Accident Analysis and Prevention 152 (2021) 106007

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Prediction of rear-end conflict frequency using multiple-location traffic parameters

Christos Katrakazas [a,*], Athanasios Theofilatos [b], Md Ashraful Islam [c], Eleonora Papadimitriou [d], Loukas Dimitriou [e], Constantinos Antoniou [f]

[a] *Department of Transportation Planning & Engineering, School of Civil Engineering, National Technical University of Athens, 15773, Greece*
[b] *School of Architecture, Civil and Building Engineering, Loughborough University, Ashby Road, Loughborough, LE11 3TU, United Kingdom*
[c] *Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Munich, 80333, Germany*
[d] *Faculty of Technology, Policy and Management, Technical University of Delft, Jaffalaan 5, Delft, 2628 BX, Netherlands*
[e] *Lab for Transport Engineering, Department of Civil and Environmental Engineering, University of Nicosia, Nicosia, 2111, Cyprus*
[f] *Chair of Transportation Systems Engineering, Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Munich, 80333, Germany*

ARTICLE INFO

Keywords:
Traffic conflicts
Safety
Surrogate safety measures
Count data modelling

ABSTRACT

Traffic conflicts are heavily correlated with traffic collisions and may provide insightful information on the failure mechanism and factors that contribute more towards a collision. Although proactive traffic management systems have been supported heavily in the research community, and autonomous vehicles (AVs) are soon to become a reality, analyses are concentrated on very specific environments using aggregated data. This study aims at investigating –for the first time- rear-end conflict frequency in an urban network level using vehicle-to-vehicle interactions and at correlating frequency with the corresponding network traffic state. The Time-To-Collision (TTC) and Deceleration Rate to Avoid Crash (DRAC) metrics are utilized to estimate conflict frequency on the current network situation, as well as on scenarios including AV characteristics. Three critical conflict points are defined, according to TTC and DRAC thresholds. After extracting conflicts, data are fitted into Zero-inflated and also traditional Negative Binomial models, as well as quasi-Poisson models, while controlling for endogeneity, in order to investigate contributory factors of conflict frequency. Results demonstrate that conflict counts are significantly higher in congested traffic and that high variations in speed increase conflicts. Nevertheless, a comparison with simulated AV traffic and the use of more surrogate safety indicators could provide more insight into the relationship between traffic state and traffic conflicts in the near future.

1. Introduction

The safety performance of a transportation network is directly expressed by the number of occurred collisions. In order to evaluate safety performance, researchers and practitioners usually correlate crashes with macroscopic traffic characteristics (e.g. Average Annual Daily Traffic; AADT) and geometrical or environmental attributes in order to predict an estimate of collision frequency and identify collision contributory factors on a link, an intersection or more generally an entire transportation network (Abdel-Aty and Pande, 2007). As it can be understood, data for such analyses are aggregated and therefore information on the failure mechanism that leads to collisions might be lost in the aggregation. Furthermore, as proactive traffic management systems gain attention in the research and practitioners community (Hossain et al., 2019), collisions can be efficiently predicted for a short-time window in real-time using microscopic data. Nevertheless, even in real-time safety modelling, aggregation of data causes problems in efficiently understanding conditions that lead to collisions (Roshandel et al., 2015).

To overcome the issue of data aggregation, and the lack of high-quality microscopic data, recent studies (Dimitriou et al., 2018; Katrakazas et al., 2018; Stylianou and Dimitriou, 2018) have been oriented towards investigating the relationship between traffic conflicts and collisions for microscopic collision modelling. Traffic conflicts are scenarios of dangerous vehicle interactions, which will evolve into collisions if no action is taken and the motion of vehicles remains

* Corresponding author.
E-mail addresses: ckatrakazas@mail.ntua.gr (C. Katrakazas), a.i.theofilatos@lboro.ac.uk (A. Theofilatos), Ashraful.islam@tum.de (M.A. Islam), e.papadimitriou@tudelft.nl (E. Papadimitriou), lucdimit@ucy.ac.cy (L. Dimitriou), c.antoniou@tum.de (C. Antoniou).

https://doi.org/10.1016/j.aap.2021.106007
Received 12 December 2019; Received in revised form 9 December 2020; Accepted 17 January 2021
Available online 5 February 2021

uninterrupted (Hyden, 1987). As traffic conflicts are a popular technique for exploiting traffic microsimulation in safety analyses (Young et al., 2014), they have also become one of the most cost-effective tools for investigating the safety effect of Autonomous Vehicles (AVs) (Papadoulis et al., 2019).

The state-of-the-art in conflict frequency prediction is limited in corridor-level analyses (Papadoulis et al., 2019), simulation-based traffic data (Katrakazas et al., 2018; Papadoulis et al., 2019) or does not take effects of autonomous vehicles into account (Dimitriou et al., 2018; Stylianou and Dimitriou, 2018). This paper aims to add to current knowledge, by attempting to correlate traffic conflicts on a transportation network-level with the traffic state of network links and develop conflict prediction models according to the characteristics of the network. To that aim, detailed Vehicle-by-Vehicle (VbV) data from 16 inductive loop detectors in the urban network of Nicosia, Cyprus are used in order to extract rear-end conflicts using the Time-To-Collision (TTC) and Deceleration Rate to Avoid a Crash (DRAC) metrics. Following the estimation of all conflicts, traffic and network attribute variables become input to Zero inflated Poisson (ZIP) and Zero-inflated Negative Binomial (ZINB) models, in order to predict conflict frequency. Furthermore, a discussion on the potential impact on conflicts if traffic was fully automated (i.e. SAE Level 5; SAE International, 2016) is attempted.

The paper is structured as follows: initially the literature with regard to conflicts, frequency modelling and autonomous vehicles is reviewed, and the methodology of the current work is presented. This is followed by a detailed description of the data used and the way the conflicts were extracted for all scenarios. Finally, conflict frequency models are developed and the results are discussed in order to draw conclusions for researchers and practitioners.

2. Literature review

Traffic conflicts have been used in numerous safety studies aiming to evaluate the association of traffic characteristics and vehicle trajectory data on collision risk (Dimitriou et al., 2018; Oh et al., 2006; Weng and Meng, 2011) usually focusing on longitudinal data to assess rear-end collisions. Numerous simulator studies also use Surrogate Safety Measures (SSMs) such as TTC for collision avoidance or incident reaction time studies (e.g. Payre et al., 2016). Traffic conflicts have recently become a tool to study various human factors involved in automation. For instance, a traffic conflicts study (de Winter et al., 2019) showed that in a car-following field experiment (test vehicle following a robot-controlled lead vehicle), participants who had a collision event had mis-aligned high levels of trust to the automated car. As AVs aim to bring dramatic safety improvements by minimising the role of human drivers and thus eliminating human error, safety is the main “banner value” to promote AVs technology. Nevertheless, existing evidence is insufficient and the positive safety impacts of AVs remain to be proved (Calvert et al., 2017; Crist and Voege, 2018; Kalra and Paddock, 2016).

Given the unavailability of actual collision data and the need for ex-ante evaluation of the expected safety impacts new technologies, traffic conflicts techniques and SSMs come with advantages due to their higher event frequency and observability and a robust conceptual and actual correlation with collisions (El-Basyouny and Sayed, 2013; Minderhoud and Bovy, 2001). SSMs allow to estimate collision risk on the basis of information on vehicles’ basic kinematics (e.g. position, velocity, acceleration; Dimitriou et al., 2018). Indicators typically used are TTC, Time Exposed Time to Collision (TET), Time Integrated Time to Collision (Kuang et al., 2015), deceleration rates, as well as probabilistic indicators, such as the Crash Potential Index (CPI) (Mullakkal-Babu et al., 2017)– which are however restricted to specific driving regimes (Oh and Kim, 2010).

The majority of studies concerned with traffic conflicts are aiming at the estimation of traffic conflicts and the development of Safety Performance Functions (SPFs) based on these estimations (Charly and Mathew, 2019; Essa and Sayed, 2019, 2018). Another group of traffic conflict studies are looking into the crash potential of a site based on the number of conflicts(Rahman et al., 2019; Xing et al., 2019). Fewer studies utilize the extracted conflicts for real-time safety assessment (Dimitriou et al., 2018; Katrakazas et al., 2018; Papadoulis et al., 2019). Nevertheless, a common characteristic of all these studies is that their study area is limited to the site of an intersection (Essa and Sayed, 2019, 2018; Ulak et al., 2019), toll facilities (Xing et al., 2019), work zones (Weng and Meng, 2011), tunnels (Meng and Qu, 2012) or restricted freeway segments (Katrakazas et al., 2018; Kuang et al., 2015). The only study that overviewed a network-wide effect was the one of (Dimitriou et al., 2018), but their focus was mostly on crash-potential, rather than the frequency of conflicts and the correlation with the traffic states of the network. Hence, it was demonstrated that a study that predicts conflict frequency with regards to the traffic state of an entire urban network is yet to be realized.

Methodologically, modelling of the frequency of traffic conflicts follows the rationale of collision frequency modelling. Traditionally, this is carried out on the basis of Safety Performance Functions relating collision counts to traffic risk exposure, expressed on the basis of AADT or similar traffic information (Hauer, 1995). Road collisions are modelled as Poisson-family-distributed rare events, however in recent years considerable research is carried out to identify techniques that are pertinent for handling specific properties of collision counts. Negative Binomial Models to handle over-dispersion in collision counts are the most common specifications (Lord et al., 2005), with several extensions allowing to handle zero-inflated counts (Lord et al., 2007) or heterogeneity (Karlaftis and Tarko, 1998; Quddus, 2008), often in a Bayesian context (Aguero-Valverde, 2013). Latent variables representations are also proposed (Castro et al., 2012). An important amount of recent work lies on hierarchical or multilevel modelling approaches (Dupont et al., 2013; Huang and Abdel-Aty, 2010) allowing to handle spatial, temporal or other unobserved dependences in collision counts.

3. Data description and conflict extraction

The available VbV data were obtained from a large-scale urban network, located in Nicosia, Cyprus. The data were collected from 16 Inductive Loop Detectors (ILD), which cover either one, two or three lanes in each direction (depending on the carriageway design). The locations of the ILDs within the city of Nicosia is depicted in Fig. 1. Data were available for a two-month period (18/09/2015 to 18/11/2015), totalling more than 21 million observations. Due to the large size of the raw observations, traffic data were aggregated over 5 min intervals, as this is usually the aggregation level used in microscopic safety analyses (Roshandel et al., 2015). From the database, the flow, speed, headway, gap and vehicle composition information were extracted and aggregated into the 5-minute intervals.

The next step of the processing procedure included the extraction of additional variables to be used for rear-end conflict estimation according to (Dimitriou et al., 2018) and conflict frequency modelling. These variables included:

f0d8; Traffic Density (calculated through $\frac{Flow}{Speed}$),

f0d8; Traffic state (Free Flow, Unstable flow, Congestion) using the fundamental diagram of flow and density and the following thresholds
- o Free Flow, when Density < 20veh/km and Flow < 80 vehicles
- o Transitional Flow, when 20< Density < 45 veh/km and Flow >80 vehicles
- o Congestion, when Density is > 45 veh/km and Flow>80 vehicles

It should be noted here that the roads’ crossections have similar characteristics, the thresholds are data-driven and were obtained after

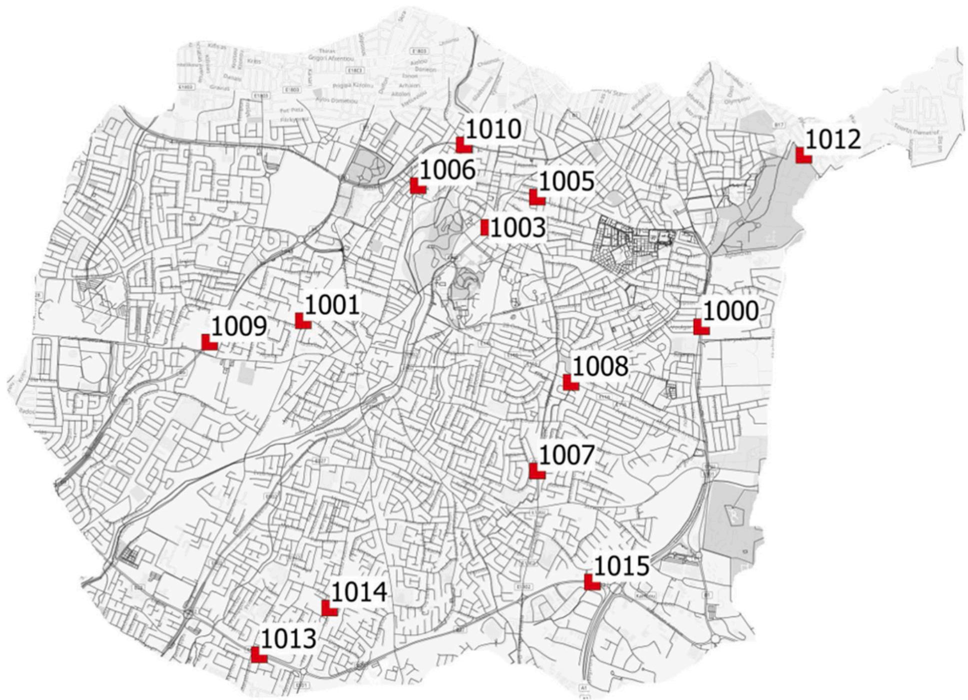

Fig. 1. The location of the ILDs within the city of Nicosia, Cyprus.

plotting the fundamental diagram for each of the detectors as shown in Fig. 2.

Furthermore, for extracting conflicts, the measures of TTC and DRAC were calculated. TTC was chosen, because it is one of the most popular surrogate safety measures for detecting rear-end conflicts and DRAC was chosen due to its explicit recognition as a safety performance metric (Cunto and Saccomanno, 2008), which considers speed differentials and deceleration profiles during dangerous vehicle encounters (Archer, 2005). DRAC reflects the necessary deceleration profile needed to make a timely stop and hence avoid a rear-end collision. The two metrics are calculated as follows:

- TTC= $\frac{X_i^t - X_{i-1}^t - L_{i-1}}{V_i^t - V_{i-1}^t}$, where X_i^t is the position of the following vehicle *i* in time *t*, V_i^t is its speed at time *t* and L_i is the length of a vehicle. The corresponding variables with the notation *i-1* (i.e. X_{i-1}^t and V_{i-1}^t) denote the position and speed of the leading vehicle passing from the same detector.
- DRAC= $\frac{\left(V_i^t - V_{i-1}^t\right)^2}{2\left[\left(X_i^t - X_{i-1}^t\right) - L_{i-1}\right]}$

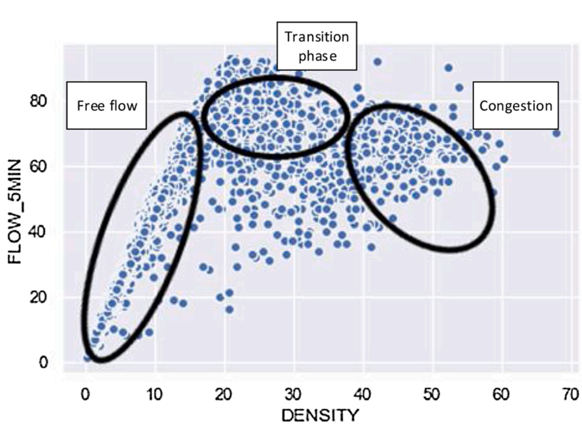

Fig. 2. Example of defining traffic state based on the flow-density diagram.

As the provided data were VbV, the lengths, distances and speeds of consecutive vehicles could easily be obtained. TTC and DRAC were chosen as the most representative indicators because they are heavily used in conflict-based safety analyses including potential automated vehicle traffic scenarios (e.g. Dijkstra, 2013; El-Basyouny and Sayed, 2013; Essa and Sayed, 2019; Papadoulis et al., 2019; Rahman and Abdel-aty, 2018).

Table 1 summarizes the descriptive statistics of the variables recorded and estimated from the available dataset.

In order to extract rear-end conflicts, for every vehicle interaction near the location of the detectors, appropriate thresholds were applied to the TTC and DRAC in order to get the most "dangerous" scenarios of interactions between vehicles. These thresholds were chosen in order to estimate conflicts in typical scenarios, as well as scenarios that "emulate" a fully and highly automated traffic, i.e. they are harsher thresholds for typical traffic. These stricter thresholds were applied, as it

Table 1
Descriptive Statistics of Traffic Variables.

Variable	Description	Mean	Std. Dev
Flow	Number of vehicles aggregated in 5 min interval per cross section	41.120	17.294
Spot Speed	Individual vehicle speed (km/h)	51.930	12.708
Headway	Temporal headway between two consecutive vehicles (in seconds)	12.729	28.934
Gap	Distance between two consecutive vehicles (in meters)	57.045	36.713
DRAC	Deceleration rate to avoid a crash (in m/s^2)	1.060	26.308
TTC	Time to Collision (in seconds)	2.311	82.272
		Percentage	
Vehicle Composition (%)	Cars and small vans	96.642	
	HGVs and Buses	1.890	
	Motorcycles	1.444	
	Unclassified vehicles	0.024	
Traffic State Cases (%)	Congestion	1.83	
	Free Flow	88.78	
	Unstable	9.39	
Total Observations	21,714,292		

is envisioned that critical scenarios for automated vehicles will be the ones dealing with very small headways with preceding vehicles as well as rapid evasive manoeuvres leading to harshest decelerations to avoid a collision (Papadoulis et al., 2019; Rahman et al., 2019).

Therefore, three critical points were defined in order to be used for the analysis:

- o **Conflict point A**: TTC < 1.5 s and DRAC > 3 m/s^2, taken from Deligianni et al. (2017); Dimitriou et al. (2018); Li et al. (2017)
- o **Conflict point B**: TTC < 1 s and DRAC > 6 m/ s^2, obtained from fitting a gaussian distribution to the data, and taking values for TTC and DRAC that did not belong to the 90 % confidence interval
- o **Conflict point C**: TTC < 0.5 s and DRAC > 10 m/sec^2, taken from extreme values used in Asljung et al. (2017); Deligianni et al. (2018); Mahmassani (2016); Rummelhard et al. (2016), which investigated the safety impacts of automated driving.

The conflict cases for each critical point, were summarized per 5-minute interval in order to correspond to the 5-minute interval of the traffic variables captured.

4. Methods of analysis

4.1. Count data modelling

As observed from the literature review, in order to model the frequency of conflicts in the urban network, collision frequency modelling deems as the most appropriate, since conflicts and collisions share similar failure mechanisms. Since collision frequency data, belong to the count data category, and more specifically to non-negative count data, linear regression modelling is inappropriate, and other approaches such as Poisson regression, negative binomial, zero-inflated Poisson regression and zero-inflated negative binomial regression have become the state-of-the-art in modelling such data (Washington et al., 2010).

A notable characteristic of crash frequency data is that the variance usually exceeds the mean of the crash counts (Lord and Mannering, 2010). When overdispersed data are present, the negative binomial as well as the quasi-Poisson model can be used to overcome this issue. The variance of a quasi-Poisson model is a linear function of the mean while the variance of a negative binomial model is a quadratic function of the mean. However, when the mean equals variance, the Poisson model is used.

Due to the serious prevalence of zeros in such crash databases, zero-inflated models have emerged in modelling collision frequency, as they address the excessive zero density by leading the modelling procedure into a normal-operations vs collision prone propensity condition for every road segment. In zero-inflated negative binomial models, each observation is assumed to have two possible states; *State 1* which if true denotes that collision counts are zero, and *State 2*, which when true, initiates the generation of traffic counts according to the negative binomial distribution. If π denotes the probability of occurrence for State 1, and hence the probability of State 2 occurrence is *1-*π, the probability distribution of a zero-inflated negative binomial random variable y_i can be estimated as:

$$\Pr(y_i = j) = \begin{cases} \pi_i + (1+\pi_i) * g(y_i = 0) \ if \ j = 0 \\ (1+\pi_i) * g(y_i) \ if \ j > 0 \end{cases} \tag{1}$$

where π_i is the logistic link function defined below and $g(y_i)$ is the negative binomial distribution given by:

$$g(y_i) = \Pr(Y = y_i \mu_i, a) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i+1)} \left(\frac{1}{1+\alpha\mu_i}\right)^{a^{-1}} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \tag{2}$$

After defining the probability distribution of the zero-inflated negative binomial variable, the probability of a roadway segment or entity to be in zero or non-zero state can be determined by a binary logit or probit model (Lambert, 1992; Washington et al., 2010).

However, it is not always obvious that the zero-inflated models are appropriate. As a consequence, statistical tests, such as the Vuong test (Vuong, 1989) provide more insight on the selection of the most appropriate model when dealing with excessive zeros in crash counts (Chen et al., 2016; Jiang et al., 2011)

Summing up, this study explores a series of Negative Binomial (NB), quasi-Poisson as well as and zero inflated Negative Binomial (ZINB) models after testing the datasets using the Vuong test. Model estimations were carried out by using the R-package *pscl* (Jackman, 2017). The Vuong test in our study is conducted by using the R-package *nonnest2* (Merkle and You, 2018).

4.2. Controlling for endogeneity

Lord and Mannering (2010) critically discuss endogeneity issues in crash data modelling, stating that when endogenous explanatory variables are included in models their values may depend on the frequency of crashes. Similar issues are likely to arise when analyzing traffic conflicts. Although it is relatively straightforward to account endogeneity in ordinary least-squares models (Lord and Mannering, 2010), considerable complexity is added to count-data models when attempting to address endogeneity (Kim and Washington, 2006). To the best of our knowledge, the present research paper is one of the few studies attempting to address this issue when dealing with crash counts.

One popular way to deal with unmeasured confounding is to use Instrumental Variable (IV) methods. The following methodological process is adopted from Sjolander and Martinussen (2019). First, let us assume that Z and X are the IV and the exposure respectively, whilst L is a set of covariates that will be controlled for the analysis. 1d44c;0 is the potential outcome, when the exposure is set to 0 (Pearl, 2009). Afterwards, the following causal model is considered:

$$\{E(Y|L,\ Z,\ X)\}\ - \{(Y0|L,\ Z,\ X)\}\ =\ m\,T\,(L)\,X\psi \tag{3}$$

where η is either the identity, log or logit link, and the vector function m (L) allows for interactions between X and L. The vector parameter ψ measures the causal effect of a particular exposure level conditionally on (L, Z).

In this paper, a two-step process is followed in order to apply IV analysis for traffic conflict modelling. According to Sjolander and Martinussen (2019), in the first stage, a regression model is fitted for the exposure, using L and Z as regressors. This regression model is used to formulate a prediction $X = E(X|Z,\ L)$ for each subject. In the second stage, another regression model is fitted for the outcome, using L and $m(L)X$ as regressors.

Moreover, in order to account for potential biases arising from this method (Greenland et al., 1999), a control function $R = X - E(X|Z,\ L)$ is added in the second stage model (Tchetgen Tchetgen et al., 2015; Vansteelandt et al., 2011).

The two-stage method is computationally feasible and can be extended to generalized linear models, such as negative binomial and zero-inflated models. IV models were estimated by using the *ivtools* package (Sjolander and Martinussen, 2019) in R software.

5. Results

5.1. Preliminary analysis

Firstly, a preliminary analysis and visual inspection of the dependent variables was carried out. More specifically, the distributions of conflicts were plotted. Fig. 3 illustrates the distribution of all three conflict points (i.e. A, B and C). In both figures, it is observed that there are excessive zeros in conflicts. Hence, the zero inflated specification is appropriate for modelling the number of conflicts. The mean and variance for the

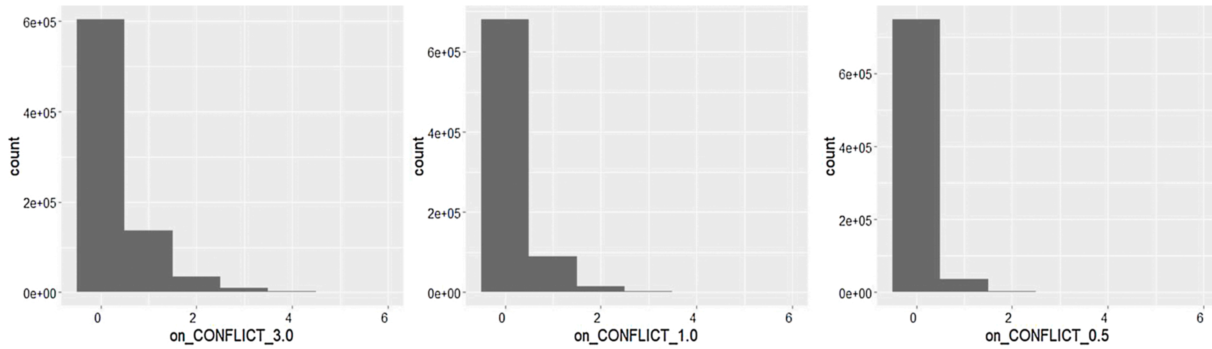

Fig. 3. Probability Distribution of conflict cases for conflict points A (left), B (center) and C (right).

conflict point A conflicts were 0.319 and 0.452, whilst the respective numbers for conflict point B were 0.169 and 0.225 and for conflict point C were 0.059 and 0.074. Because the means and variances do not differ in a great extent, both Poisson and Negative Binomial specifications will be tested.

As a second step, all candidate independent variables must be tested for potential correlations. The results of the correlation matrix for the continuous variables are illustrated in Fig. 4 and assisted in selecting the variables which had no correlation between each other. In order to test the correlation of discrete and continuous variables, the Kruskal-Wallis test which is a relevant non-parametric method was utilized. The Kruskal-Wallis test showed no correlation among the Traffic state variable and the continuous variables.

5.2. Conflict prediction modelling

The following tables provide a summary of the best fitting models for conflict points A, B and C respectively. As stated earlier, both Negative Binomial, quasi-Poisson and Poisson models were considered; models having the best fit were ultimately retained. Various random effects model specifications were also tested, however, the variance of the random effects (intercepts and slopes) were not significant in any of the three models. Hence, only fixed effect models were included. In addition to AIC and BIC test, the Vuong test was also used to provide insight on the selection of the most appropriate structure (typical or the Zero Inflated structure). The low VIF values in all six models showed that no multicollinearity issues were present in our analyses.

5.2.1. Conflict point A

For conflict point A, the alternative hypothesis that the Zero Inflated model is better, was accepted (Z-value = 15.209, p-value < 0.001). Hence, for conflict point A the ZINB was the best-fitting model. Table 2 summarizes the ZINB model for rear-end conflict point A cases (i.e. TTC < 3.0 s). VIF tests indicated low collinearity, as the values were lower than five. It is noted that in the conditional part of the model, VIF test is not applicable because there is only one predictor.

The likelihood ratio test was significant at 95 % level, indicating an adequate fit. The conditional model (count model) suggests that free flows are associated with fewer conflicts, whilst unstable flow state cause an increase to conflict likelihood. In other words, congested and unstable flows increase conflict numbers. Although there is not much research on the impact of congested conditions (on a real-time basis) on collision frequency, our findings can be considered in line with a few past studies regarding real-time collision risk and congested conditions such as high occupancy (Xu et al., 2013; Yu and Abdel-Aty, 2013). Moreover, a non-linear impact of flow on collision numbers have been recently suggested by Yu et al. (2019). However, Yu et al. (2019), found that moderate levels of traffic volume have higher crash probabilities. Regarding unstable flow state, this finding is consistent with past

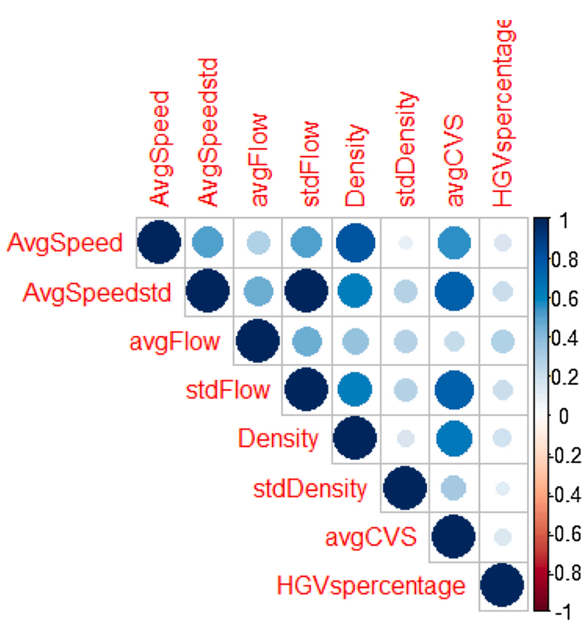

Fig. 4. Correlation Matrix for Continuous Independent Variables.
(Avg: Average, Std: Standard deviation, CVS: Coefficient of variation of Speed, HGV: Heavy Goods Vehicles)

Table 2
Summary of the ZINB model for conflict point A cases frequency.

	Variable	Beta coefficient	Standard error	p-value	VIF
Conditional model	Constant term	−1.766	0.018	<0.001 ***	–
	Traffic_state (Free flow)	−0.197	0.016	<0.001 ***	
	Traffic_state (Unstable flow)	0.132	0.017	<0.001 ***	1.14
	Traffic_state (Congested flow)	ref.cat	ref.cat	ref.cat	
	Flow	0.029	1.17E-04	<0.001 ***	1.14
Zero-inflation model	Constant term	−3.130	0.038	<0.001 ***	–
	AvgSpeed	0.032	4.10E-04	<0.001 ***	–
Dispersion parameter	Log(theta)	1.449048	0.029522	<0.001 ***	–
Log-likelihood at zero	−575922.90				
Log-likelihood of the model	−521200.00				

***: Significant at 99 %, **: Significant at 95 %, *: Significant at 90 %, n.s.: non-significant.

literature in the field of real-time collision risk as traffic variations are considered as a very common risk factor (Ahmed et al., 2012; Theofilatos, 2017; Xu et al., 2013; Yu and Abdel-Aty, 2013).

When considering the zero state, several interesting remarks can be observed. The positive sign of the beta coefficient of average speed implies a negative correlation with increased rear-end conflict numbers. In other words, as average speeds rise, rear-end conflict numbers are more probable to be lower. This is in-line with previous research findings on the relationship between speed and traffic conflicts(Tarko, 2020; Xing et al., 2019)

When controlling for potential endogeneity, a few notable differences are observed (see Table 3). As mentioned in the methodological framework, a two-step process was followed. In the first stage, the Flow variable was regressed on ttc, indicating a significant positive relationship (beta coefficient = −0.002, p-value = <0.001). Traffic state was not found to be influenced by ttc. In the second stage, a count data model was estimated to control for the existing endogenous relationships. The quasi-Poisson model had the best fit and the respective estimation results are illustrated in Table 3. The control function R which was used to reduce potential bias of the two-step process was significant, hence it is retained in the estimation process.

A few notable differences are identified in the IV analysis model. Traffic state variable is not found to statistically influence conflicts as observed in the zero-inflated model and was not retained in the final model. Secondly, the logarithm of flow is significant in the IV model, indicating that as flow decreases, more conflicts are expected to occur. Thirdly, a negative correlation between percentage of heavy vehicles and conflicts was identified when endogeneity was accounted for, possibly due to risk compensation by drivers when heavy vehicles are present in traffic. Although this factor is seldom explored in proactive safety evaluation, Theofilatos et al. (2018) found no influence of truck proportions in collision risk.

5.2.2. Conflict point B

In order to explore conflict point B cases frequency, the Vuong test and the AIC/BIC tests indicated that despite excessive zeros the traditional NB model had better goodness of fit, as both AIC and BIC values were lower than the Zero Inflated model. Moreover, according to the Vuong test, the alternative hypothesis that the traditional Negative Binomial model is better the Zero Inflated model was accepted (Z-value = −36.294, p-value < 0.001). The findings are very similar to the previous model for conflict point A, however the model structure is different, as the zero-state part does not exist in this case. Overall, it is suggested that congested flows, low speeds as well as variations in speed are associated with increased numbers of rear-end conflicts described by conflict point B. The impact of speed variation for conflict point B (which was not significant for conflict point A frequency model without accounting for endogeneity) is in accordance with past studies in the field (Theofilatos et al., 2017; Yu and Abdel-Aty, 2013). Interestingly, unstable flows are associated with fewer conflicts than congested flows. However, this finding should be interpreted with care, because the unstable flow is category is compared to the congested flow category, due to the categorical nature of traffic state variable.

When controlling for potential endogeneity, there are some differences compared to the NB model (see Table 4 and 5). The two-step process was followed, having traffic flow as endogenous variable regressed on TTC (p-value <0.001). The quasi-Poisson model had the best fit and the respective estimation results are illustrated in Table 5. When compared with the NB model without endogeneity, a few additional variables were found to significantly influence conflict counts. For example, the standard deviation of flow is positively correlated with conflict occurrence. On the other hand, this model suggests that high percentages of heavy vehicles are associated with fewer conflicts. Moreover, the logarithm of flow was found to decrease traffic conflicts. This finding might be considered contradictory with the impact of the traffic state variable; however, the traffic state variable is significant only at 90 % level. Generally, more research is needed towards that direction.

5.2.3. Conflict point C

The last model is concerned with the frequency of rear-end conflicts obtained by conflict point C (AVs conditions). The best fit was achieved by the traditional Negative Binomial specification. The most significant variables are the traffic state and percentage of heavy goods vehicles (HGV), average speed and the logarithm of the average coefficient of variation of speed. The findings are in line with the previous models, suggesting that free flows, high percentage of HGVs as well as increased speeds are associated with reduced numbers of point C rear-end conflicts. Similarly, increased variations in speed (as expressed by the coefficient variation of speed) increase traffic conflicts.

When attempting to control for potential endogeneity by regressing flow on TTC (as in conflict point A and B processes), the quasi-Poisson model had the best fit and the respective estimation results are illustrated in Tables 6 and 7. Firstly, traffic state was not significant. Moreover, the beta coefficient of heavy vehicle percentage was found to have

Table 3
Summary of the quasi-Poisson model with IVs for conflict point A cases frequency.

	Variable	Beta coefficient	Standard error	p-value	VIF
IV model	Constant term	11.724	3.040	<0.001 ***	–
	avgCVS	3.293	0.094	<0.001 ***	1.23
	log(Flow)	−6.214	1.302	0.001 ***	1.00
	HGVspercentage	−0.011	0.002	<0.001 ***	1.23
	R (control function)	0.047	3.41E-04	<0.001 ***	–
Log-likelihood at zero	−336000.00				
Log-likelihood of the model	−263000.00				

*** : Significant at 99 %, **: Significant at 95 %, *: Significant at 90 %, n.s.: non-significant.

Table 4
Summary of the NB model for conflict point B cases frequency.

	Variable	Beta coefficient	Standard error	p-value	VIF
Conditional model	Constant term	−1.558	0.023	<0.001 ***	–
	Traffic_state (Free flow)	−0.585	0.021	<0.001 ***	1.43
	Traffic_state (Unstable flow)	−0.074	0.021	<0.001 ***	
	Traffic_state (Congested flow)	ref.cat	ref.cat	ref.cat	
	log(avgCVS)	3.820	0.019	<0.001 ***	1.52
	AvgSpeed	−0.011	2.39E-04	<0.001 ***	1.78
Dispersion parameter	Theta	1.3728	0.0166	<0.001 ***	–
Log-likelihood at zero	−225401.50				
Log-likelihood of the model	−187715.00				

*** : Significant at 99 %, **: Significant at 95 %, *: Significant at 90 %, n.s.: non-significant.

Table 5
Summary of the quasi-Poisson model with IVs for conflict point B cases frequency.

	Variable	Beta coefficient	Standard error	p-value	VIF
IV model	Constant term	7.670	3.259	0.019**	–
	Traffic_state (Free flow)	−0.556	0.320	0.082*	
	Traffic_state (Unstable flow)	−0.549	0.331	0.097*	1.13
	Traffic_state (Congested flow)	ref.cat	ref.cat	ref.cat	
	stdFlow	0.057	0.003	<0.001 ***	2.82
	log(avgCVS)	2.186	0.105	<0.001 ***	2.81
	log(Flow)	−4.785	1.391	<0.001 ***	1.13
	HGVspercentage	−0.003	0.002	0.081*	1.22
	R (control function)	0.055	4.83E-04	<0.001 ***	–
Log-likelihood at zero	−53600.00				
Log-likelihood of the model	−42055.00				

*** : Significant at 99 %, **: Significant at 95 %, *: Significant at 90 %, n.s.: non-significant.

Table 6
Summary of the NB model for conflict point C cases.

	Variable	Beta coefficient	Standard error	p-value	VIF
Conditional model	Constant term	−2.304	0.033	<0.001 ***	–
	Traffic_state (Free flow)	−0.709	0.030	<0.001 ***	
	Traffic_state (Unstable flow)	−0.122	0.030	<0.001 ***	1.57
	Traffic_state (Congested flow)	ref.cat	ref.cat	ref.cat	
	HGVspercentage	−0.003	0.001	<0.001 ***	1.22
	AvgSpeed	−0.017	4.29E-04	<0.001 ***	2.05
	log(avgCVS)	4.020	0.027	<0.001 ***	1.77
Dispersion parameter	Theta	0.9182	0.0171	<0.001 ***	–
Log-likelihood at zero	−117383.5				
Log-likelihood of the model	−93231.00				

*** : Significant at 99 %, **: Significant at 95 %, *: Significant at 90 %, n.s.: non-significant.

a positive sign, contrary to the previous models. However, it is noted that flow (which is an endogenous variable) was not found to be significant, hence, endogeneity for conflict C cases might not be the case.

6. Conclusions

Collision frequency is the traditional indicator of the safety level of a transportation network. However, to-date, collision frequency prediction models fail in utilizing highly disaggregated vehicle-by-vehicle data and are usually concerned with restricted areas (e.g. limited sections of a highway, weaving sections or intersections). This paper aimed at

Table 7
Summary of the quasi-Poisson model with IVs for conflict point C cases frequency when controlling for endogeneity.

	Variable	Beta coefficient	Standard error	p-value	VIF
IV model	Constant term	−0.218	4.444	0.961 n.s.	–
	avgCVS	3.505	0.091	<0.001 ***	1.27
	HGVspercentage	0.005	0.002	0.024**	1.26
	log(Flow)	−1.959	1.902	0.303 n.s.	1.01
	R	0.050	0.001	<0.001 ***	–
Log-likelihood at zero	−23940				
Log-likelihood of the model	−19990				

*** : Significant at 99 %, **: Significant at 95 %, *: Significant at 90 %, n.s.: non-significant.

extending the state-of-the-art- in frequency modelling, by predicting traffic conflict frequency in an urban network-wide area according the prevailing traffic state and including scenarios that imitate fully automated traffic. The novelty of the work lies in the utilization of highly disaggregated Vehicle-by-Vehicle (VbV) traffic data obtained from 16 inductive loop detectors in Nicosia, Cyprus and containing more than 2 million observations. Using the widely utilized TTC and DRAC metrics, conflicts were extracted an afterwards Zero-Inflated Negative Binomial (ZINB), traditional Negative Binomial (NB) as well as quasi-Poisson models, were applied in order to predict conflict frequency in three scenarios; two for current vehicle settings and one that could be applied in a potential automated traffic scenario.

The model results demonstrated that VbV data can be effectively utilized for conflict frequency prediction and provided insight on the variables that influence conflict occurrence. More specifically, conflict numbers are increased especially during congested traffic conditions and are more likely to be fewer during free flow traffic. Furthermore, for two out of three types of conflicts (Points A and B), increased speeds led to fewer conflicts, while high variations of speeds in conflict point B and C led to more frequent conflicts. However, when attempting to address endogeneity, the results show a few notable differences exist. For instance, the impact of traffic flow is not clear as mixed findings occur. Secondly, unstable flow as expressed by traffic state was only important for conflicts described by conflict point B. Since it is the first study addressing the issue of endogeneity when modelling traffic conflicts, additional research is needed.

In the hypothetical scenario of fully automated traffic (conflict point C), high percentage of heavy vehicles was an additional parameter found to increase traffic conflicts. In this case, it seems that endogeneity is not occurring. It is noted that this scenario is developed on the basis of certain assumptions and threshold values, and since research in this field is evolving, it is possible that some of these assumptions are imperfect. However, the main purpose of this research is to demonstrate the potential for comparison of results between conventional traffic and traffic moving with greater decelerations and shorter headways, as it would be if traffic was fully automated. Of course, as the data utilized did not come from actual automated vehicles, results should be interpreted with care with regards to conflict point C. This type of analysis may eventually lead to the estimation of safety benefits of AVs compared to current traffic conditions. In this context, the microscopic analysis allowed by traffic conflicts estimation is clearly advantageous, as there is need for proactive estimation of the safety impacts of a level of automated traffic, before actual crashes happen.

Nevertheless, the current study is not without limitations. The data utilized came explicitly from loop detectors. Such data are usually discontinuous both spatially and temporally and are less informative that actual vehicle trajectories (Roshandel et al., 2015). In order to perform an actual comparison with simulated AV traffic to be validated along with the corresponding conflicts, conflicts extracted from micro-simulation need to be obtained. Moreover, the scope of the present study is focused to rear-end collisions, and further research should explore other critical conflict types in urban road networks e.g. intersections – provided that appropriate relevant data could be available. It would be also very interesting to examine intermediate penetration levels of AVs i. e. mixed traffic conditions and this will be pursued in future research. As only microscopic surrogate safety indicators were used (i.e. TTC and DRAC), driving behaviour could not be directly captured. Indicators corresponding to fatigue, impairment and distraction could enhance the connection between predicted conflicts and actual collisions. Furthermore, the utilization of other surrogate safety indicators such as modified TTC and DRAC (Charly and Mathew, 2019; Zheng et al., 2019) as well as controlling for either TTC or DRAC similar to the work of (Wang and Stamatiadis, 2016, 2014) would further validate the results of modelling and will be considered in future work.

This study utilized fixed effects models, because the random effects terms were not statistically significant. However, future studies should utilize more network information in order to account for potential spatial heterogeneity and autocorrelation. Lastly, Machine Learning and Deep Learning could also be utilized and provide further insights.

Declaration of Competing Interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

CRediT authorship contribution statement

Christos Katrakazas: Conceptualization, Methodology. **Athanasios Theofilatos:** Methodology, Software. **Md Ashraful Islam:** Software. **Loukas Dimitriou:** Conceptualization, Resources. **Constantinos Antoniou:** Supervision.

References

Abdel-Aty, M., Pande, A., 2007. Crash data analysis: collective vs. individual crash level approach. J. Safety Res. 38, 581–587.
Aguero-Valverde, J., 2013. Full bayes poisson gamma, poisson lognormal, and zero inflated random effects models: comparing the precision of crash frequency estimates. Accid. Anal. Prev. 50, 289–297.
Ahmed, M., Abdel-Aty, M., Yu, R., 2012. Assessment of the interaction between crash occurrence, mountainous freeway geometry, Real-time weather and AVI traffic data. TRB 2012 Annu. Meet. 2450.
Archer, J., 2005. Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: a study of urban and suburban intersections. Acad. thesis, R. Inst. Technol. Dep. Infrastructure, Div. Transp. Logist. Cent. Transp. Res.
Asljung, D., Nilsson, J., Fredriksson, J., 2017. Using extreme value theory for vehicle level safety validation and implications for autonomous vehicles. IEEE Trans. Intell. Veh. 2, 288–297.
Calvert, S.C., Schakel, W.J., van Lint, J.W.C., 2017. Will automated vehicles negatively impact traffic flow? J. Adv. Transp. 2017.
Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. Transp. Res. Part B Methodol. 46, 253–272.
Charly, A., Mathew, T.V., 2019. Estimation of traffic conflicts using precise lateral position and width of vehicles for safety assessment. Accid. Anal. Prev. 132, 105264.
Chen, F., Chen, S., Ma, X., 2016. Crash frequency modeling using real-time environmental and traffic data and unbalanced panel data models. Int. J. Environ. Res. Public Health 13, 1–16.
Crist, P., Voege, T., 2018. Safer Roads With Automated Vehicles?[1] P. Crist and T. Voege, "Safer Roads With Automated Vehicles?,", 2018, pp. 1–46, 1–46.
Cunto, F., Saccomanno, F.F., 2008. Calibration and validation of simulated vehicle safety performance at signalized intersections. Accid. Anal. Prev. 40, 1171–1179.
de Winter, J.C.F., Dodou, D., Happee, R., Eisma, Y.B., 2019. Will vehicle data be shared to address the how, where, and who of traffic accidents? Eur. J. Futur. Res. 7, 1–9.
Deligianni, S.P., Quddus, M., Morris, A., Anvuur, A., Reed, S., 2017. Analyzing and modeling drivers' deceleration behavior from normal driving. Transp. Res. Rec. J. Transp. Res. Board 2663, 134–141.
Deligianni, S.P., Quddus, M., Morris, A., Anvuur, A., 2018. A normal driving based deceleration behaviour study towards autonomous vehicles. Proc. 6th Humanist Conf.
Dijkstra, A., 2013. Assessing the safety of routes in a regional network. Transp. Res. Part C Emerg. Technol. 32, 103–115.
Dimitriou, L., Stylianou, K., Abdel-Aty, M.A., 2018. Assessing rear-end crash potential in urban locations based on vehicle-by-vehicle interactions, geometric characteristics and operational conditions. Accid. Anal. Prev. 118, 221–235.
Dupont, E., Papadimitriou, E., Martensen, H., Yannis, G., 2013. Multilevel analysis in road safety research. Accid. Anal. Prev. 60, 402–411.
El-Basyouny, K., Sayed, T., 2013. Safety performance functions using traffic conflicts. Saf. Sci. 51, 160–164.
Essa, M., Sayed, T., 2018. Traffic conflict models to evaluate the safety of signalized intersections at the cycle level. Transp. Res. Part C Emerg. Technol. 89, 289–302.
Essa, M., Sayed, T., 2019. Full Bayesian conflict-based models for real time safety evaluation of signalized intersections. Accid. Anal. Prev. 129, 367–381.
Greenland, S., Robins, J.M., Pearl, J., 1999. Confounding and collapsibility in causal inference. Stat. Sci. 14, 29–46.
Hauer, E., 1995. On exposure and accident rate. Traffic Eng. Control 36, 134–138.
Hossain, M., Abdel-Aty, M., Quddus, M.A., Muromachi, Y., Sadeek, S.N., 2019. Real-time crash prediction models: state-of-the-art, design pathways and ubiquitous requirements. Accid. Anal. Prev. 124, 66–84.
Huang, H., Abdel-Aty, M., 2010. Multilevel data and bayesian analysis in traffic safety. Accid. Anal. Prev. 42, 1556–1565.
Hyden, C., 1987. The Development of a Method for Traffic Safety Evaluation: the Swedish Traffic Conflict Technique. Bull. 70, Lund Univ. Technol. Lund, pp. 1–8.
Jackman, S., 2017. pscl: Classes and Methods for R [WWW Document]. URL. https://github.com/atahk/pscl/.
Jiang, X., Yan, X., Huang, B., Richards, S.H., 2011. Influence of curbs on traffic crash frequency on high-speed roadways. Traffic Inj. Prev. 12, 412–421.
Kalra, N., Paddock, S.M., 2016. Driving to safety: how many miles of driving would it take to demonstrate autonomous vehicle reliability? Transp. Res. Part A Policy Pract. 94, 182–193.
Karlaftis, M.G., Tarko, A.P., 1998. Heterogeneity considerations in accident modeling. Accid. Anal. Prev. 30, 425–433.
Katrakazas, C., Quddus, M., Chen, W.H., 2018. A simulation study of predicting real-time conflict-prone traffic conditions. IEEE trans. Intell. Transp. Syst. 19, 3196–3207.
Kim, D.G., Washington, S., 2006. The significance of endogeneity problems in crash models: an examination of left-turn lanes in intersection crash models. Accid. Anal. Prev. 38, 1094–1100.
Kuang, Y., Qu, X., Wang, S., 2015. A tree-structured crash surrogate measure for freeways. Accid. Anal. Prev. 77, 137–148.
Lambert, D., 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 34 (1), 37–41.
Li, Y., Xing, L., Wang, W., Wang, H., Dong, C., Liu, S., 2017. Evaluating impacts of different longitudinal driver assistance systems on reducing multi-vehicle rear-end crashes during small-scale inclement weather. Accid. Anal. Prev. 107, 63–76.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. Part A Policy Pract. 44, 291–305.
Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. Accid. Anal. Prev. 37, 35–46.
Lord, D., Washington, S., Ivan, J.N., 2007. Further notes on the application of zero-inflated models in highway safety. Accid. Anal. Prev. 39, 53–57.
Mahmassani, H.S., 2016. 50th anniversary invited article—autonomous vehicles and connected vehicle systems: flow and operations considerations. Transp. Sci. 50, 1140–1162.
Meng, Q., Qu, X., 2012. Estimation of rear-end vehicle crash frequencies in urban road tunnels. Accid. Anal. Prev. 48, 254–263.
Merkle, E.C., You, D., 2018. Package "nonnest2": Tests of Non-Nested Models, pp. 1–7.
Minderhoud, M.M., Bovy, P.H., 2001. Extended time-to-collision measures for road traffic safety assessment. Accid. Anal. Prev. 33 (1), 89–97.
Mullakkal-Babu, F.A., Wang, M., Farah, H., van Arem, B., Happee, R., 2017. Comparative assessment of safety indicators for vehicle trajectories on highways. Transp. Res. Rec. J. Transp. Res. Board 2659, 127–136.
Oh, C., Kim, T., 2010. Estimation of rear-end crash potential using vehicle trajectory data. Accid. Anal. Prev. 42, 1888–1893.
Oh, C., Park, S., Ritchie, S.G., 2006. A method for identifying rear-end collision risks using inductive loop detectors. Accid. Anal. Prev. 38, 295–301.
Papadoulis, A., Quddus, M., Imprialou, M., 2019. Evaluating the safety impact of connected and autonomous vehicles on motorways. Accid. Anal. Prev. 124, 12–22.
Payre, W., Cestac, J., Delhomme, P., 2016. Fully automated driving: impact of trust and practice on manual control recovery. Hum. Factors 58, 229–241.
Pearl, J., 2009. Causality: models, reasoning, and inference. Causality: Models, Reasoning, and Inference, second edition. Second Edition.

Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accid. Anal. Prev. 40, 1486–1497.
Rahman, S., Abdel-aty, M., 2018. Longitudinal safety evaluation of connected vehicles' platooning on expressways. Accid. Anal. Prev. 117, 381–391.
Rahman, M.S., Abdel-Aty, M., Lee, J., Rahman, M.H., 2019. Safety benefits of arterials' crash risk under connected and automated vehicles. Transp. Res. Part C Emerg. Technol. 100, 354–371.
Roshandel, S., Zheng, Z., Washington, S., 2015. Impact of real-time traffic characteristics on freeway crash occurrence: systematic review and meta-analysis. Accid. Anal. Prev. 79, 198–211.
Rummelhard, L., Nègre, A., Perrollaz, M., Laugier, C., 2016. Probabilistic grid-based collision risk prediction for driving application. Springer Tracts Adv. Robot. 109, 821–834.
SAE International, 2016. Surface vehicle recommended practice: taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. SAE Stand. J3016 30.
Sjolander, A., Martinussen, T., 2019. Instrumental variable estimation with the r package ivtools. Epidemiol. Method. 8, 1–20.
Stylianou, K., Dimitriou, L., 2018. Analysis of rear-end conflicts in urban networks using bayesian networks. Transp. Res. Rec.
Tarko, A.P., 2020. Traffic conflicts as crash surrogates. Meas. Road Saf. Using Surrog. Events 31–45.
Tchetgen Tchetgen, E.J., Walter, S., Vansteelandt, S., Martinussen, T., Glymour, M., 2015. Instrumental variable estimation in a survival context Eric. Epidemiology 26, 402–410.
Theofilatos, A., 2017. Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. J. Safety Res. 61.
Theofilatos, A., Yannis, G., Antoniou, C., Chaziris, A., Sermpis, D., 2017. Time series and support vector machines to predict powered-two-wheeler accident risk and accident type propensity: a combined approach. J. Transp. Saf. Secur. 9962, 1–20.
Theofilatos, A., Yannis, G., Kopelias, P., Papadimitriou, F., 2018. Impact of real-time traffic characteristics on crash occurrence: preliminary results of the case of rare events. Accid. Anal. Prev. 1–9.
Ulak, M.B., Ozguven, E.E., Moses, R., Sando, T., Boot, W., AbdelRazig, Y., Sobanjo, J.O., 2019. Assessment of traffic performance measures and safety based on driver age and experience: a microsimulation based analysis for an unsignalized T-intersection. J. Traffic Transp. Eng. (English Ed. 6, 455–469.
Vansteelandt, S., Bowden, J., Babanezhad, M., Goetghebeur, E., 2011. On instrumental variables estimation of causal odds ratios. Stat. Sci. 26, 403–422.
Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econom 57 (2), 307–333, 57.
Wang, C., Stamatiadis, N., 2014. Evaluation of a simulation-based surrogate safety metric. Accid. Anal. Prev. 71, 82–92.
Wang, C., Stamatiadis, N., 2016. Sensitivity analysis on new simulation-based conflict metrics. Saf. Sci. 82, 399–409.
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2010. Statistical and Econometric Methods for Transportation Data Analysis. CRC Press.
Weng, J., Meng, Q., 2011. Rear-end crash potential estimation in the work zone merging areas. J. Adv. Transp. 47, 512–525.
Xing, L., He, J., Abdel-Aty, M., Cai, Q., Li, Y., Zheng, O., 2019. Examining traffic conflicts of up stream toll plaza area using vehicles' trajectory data. Accid. Anal. Prev. 125, 174–187.
Xu, C., Tarko, A.P., Wang, W., Liu, P., 2013. Predicting crash likelihood and severity on freeways with real-time loop detector data. Accid. Anal. Prev. 57, 30–39.
Young, W., Sobhani, A., Lenne, M.G., Sarvi, M., 2014. Simulation of safety: a review of the state of the art in road safety simulation modelling. Accid. Anal. Prev. 66, 89–103.
Yu, R., Abdel-Aty, M., 2013. Investigating the different characteristics of weekday and weekend crashes. J. Safety Res. 46, 91–97.
Yu, R., Wang, Y., Quddus, M., Li, J., 2019. A marginalized random effects hurdle negative binomial model for analyzing refined-scale crash frequency data. Anal. Methods Accid. Res., 100092
Zheng, L., Sayed, T., Essa, M., 2019. Validating the bivariate extreme value modeling approach for road safety estimation with different traffic conflict indicators. Accid. Anal. Prev. 123, 314–323.