estimating the model parameters that could cope with such a large dataset.

Note

1. To support this point, we randomly took three times the 50% of the January 2021 sample and recalibrated the model, obtaining almost the same results for the estimated coefficient.

Acknowledgments

Franco Basso gratefully acknowledges the financial support from both the Complex Engineering Systems Institute, ISCI (grant ANID PIA/PUENTE AFB230002) and from ANID-FONDECYT 1241335. Raúl Pezoa acknowledges the financial support from ANID-FONDECYT 1241913. Mauricio Varas acknowledges the financial support from ANID-FONDECYT 1241843.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Asociación Nacional de Investigación y Desarrollo [grant number: FONDECYT 1241335]; Asociación Nacional de Investigación y Desarrollo [grant number: FONDECYT 1241843]; Asociación Nacional de Investigación y Desarrollo [grant number: FONDECYT 1241913].

References

Adanu, E., A. Lidbe, E. Tedla, and S. Jones. 2021. "Factors Associated with Driver Injury Severity of Lane Changing Crashes Involving Younger and Older Drivers." *Accident Analysis & Prevention* 149: 105867.

Allen, B., B. Shin, and P. Cooper. 1978. "Analysis of Traffic Conflicts and Collisions." *Transportation Research Record* 667: 67–74.

Arun, A., M. M. Haque, A. Bhaskar, S. Washington, and T. Sayed. 2021. "A Bivariate Extreme Value Model for Estimating Crash Frequency by Severity Using Traffic Conflicts." *Analytic Methods in Accident Research* 32: 100180.

Bai, L., P. Liu, Y. Guo, and C. Xu. 2015. "Understanding Factors Affecting Frequency of Traffic Conflicts Between Electric Bicycles and Motorized Vehicles at Signalized Intersections." *Transportation Research Record* 2514 (1): 68–78.

Basso, F., L. J. Basso, F. Bravo, and R. Pezoa. 2018. "Real-time Crash Prediction in an Urban Expressway Using Desaggregated Data." *Transportation Research Part C: Emerging Technologies* 86: 202–219.

Basso, F., L. J. Basso, and R. Pezoa. 2020. "The Importance of Flow Composition in Real-Time Crash Prediction." *Accident Analysis & Prevention* 137: 105436.

Basso, F., Á Cifuentes, F. Cuevas-Pavincich, R. Pezoa, and M. Varas. 2022. "Assessing Influential Factors for Lane Change Behavior Using Full Real-World Vehicle-by-Vehicle Data." *Transportation Letters* 14 (10): 1126–1137.

Basso, F., R. Pezoa, M. Varas, and M. Villalobos. 2021. "A Deep Learning Approach for Real-Time Crash Prediction Using Vehicle-by-Vehicle Data." *Accident Analysis & Prevention* 162: 106409.

Comisión Nacional de Seguridad de Tránsito. 01 de agosto de 2018. *Libro del nuevo conductor*. Obtenido de https://www.conaset.cl/wp-content/uploads/2018/08/LNC-AUTOMOVILISTAS_ESPA%C3%91OL-01-08-2018.pdf

Comisión Nacional de Seguridad de Tránsito. mayo de 2021. *Siniestralidad vial y jóvenes*. Obtenido de https://www.conaset.cl/wp-content/uploads/2021/05/Costos-accidentes-2020.pdf

Comisión Nacional de Seguridad de Tránsito. 2022a. *Datos ordenados según los tipos y causas de carabineros*. Obtenido de Evolución de siniestros de tránsito Chile (1972-2021): https://www.conaset.cl/programa/observatorio-datos-estadistica/biblioteca-observatorio/estadisticas-generales/.

Comisión Nacional de Seguridad de Tránsito. 2022b. *Datos ordenados según los tipos y causas de carabineros.* Obtenido de Causas de siniestros (2000-2021): https://www.conaset.cl/programa/observatorio-datos-estadistica/biblioteca-observatorio/estadisticas-generales/.

Cooper, F., and N. Ferguson. 1976. "Traffic Studies at T-Junctions-A Conflict Simulation Model." *Traffic Engineering Control* 17: 306–309.

Das, S., and A. K. Maurya. 2019. "Defining Time-to-Collision Thresholds by the Type of Lead Vehicle in non-Lane-Based Traffic Environments." *IEEE Transactions on Intelligent Transportation Systems* 21 (12): 4972–4982.

Dimitriou, L., K. Stylianou, and M. Abdel-Aty. 2018. "Assessing Rear-end Crash Potential in Urban Locations Based on Vehicle-by-Vehicle Interactions, Geometric Characteristics and Operational Conditions." *Accident Analysis & Prevention* 118: 221–235.

Essa, M., and T. Sayed. 2018. "Traffic Conflict Models to Evaluate the Safety of Signalized Intersections at the Cycle Level." *Transportation Research Part C: Emerging Technologies* 89: 289–302.

Formosa, N., M. Quddus, S. Ison, M. Abdel-Aty, and J. Yuan. 2020. "Predicting Real-Time Traffic Conflicts Using Deep Learning." *Accident Analysis & Prevention* 136: 105429.

Guo, Y., P. Liu, L. Bai, C. Xu, and J. Chen. 2014. "Red Light Running Behavior of Electric Bicycles at Signalized Intersections in China." *Transportation Research Record* 2468(1): 28–37.

Guo, Y., T. Sayed, and M. Essa. 2020. "Real-time Conflict-Based Bayesian Tobit Models for Safety Evaluation of Signalized Intersections." *Accident Analysis ans Prevention* 144: 105660.

Güttinger, V. 1984. "Conflict Observation in Theory and in Practice." International Calibration Study of Traffic Conflict Techniques, 17–24.

Hayward, J. 1971. *Near Misses as a Measure of Safety at Urban Intersections*. Pennsylvania Transportation and Traffic Safety Center.

Hossain, M., M. Abdel-Aty, M. Quddus, Y. Muromachi, and S. Sadeek. 2019. "Real-time Crash Prediction Models: State-of-the-art, Design Pathways and Ubiquitous Requirements." *Accident Analysis & Prevention* 124: 66–84.

Hu, Y., Y. Li, H. Huang, J. Lee, C. Yuan, and G. Zuo. 2022. "A High-Resolution Trajectory Data Driven Method for Real-Time Evaluation of Traffic Safety." *Accident Analysis & Prevention* 165: 106503.

Huang, H., Y. Wei, C. Han, J. Lee, S. Mao, and F. Gao. 2022. "Travel Route Safety Estimation Based on Conflict Simulation." *Accident Analysis & Prevention* 171: 106666.

Kamel, A., T. Sayed, and C. Fu. 2023. "Real-time Safety Analysis Using Autonomous Vehicle Data: A Bayesian Hierarchical Extreme Value Model." *Transportmetrica B: Transport Dynamics* 11 (1): 826–846.

Li, X. Q., J. Jiang, and J. Lu. 2011. "Safety Differences Between Novice and Experienced Drivers Under Car-Following Situations." ICCTP 2011: Towards Sustainable Transportation Systems, 2196–2207.

Ministerio de Salud. (Accessed March 25th, 2021). *Ministerio de Salud*. https://www.minsal.cl/reporte-covid-19-region-metropolitana-estara-en-cuarentena-total-a-partir-del-sabado-25-de-marzo/.

Mitra, S. 2009. "Spatial Autocorrelation and Bayesian Spatial Statistical Method for Analyzing Intersections Prone to Injury Crashes." *Transportation Research Record* 2136 (1): 92–100.

Nadimi, N., S. S. NaserAlavi, and M. Asadamraji. 2022, December. "Calculating Dynamic Thresholds for Critical Time to Collision as a Safety Measure." In *Proceedings of the Institution of Civil Engineers-Transport*. (Vol. 175, No. 7, pp. 403-412). Thomas Telford Ltd.

Organización Mundial de la Salud. 20 de junio de 2022. *Traumatismos causados por el tránsito*. Obtenido de https://www.who.int/es/news-room/fact-sheets/detail/road-traffic-injuries

Orsini, F., G. Gecchele, R. Rossi, and M. Gastaldi. 2021. "A Conflict-Based Approach for Real-Time Road Safety Analysis: Comparative Evaluation with Crash-Based Models." *Accident Analysis & Prevention* 161: 106382.

Pinnow, J., M. Masoud, M. Elhenawy, and S. Glaser. 2021. "A Review of Naturalistic Driving Study Surrogates and Surrogate Indicator Viability Within the Context of Different Road Geometries." *Accident Analysis & Prevention* 157: 106185.

Shelby, S. G. 2011. "Delta-V as a Measure of Traffic Conflict Severity." *3rd International Conference on Road Safety and Simulation*, 14–16.

Stylinaou, K., L. Dimitriou, and M. Abdel-Aty. 2019. "Big Data and Road Safety: A Comprehensive Review." *Mobility Patterns. Big Data and Transport Analytics*, 297–343.

Theofilatos, Athanasios, Cong Chen, and Constantinos Antoniou. 2019. "Comparing Machine Learning and Deep Learning Methods for Real-Time Crash Prediction." *Transportation Research Record* 2673 (8): 169–178.

Wang, C., Y. Xie, H. Huang, and P. Liu. 2021. "A Review of Surrogate Safety Measures and Their Applications in Connected and Automated Vehicles Safety Modeling." *Accident Analysis & Prevention* 157: 106–157.

Wang, L., L. Zou, M. Abdel-Aty, and W. Ma. 2023. "Expressway Rear-end Crash Risk Evolution Mechanism Analysis Under Different Traffic States." *Transportmetrica B: Transport Dynamics* 11 (1): 510–527.

Wouters, P. I., and J. M. Bos. 2000. "Traffic Accident Reduction by Monitoring Driver Behavior with in-car Data Recorders." *Accident Analysis & Prevention* 32: 643–650.

Yang, K., R. Yu, X. Wang, M. Quddus, and L. Xue. 2018. "How to Determine an Optimal Threshold to Classify Real-Time Crash-Prone Traffic Conditions?" *Accident Analysis & Prevention* 117: 250–261.

Yuan, C., Y. Li, H. Huang, S. Wang, Z. Sun, and Y. Li. 2022a. "Using Traffic Flow Characteristics to Predict Real-Time Conflict Risk: A Novel Method for Trajectory Data Analysis." *Analytic Methods in Accident Research* 35: 100217.

Yuan, C., Y. Li, H. Huang, S. Wang, Z. Sun, and H. Wang. 2022b. "Application of Explainable Machine Learning for Real-Time Safety Analysis Toward a Connected Vehicle Environment." *Accident Analysis & Prevention* 171: 106681.

Yuan, Q., T. Zhang, X. Xu, B. Scott-Parker, and O. Viedo-Tresoalacios. 2021, October. "Comparisons in Crash Rate and Risky Behaviors Between Novice Drivers and Experienced Drivers in China." In *2021 6th International Conference on Transportation Information and Safety (ICTIS)*, 1448–1457. IEEE.

Zheng, L., and T. Sayed. 2020. "A Bivariate Bayesian Hierarchical Extreme Value Model for Traffic Conflict-Based Crash Estimation." *Analytic Methods in Accident Research* 25: 100111.

Zheng, L., T. Sayed, and F. Mannering. 2021. "Modeling Traffic Conflicts for use in Road Safety Analysis: A Review of Analytic Methods and Future Directions." *Analytic Methods in Accident Research* 29: 100142.

Zheng, L., C. Zhu, Z. He, T. He, and S. Liu. 2018. *Empirical Validation of Vehicle Type-Dependent car-Following Heterogeneity from Micro-and Macro-Viewpoints*. Transportmetrica B: Transport Dynamics.