Transportation Research Interdisciplinary Perspectives 37 (2026) 102013

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.sciencedirect.com/journal/transportation-research-interdisciplinary-perspectives

What factors influence driver behaviour characteristics when pedestrians cross the road? Insights from the UDRIVE naturalistic driving study

Silvia F. Varotto [a,b,*], Girish Kumaar Srinivasan Ravi Kumar [c], Ruomei Liu [c], Matheus H.W. Stratermans [c], Reyzha Wikki Zaninshi [c], Eleonora Papadimitriou [d], Meng Wang [c,e]

[a] *SWOV Institute for Road Safety Research, Bezuidenhoutseweg 62, 2594 AW The Hague, Netherlands*
[b] *Energy & Mobility Laboratory (EMob-Lab), École nationale des travaux publics de l'État, Université Gustave Eiffel, Rue Maurice Audin 3, 69120 Vaulx-en-Velin, France*
[c] *Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, Netherlands*
[d] *Department of Values, Technology and Innovation, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, Netherlands*
[e] *Chair of Traffic Process Automation, "Friedrich List" Faculty of Transport and Traffic Sciences, Technische Universität Dresden, Hettnerstraße 3, 01069 Dresden, Germany*

ARTICLE INFO

Keywords:
Naturalistic driving
Driver behaviour
Pedestrians
Regression model

ABSTRACT

In urban areas, drivers frequently interact with vulnerable road users. On-road studies have shown that drivers are more likely to have safety-relevant interactions with pedestrians when they are inattentive and when pedestrians behave unexpectedly. Notwithstanding these behavioural effects, most microscopic traffic flow models do not accurately describe driver response to pedestrian crossing behaviour.

This study investigates the factors influencing driver behaviour characteristics when pedestrians cross the road in front of the vehicle. The data were collected in the UDRIVE naturalistic driving study in France and the UK. The interactions with pedestrians in daylight were identified using the MobilEye® smart camera. The minimum time to zebra and the maximum deceleration during each interaction were investigated in regression models.

The results showed that, controlling for the initial speed of the subject vehicle, the minimum time to zebra during interactions was significantly shorter when the pedestrian crossed while the driver had a green traffic light, the vehicle segment was medium, and other pedestrians had already crossed. Controlling for initial speed and acceleration, the maximum deceleration during interactions was lower when the pedestrian crossed while the driver had a green traffic light, no other pedestrians had already crossed, the pedestrian was not a child, teenager or elderly person, and the pedestrian did not glance toward the vehicle. These factors can be incorporated into traffic simulations to describe driver responses more realistically. Further research is needed to understand the influence of the driver's state because most drivers looked toward pedestrians.

1. Introduction

Driving in urban environments involves complex traffic situations, including interacting with vulnerable road users such as pedestrians. Human errors, driver inattention, and unexpected pedestrian behaviour contribute to safety-relevant traffic interactions. Notwithstanding these behavioural effects, most mathematical models used to assess traffic flow efficiency and safety do not sufficiently represent interactions between drivers and pedestrians based on empirical findings. Few studies have implemented mathematical models describing driver decisions to yield to pedestrian crossings into traffic simulations (Chen et al., 2019; Lu et al., 2016) and considered a limited number of traffic and road geometry factors that could influence the driver behaviour characteristics (e.g., speed) when pedestrians cross (Chen et al., 2019). The driver behaviour characteristics when pedestrians cross could change over time and be influenced by the driver's state, traffic conditions, pedestrian behaviour, the vehicle characteristics, and environmental characteristics. To increase the validity and predictive ability of current mathematical models, findings from human factors and traffic psychology should be integrated (Markkula et al., 2018; Van Lint & Calvert, 2018). By incorporating these findings, one could enhance the forecasting accuracy of driver assistance systems and microscopic

* Corresponding author.
E-mail address: silviafrancesca.varotto@entpe.fr (S.F. Varotto).

https://doi.org/10.1016/j.trip.2026.102013
Received 15 October 2024; Received in revised form 4 March 2026; Accepted 19 April 2026
Available online 28 April 2026

simulations investigating the effect of interactions between drivers and pedestrians on traffic operations.

The factors influencing the behaviour of individual drivers with pedestrians can be unravelled in on-road studies. On-road studies offer the opportunity to investigate all observable factors influencing driver responses before and during interactions with pedestrians via cameras and sensors installed inside the vehicle. For example, drivers may fail to respond to pedestrians when engaged in non-driving tasks or distracted. On-road studies have shown that most drivers were inattentive or engaged in non-driving tasks in near crashes with pedestrians (Dingus et al., 2006). These studies are particularly suitable for investigating how participants engage in different tasks because participants do not receive direct instructions as in test-track and driving simulator experiments (Carsten et al., 2013).

Previous on-road studies have shown that driver responses to pedestrians were influenced by driver characteristics, pedestrian behaviour, and road characteristics. The most frequent contributing factors to safety-relevant interactions with pedestrians were driver inattention (Habibovic et al., 2013; Sheykhfard et al., 2021), visual obstruction at intersections (Habibovic et al., 2013), and unexpected pedestrian behaviour outside intersections (Habibovic et al., 2013; Sheykhfard & Haghighi, 2018). Potential conflicts (i.e., contact or proximity will occur if the vehicle or the pedestrian does not change speed or direction) happened significantly more often when the pedestrians were in a group (Tian et al., 2015), were on the road (Tian et al., 2014, 2015), were crossing the road or walking at high speed (Tian et al., 2015), when drivers were turning to the right or the left (Tian et al., 2015), in rural environments (Tian et al., 2015), at mid-walks crosswalks (Tian et al., 2015), on roads without median (Tian et al., 2015) and without traffic control devices (Tian et al., 2014). Although these studies identified several contributing factors using descriptive statistics and tests, further investigations are needed to measure the driver behaviour characteristics (e.g., acceleration and distance headway) during the interaction and quantify the impact of these factors on the driver behaviour characteristics using statistical tests. Quantifying the impact of these factors is necessary for the incorporation into microscopic traffic flow simulation and driving assistance systems. Section 1.1 presents an overview of studies that analyse driver behaviour characteristics during interactions with pedestrians. Section 1.2 summarises the research gaps and the research objectives.

1.1. Background

This section discusses on-road, test-track, and driving simulator studies that analyse driver behaviour when pedestrians cross in front of drivers. This type of interaction was chosen because of the high risk of traffic disruptions (e.g., queue formation) and severe accidents (e.g., frontal collision). On-road and driving simulator studies allow one to observe driver behaviour with greater detail (e.g., driver behaviour characteristics, driver characteristics and states) and across a broader range of traffic situations (e.g., different situations for each driver over time and locations with different road characteristics) than fixed videography methods. For a review of data-collection methods for investigating interactions between drivers and pedestrians, the reader is referred to Sheykhfard et al. (2021). In this section, longitudinal and lateral distances reported in prior studies are described with respect to the travel direction of the vehicle.

A test-track study and a few on-road studies have analysed the driver behaviour characteristics when pedestrians cross. In a test-track study, Lubbe and Rosén (2014) analysed the time to collision and the longitudinal and lateral distances between the driver and the pedestrian when the driver started braking. The time to collision ranged between 2.1 and 4.3 s and was not influenced by the driver’s speed. Tian et al. (2019) investigated the distance and the time of collision when the pedestrian first appeared to the driver in an on-road study. When the pedestrian first appeared, the distance was most often between 10 and 40 m, while the time to collision was between 2.5 and 5 s. Sun et al. (2022) analysed the time to collision when the driver started decelerating due to pedestrians crossing. When the driver started decelerating, the time to collision ranged from 1.68 to 6.28 s. Sheykhfard et al. (2023) investigated the time to collision when either the driver or the pedestrian executed an evasive manoeuvre (i.e., time to accident) and the elapsed time between the first and the second user passing through a point (i.e., post encroachment time). Descriptive statistics suggested that both indicators were higher when a zebra crossing was present. Notably, these studies did not analyse the driver behaviour characteristics during the whole interaction, which can be considered more informative of the driver response. Further investigations are also required to understand the impact of a broader range of factors influencing driver response using statistical methods.

Some driving simulator studies have analysed the effect of one or two factors on the driver behaviour characteristics at the beginning of and during the interaction using statistical tests. Lubbe and Davidsson (2015) found that the time to collision and the longitudinal distance at braking were shorter, and the lateral distance was larger, when the pedestrian's speed was higher. Bella and Silvestri (2015) showed that the minimum speed during the interaction was higher, and that the distance where the deceleration began and ended was longer with a curb extension. In a follow-up study, Bella and Silvestri (2021) showed that the time to collision at braking was shorter and the speed reduction was higher when the pedestrian crossed outside of a crosswalk. Portera et al. (2024) found that the minimum time to collision during the interaction and the reaction distance at braking were shorter with a conventional crosswalk than with an LED-based crosswalk. The speed was lower, and the reaction distance at braking was shorter when the complexity of the cognitive non-driving task was higher.

Other driving simulator studies have investigated the impact of various factors on the driver behaviour characteristics using regression models. Wu et al. (2018) found that the maximum deceleration was higher at night, without crosswalks, with two lanes and when pedestrians wore dark clothing. The post-encroachment time was lower at night and when pedestrians wore dark clothing. The minimum time to collision was lower at night, without crosswalks, with one lane, and when pedestrians wore dark clothing. Dozza et al. (2020) showed that the minimum time to collision, the time to collision at gas release, and the time to collision at braking were shorter when the crossing side was near, the vehicle speed was high, and the crossing angle was 90 degrees. The time to arrival at gas release and braking was shorter when the crossing side was near. The longitudinal distance at gas release and braking was shorter when the crossing side was near, the vehicle speed was low, and the crossing angle was 90 degrees. The lateral distance at gas release and braking were shorter when the crossing side was near, the pedestrian speed was low, and the crossing angle was 45 degrees. Angioi and Bassani (2022) found that the minimum time to collision was shorter when drivers were familiar with the route, when there was no curb extension, when the accepted gap for the pedestrian was short, and when the driver was young. The post-encroachment time was shorter when the drivers were familiar with the route, and the gap accepted by the pedestrian was short. The maximum speed was higher when the drivers were familiar with the route, and there was no curb extension. The maximum deceleration was higher when there was no curb extension, and the gap accepted by the pedestrian was short. Recently, Yang et al. (2024) showed that the mean deceleration was higher and the maximum deceleration occurred closer to the pedestrians at zebra crossings and with shorter gaps. The mean lateral deviation was larger outside of zebra crossings, with shorter time gaps, and with encounters over time. Notably, these studies analysed the impact of a limited range of factors on the driver behaviour characteristics in controlled experiments.

1.2. Research gap and objectives

The above overview shows that previous test-track and on-road experiments have analysed driver behaviour characteristics at the onset of interactions with pedestrians crossing in front of them. Some driving simulator studies have examined the effects of a limited number of factors on driver behaviour characteristics at the beginning and during the interaction, using statistical tests and regression models. Further analysis is needed to examine the effects of a wider range of factors in more realistic and complex traffic situations based on data collected in on-road experiments. Quantifying the effects of these factors on driver behaviour characteristics using regression models represents the first step toward understanding their impact on traffic operations and designing automated vehicle functionalities that can improve traffic safety.

The research objective of this study is to analyse the main factors that influence the behaviour characteristics of drivers interacting with pedestrians crossing in on-road experiments. The minimum time to zebra (Wu et al., 2018; Dozza et al., 2020; Angioi & Bassani, 2022; Portera et al., 2024) and the maximum deceleration (Angioi & Bassani, 2022; Wu et al., 2018) during each interaction (i.e., interaction characteristic) are investigated to capture adaptations in the longitudinal control task of drivers. This study is based on analysing and annotating the naturalistic driving data collected in the UDRIVE project (Van Nes et al., 2019). The pedestrian collision warning generated by the MobilEye® smart camera is used to identify interactions between drivers and pedestrians in daylight conditions. The distance measured by the MobilEye® camera and driver behaviour characteristics measured by the CAN are utilised to calculate the interaction characteristics. The interaction characteristics investigated in this study are often used to describe drivers' responses in microscopic traffic simulations and classify safety-relevant events. The effects of the behaviour characteristics of the subject vehicle at the beginning of the interaction, the state (i.e., glance behaviour and engagement in non-driving tasks) and characteristics of the driver, the behaviour and characteristics of the pedestrian, and the infrastructure and environment characteristics on the interaction characteristics are investigated using correlation analyses and regression models.

The remainder of the paper is organised as follows. Section 2 describes the driver behaviour and video data, the annotation procedures, and the data analysis methods in this study. Section 3 presents the regression models predicting the minimum time to zebra and the maximum deceleration during interactions with pedestrians. Finally, section 4 presents the main findings on factors affecting event characteristics and discusses recommendations for practice and future research.

2. Method

2.1. UDRIVE database

The driver behaviour data were collected in the United Kingdom and France using private passenger vehicles within the UDRIVE project (Van Nes et al., 2019), which aimed to investigate driver behaviour across European countries. Ninety-six participants were recruited from the local population using social media, advertisements, and networks. The recruitment criteria included owning specific vehicle makes and models (Renault Clio and Megane), a minimum annual mileage of 10,000 km, and a minimum quota of 40% per gender. The vehicle models were chosen to represent the small-vehicle (Clio) and medium-vehicle (Megane) segments, which covered 49% of the passenger vehicle market share in Europe in 2011 (Bärgman et al., 2017). In total, 46 participants were males and 50 were females. Twelve participants were 18–29 years old, twenty-seven were 30–39, twenty-three were 40–49, and thirty-four were 50–70. All participants signed a written informed consent form after being informed about the project, their rights and their duties.

The data acquisition system (DAS) designed in the UDRIVE project was installed in sixty private passenger vehicles. A single vehicle was shared among participants from the same household. Overall, 31 participants drove small vehicles, and 65 participants drove medium vehicles. The data were collected over 12 to 18 months between 2015 and 2017. The DAS registered date and time, GPS coordinates, CAN data (e.g., speed and acceleration), data from MobilEye® smart camera (e.g., position and relative speed of road users in front), and video data from seven cameras recording the surrounding environment and the driver (Fig. 1). The data were registered at a frequency of 1 Hz (GPS coordinated) and 10 Hz (e.g., acceleration). In addition, the drivers' demographic characteristics were collected via questionnaires. The present study is based on data from eighty-nine participants who drove 401,409 km in 9,164 h (Jansen et al., 2017).

2.2. Selection of interactions with pedestrians

Interactions between drivers and pedestrians were identified using data from a MobilEye® smart camera (Jansen et al., 2017). These smart cameras could automatically identify and classify surrounding road users using computer vision algorithms (MobilEye®). The MobilEye® smart camera requires sufficient lighting to detect and record interactions. The data included the type (vehicle, cyclist or pedestrian), position (lateral and longitudinal distances relative to the subject vehicle) and relative speed of the closest road users in front of the subject vehicle. In addition, the MobilEye® smart camera generated a pedestrian collision warning when the estimated time to collision with a pedestrian fell below two seconds. The warning terminated when the estimated time to collision exceeded two seconds or when the pedestrian was outside the field of view. Notably, this warning was not available to the drivers during the experiment. All events that triggered the pedestrian collision warning were manually inspected and annotated during the UDRIVE project (Jansen et al., 2017). The number and type of relevant vulnerable road users present at the onset of the pedestrian collision warning were annotated. The event was considered eligible for analysis in this study if, according to the original annotation, one or more pedestrians were within close proximity of the subject vehicle.

2.3. Video annotation

The second, fourth, and fifth authors inspected and manually coded the videos in November 2019 to extract additional information on the pedestrian, the driver, the driving context, and the road. Initially, the three authors coded ten events and discussed the variables annotated in a plenary session to achieve a mutual understanding. After this training session, the events were randomly assigned for annotation to one of the three authors. Each video was played slowly forward and backward multiple times by each annotator. The first author manually inspected the videos and reviewed the annotated variables for all events analysed in this study. To guarantee the reliability of the annotated variables, only events in which the annotators reached a joint agreement were included in the analysis.

The variables were annotated according to the definitions proposed in the UDRIVE codebook when they were available (Bärgman et al., 2017). Some categories defined in the UDRIVE codebook were absent from this dataset or were combined with others when the number of observations was too small for separate analysis. Specific definitions were developed for this study when they were unavailable in the UDRIVE codebook. Relevant events were assumed to begin at the onset of the pedestrian collision warning and to terminate at the offset of the warning. During the 5 s before the event began, driver behaviour and the presence of other pedestrians were annotated as described in Table 1. The non-driving task category includes all non-driving tasks defined in the UDRIVE codebook (e.g., phone usage, eating and drinking, smoking). Pedestrian characteristics were annotated at the beginning of the

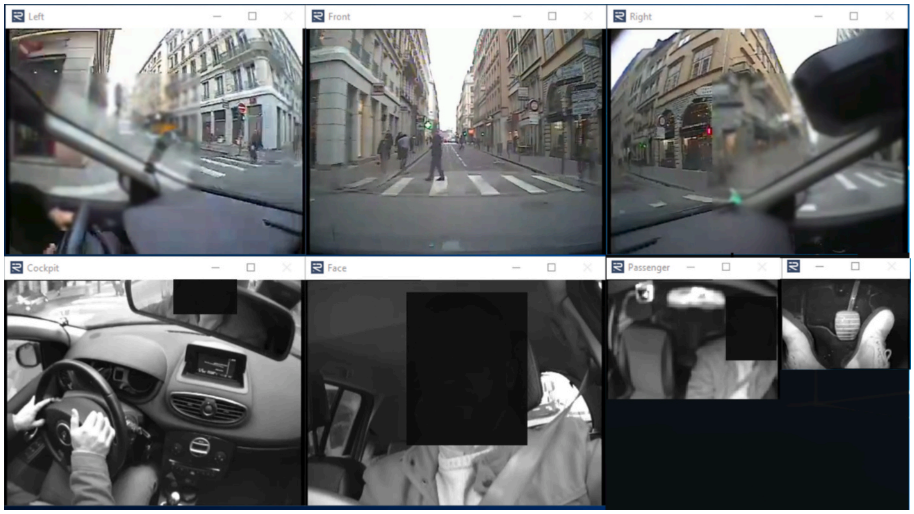

Fig. 1. Seven camera views in the UDRIVE database during an interaction with a pedestrian: front left, front centre, front right, cockpit, face of the driver, cabin, and feet. To protect the privacy of the driver, the face has been obscured.

Table 1
Driver behaviour and the presence of other pedestrians annotated before the event.

Variable	Categories	Source
Driver looked towards the pedestrian	No, yes	Study-defined
Driver engaged in a non-driving task	No, yes	UDRIVE codebook
Other pedestrians were present	No, crossing the road, on the pavement, in other locations	Study-defined

event, as shown in Table 2. When pedestrian characteristics were unclear at the beginning of the event, they were annotated based on the most explicit frames during the event. Finally, the characteristics of the driving context and the road were annotated based on the most explicit frames available before and during the event, as described in Table 3. The adverse weather category includes all adverse weather conditions in the UDRIVE codebook. The residential category includes the open and moderate residential categories in the codebook. The other intersection type category includes all intersection types different from X and T in the codebook (e.g., Y intersection and roundabout).

2.4. Definition of longitudinal interactions with pedestrians

The variables annotated and available in the UDRIVE database were used to identify longitudinal interactions with pedestrians crossing

Table 2
Pedestrian characteristics annotated at the beginning of the event.

Variable	Categories	Source
Pedestrian location	Pavement, road, median refugee	Study-defined
Pedestrian sight direction	Towards vehicle, away from vehicle, into the crossing	Study-defined
Pedestrian intention	Cross the road, walk towards vehicle, walk away from vehicle, walk perpendicular to vehicle, wait, doubtful, enter or exit parked car	Study-defined
Pedestrian number	Single, group	UDRIVE codebook
Pedestrian age group	Children, teenagers, adults, elderly, unclear	UDRIVE codebook
Pedestrian gender	Male, female, mixed group, unclear	UDRIVE codebook

Table 3
Characteristics of the driving context and the road annotated before or during the event.

Variable	Categories	Source
Daylight	No, yes	UDRIVE codebook
Adverse weather conditions	No, yes	UDRIVE codebook
Locality type	Residential, business industrial, urban	UDRIVE codebook
Presence and type of intersection	None, X intersection, T intersection, other intersection type, unknown	UDRIVE codebook
Crossing type	None, zebra crossing, zebra crossing with median, zebra crossing with warning traffic lights, traffic lights for cars only, traffic light for cars and pedestrians	Study-defined
Priority rules	Right of way only, traffic signs, warning traffic lights, traffic lights	Study-defined

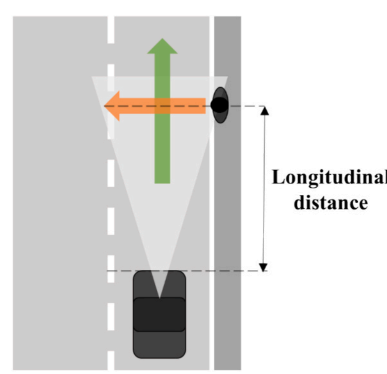

Fig. 2. Schematic representation of longitudinal interactions based on the definition proposed in this study. The grey triangle indicates the field of view of the MobilEye® smart camera.

(Fig. 2). The longitudinal and lateral distances between the subject vehicle and the pedestrian, relative to the coordinate system of the vehicle, were based on the MobilEye® data. The MobilEye® smart camera tracks multiple road users simultaneously and may misclassify them (e.g., traffic signs or light poles may be misclassified as pedestrians). We manually checked that the MobilEye® data of the pedestrians selected in each event matched those of interest in the videos, and that the estimated distances matched the traffic situation. Comparisons with other measures (e.g., vehicle speed, time needed to reach the crossing location, and lane width) indicated that the distances estimated in these traffic situations were reliable for analysis.

Only events classified as longitudinal interactions were analysed. Events that did not meet these criteria were excluded. Longitudinal interactions were identified as events in which, at the beginning, (1) the pedestrian was crossing the road in front of the vehicle and (2) the driver was moving straight ahead towards the pedestrian. Events in which, at the beginning, the driver was in a curve, in a roundabout, outside the road (e.g., parking lots), changing lanes and stopped at an intersection (i.e., red traffic light or stop sign) were removed from the analysis. The analysis was therefore restricted to the participants who had at least one valid longitudinal interaction.

The minimum time to zebra and the maximum deceleration were selected to measure adaptations in drivers' longitudinal control during these interactions. The time to zebra is selected as an indicator of the quality of the longitudinal control task. It captures the time remaining before the collision with the pedestrian crossing and was developed to measure collision risk in this specific traffic situation (Johnsson et al., 2018; Várhelyi, 1998). A time to collision shorter than 1.5 s is commonly considered a safety–critical value (El-Basyouny & Sayed, 2013). The minimum time to zebra was calculated using the vehicle speed during the interaction and the longitudinal distance to the point where the pedestrian was crossing the road. This definition allowed for the inclusion of situations in which pedestrians crossed on and off zebra crossings. While the pedestrian movement determines the initial crossing position, the drivers' longitudinal control actions actively influence the observed minimum time to zebra. This definition is equivalent to the minimum time-to-collision in this specific traffic situation (Wu et al., 2018; Dozza et al., 2020; Angioi & Bassani, 2022; Portera et al., 2024), assuming that the direction of movement of the pedestrian is perpendicular to the direction of movement of the vehicle. The maximum deceleration is chosen as an indicator of driver aggressiveness while interacting with pedestrians. A longitudinal deceleration lower than 6.5 – 7.5 m/s^2 is commonly recommended for identifying near-crashes and crashes in naturalistic driving studies (Perez et al., 2017). The maximum deceleration was calculated based on the vehicle acceleration during the interaction. This measure was analysed in previous studies (Angioi & Bassani, 2022; Wu et al., 2018).

2.5. Statistical analysis

The relationships among the minimum time to zebra, the maximum deceleration, traffic conditions, infrastructure characteristics, driver state and characteristics, and pedestrian behaviour and characteristics were explored using correlation analysis. Only variables with at least five observations per category were analysed using statistical methods. Spearman rank-order correlations (r_S) and Pearson product-moment correlations (r_P) were calculated.

The main factors influencing the minimum time to zebra and the maximum deceleration (interaction characteristic) were investigated in regression models. Regression models allow capturing the unconditional marginal effect of several explanatory variables on road user behaviour characteristics, accounting for correlations between repeated observations over time (Wu et al., 2018; Dozza et al., 2020; Angioi & Bassani, 2022). Regression models are well-suited for analysing observations in naturalistic driving experiments, where experimenters do not control the effects of several factors, and finding similar traffic situations that can serve as a baseline remains a challenge (Carsten et al., 2013). The regression models predicting the interaction characteristics (IntChar) for driver *n* at time *t* are presented in equation (1):

$$Y_n(t) = \mu + \lambda \bullet X_n(t) + \omega \bullet \upsilon_n(t) \tag{1}$$

where μ is the constant, λ is the vector of parameters associated with the explanatory variables $X_n(t)$, and ω is the parameter related to the observation-specific error term $\upsilon_n(t) \sim N(0,1)$. The explanatory variables can include the subject vehicle behaviour characteristics, driver state and characteristics, pedestrian behaviour and characteristics, and infrastructure and environmental characteristics. Equation (1) can also include a driver-specific error term and a trip-specific error term that capture unobserved preferences influencing all interactions by the individual driver over time and within the same trip. The probability density function of the interaction characteristics is given in equation (2):

$$P(IntChar_n(t)) = \frac{1}{\omega}\phi\left(\frac{IntChar_n(t) - \mu - \lambda \bullet X_n(t)}{\omega}\right) \tag{2}$$

The regression models were estimated using maximum likelihood methods in the R package 'nlme' (Pinheiro et al., 2024). This package was chosen because it allows explicitly defining and testing a residual variance–covariance structure. The package 'ggeffects' (Lüdecke, 2018) was used to calculate the estimated marginal means of the interaction characteristics when only one explanatory variable was varied, while all others were held fixed. To assess uncertainty in the estimated parameters, confidence intervals were calculated.

In this study, each interaction characteristic (i.e., minimum time to zebra and maximum deceleration) was treated as the dependent variable in a separate regression model. All the other variables available were tested as independent variables, including driver state in the five seconds before the event (e.g., engagement in non-driving tasks), driver and pedestrian behaviour characteristics at the beginning of the event (e.g., driver speed), and the characteristics of the road and environment (e.g., crossing type). Notably, the interaction characteristics were calculated as the minimum or maximum values observed during the interaction, whereas the independent variables (e.g., driver speed) were measured before or at the start of the event. This temporal separation reduces the risk of endogeneity between independent and dependent variables. The explanatory variables included in the final specifications of the regression models were selected based on statistical significance (p-value < 0.05) and interpretability (non-redundant variables). We centred each continuous variable on its mean. We compared alternative model specifications and tested the impact of each explanatory variable statistically using the likelihood ratio test. The variables with medium or high correlation with the interaction characteristics were included in the regression model first. The explanatory variables that were correlated with each other were tested separately in the model. In addition, we examined the correlation matrix and the variance inflation factors of the parameters estimated each time. Potential indicators of multicollinearity considered in this study were large changes in parameter estimates, non-significant parameters with a significant likelihood-ratio test, and variance inflation factors exceeding five when a new variable was added to the model. We combined variables with similar meanings and a non-significant difference in their impact on the interaction characteristics. Variables that did not have a significant effect were excluded from the model. A binary variable indicating missing values was included in the model, along with the original variable, when the original variable was unavailable for certain observations (e.g., the variable was unclear based on the annotation).

3. Results

3.1. Descriptive statistics

The MobilEye® data and the video annotation identified 351 interactions between drivers and pedestrians (Jansen et al., 2017). Based on the definitions proposed in Section 2.4, 69 events were classified as longitudinal interactions. The longitudinal interactions happened in 68 distinct trips by 28 drivers. The number of interactions per driver ranged from 1 to 11 (*Mdn* = 1, *M* = 2.46, *SD* = 2.46). The minimum time to zebra during these interactions was between 0.32 s and 1.96 s (*Mdn* = 1.31, *M* = 1.22, *SD* = 0.43). Most observations (69.6%) had a minimum time to zebra lower than 1.5 s, a value commonly considered safety–critical (El-Basyouny & Sayed, 2013). The maximum deceleration was between 0.20 m/s^2 and 6.49 m/s^2 (*Mdn* = 1.50, *M* = 1.74, *SD* = 1.30). Most observations showed a moderate maximum deceleration, and only one reached a maximum deceleration similar to the thresholds used to identify near-crashes (Perez et al., 2017). A one-sample Kolmogorov-Smirnov test was executed to test whether the mean and the standard deviation of the minimum time to zebra and the maximum deceleration come from a normal distribution. The test did not indicate significant deviations from normality at the 5% significance level (minimum time to zebra: test statistic = 0.091, p-value = 0.587; maximum deceleration: test statistic = 0.119, p-value = 0.283).

We calculated descriptive statistics for the explanatory variables to understand the general interaction characteristics. The statistics for the driver behaviour characteristics at the beginning of the longitudinal interactions are presented in Table 4. All interactions occurred at low speeds and close distances. The median acceleration shows that most drivers were already decelerating at the beginning of the interaction.

The descriptive statistics of pedestrian behaviour characteristics at the beginning of interactions and pedestrian characteristics are presented in Table 5. For each category, the mean and the standard deviation of the minimum time to zebra and the maximum deceleration were calculated. Most pedestrians were on the road, and all of them were crossing the road at the beginning of the event. When pedestrians were on the pavement at the beginning of the event, drivers decelerated more and had a longer minimum time to zebra. Most pedestrians were looking into the crossing at the beginning of the event. When the pedestrians were looking toward the vehicle, drivers decelerated more. Most often, no other pedestrians were visible before the event began. When other pedestrians crossed before the event began, drivers decelerated more. A certain number of pedestrians crossed the road when the driver had a green traffic light (e.g., violation of a red pedestrian traffic light or unprotected crossing at the traffic light). Drivers decelerated less and had a shorter minimum time to zebra when they had a green traffic light and pedestrians crossed. Most pedestrians were single and adults. Drivers decelerated more when approaching children, teenagers, and elderly pedestrians. More pedestrians were identified as males than females.

The descriptive statistics of road and environment characteristics are shown in Table 6. Most longitudinal interactions occurred in France. All interactions occurred in daylight, consistently with the operational limitations of the MobilEye® smart camera. No interactions were observed under low-light conditions, such as dusk or artificial lighting. Most interactions were in favourable weather conditions. Drivers decelerated less when the weather conditions were adverse. Most interactions happened in urban environments and at intersections. Most intersections were regulated by traffic signs and road markings, and the most common crossing type was a zebra crossing without traffic lights. Drivers decelerated less and had a shorter minimum time to zebra at intersections and crossings with traffic lights.

The descriptive statistics of driver characteristics and vehicle characteristics are presented in Table 7. When we compare age groups, we notice that drivers aged 40–49 had the highest percentage of interactions. Drivers aged 30–39 decelerated more and had a shorter minimum time to zebra. The number of interactions and the interaction characteristics were similar across genders. Notably, the driver was looking toward the pedestrian and was not engaged in non-driving tasks before the beginning of almost all interactions. The observation with the driver not looking toward the pedestrian showed the harshest maximum deceleration. The vehicle segment was small more often than medium. When the vehicle segment was medium, drivers decelerated less and had a shorter minimum time to zebra.

Table 4
Descriptive statistics of driver behaviour characteristics at the beginning of longitudinal interactions.

Variable	Description	Min.	Mdn	M	Max.	SD
Speed	Speed of the subject vehicle in km/h	9.39	20.19	23.13	42.03	8.24
Accel.	Acceleration of the subject vehicle in m/s^2	−3.57	−0.70	−0.97	1.21	1.00
LongDist	Longitudinal distance to the closest pedestrian in m	4.14	9.56	10.74	21.82	4.63
LatDist	Lateral distance to the closest pedestrian in m	0.00	0.86	0.96	1.44	2.15

Table 5
Descriptive statistics of pedestrian behaviour characteristics and pedestrian characteristics.

Variable	Levels	Observations (percentage per event)	M (SD) of minimum time to zebra	M (SD) of maximum deceleration
Location at the beginning of the event	Pavement	3 (4.3%)	1.66 (0.37)	2.85 (0.69)
	Road	66 (95.7%)	1.20 (0.42)	1.69 (1.30)
Glance direction at the beginning of the event	Towards the subject vehicle	7 (10.1%)	1.24 (0.49)	2.31 (1.08)
	Away from the subject vehicle	8 (11.6%)	1.10 (0.49)	1.47 (0.97)
	Into the crossing	49 (71.0%)	1.20 (0.41)	1.63 (1.35)
	Unclear	5 (7.2%)	1.59 (0.33)	2.46 (1.39)
Presence of other pedestrians before the beginning of the event	Pedestrians crossing	14 (20.3%)	1.02 (0.61)	2.33 (1.89)
	Pedestrian not crossing	14 (20.3%)	1.23 (0.43)	1.40 (0.77)
	No other pedestrians	41 (59.4%)	1.29 (0.34)	1.65 (1.16)
Crossing when the driver has a green traffic light	No	57 (82.6%)	1.29 (0.40)	1.85 (1.31)
	Yes	12 (17.4%)	0.91 (0.42)	1.20 (1.15)
Number	Single	56 (81.2%)	1.22 (0.43)	1.72 (1.35)
	Group	13 (18.8%)	1.22 (0.42)	1.84 (1.13)
Age	Children	1 (1.4%)	1.55 (NA)	2.00 (NA)
	Teenagers	5 (7.2%)	1.52 (0.28)	2.40 (1.26)
	Adults	54 (78.3%)	1.22 (0.43)	1.74 (1.34)
	Elderly	1 (1.4%)	1.31 (NA)	3.25 (NA)
	Unclear	8 (11.6%)	0.98 (0.46)	1.08 (0.97)
Gender	Female	17 (24.6%)	1.36 (0.39)	1.75 (0.96)
	Male	26 (37.7%)	1.20 (0.43)	2.03 (1.66)
	Mixed group	2 (2.9%)	1.57 (0.40)	2.09 (1.55)
	Unclear	24 (34.8%)	1.12 (0.44)	1.39 (1.02)

3.2. Statistical analysis

3.2.1. Minimum time to zebra

The relationships between minimum time to zebra, traffic conditions, infrastructure characteristics, driver state and characteristics, and pedestrian behaviour, state and characteristics were explored using correlation analysis. A short minimum time to zebra was observed when the speed at the beginning of the interaction was high (r_P = -0.25, p_P = 0.042; r_S = -0.24, p_S = 0.048), the pedestrian crossed when the driver had a green traffic light (r_P = -0.34, p_P = 0.005; r_S = -0.32, p_S = 0.007), other pedestrians had already crossed before (r_P = -0.24, p_P = 0.043; r_S = -0.19; p_S = 0.118), and the vehicle segment was medium (r_P = -0.22, p_P = 0.067; r_S = -0.22, p_S = 0.075). A long minimum time to zebra was observed when the intersection priority rules were based on traffic signs and road markings (r_P = 0.22, p_P = 0.069; r_S = 0.21, p_S = 0.079) and when the pedestrian crossing

Table 6
Descriptive statistics of environment and road characteristics.

Variable	Levels	Observations (percentage per event)	M (SD) of minimum time to zebra	M (SD) of maximum deceleration
Country	France	57 (82.6%)	1.20 (0.43)	1.68 (1.28)
	United Kingdom	12 (17.3%)	1.34 (0.42)	2.03 (1.34)
Adverse weather conditions	No	64 (92.8%)	1.22 (0.43)	1.80 (1.32)
	Yes	5 (7.2%)	1.30 (0.42)	0.93 (0.62)
Locality type	Residential	15 (21.7%)	1.33 (0.36)	1.91 (1.09)
	Business industrial	5 (7.2%)	0.97 (0.46)	1.99 (1.86)
	Urban	49 (71.0%)	1.21 (0.44)	1.66 (1.32)
Intersection type	No intersection	26 (37.7%)	1.22 (0.47)	1.82 (1.36)
	X intersection	17 (24.6%)	1.16 (0.51)	1.67 (1.57)
	T intersection	21 (30.4%)	1.29 (0.30)	1.73 (1.09)
	Other intersection types	4 (5.8%)	1.19 (0.47)	1.94 (1.00)
	Unclear	1 (1.4%)	0.79 (NA)	0.20 (NA)
Intersection priority rules	Law only	4 (5.8%)	1.45 (0.52)	1.57 (0.52)
	Traffic signs and road markings	36 (52.2%)	1.31 (0.37)	2.01 (1.41)
	Traffic lights	12 (17.4%)	0.91 (0.42)	1.20 (1.15)
	Unclear	17 (24.6%)	1.20 (0.45)	1.58 (1.22)
Pedestrian crossing type	None	25 (36.2%)	1.22 (0.44)	1.46 (1.05)
	Zebra crossing and traffic signs	28 (40.6%)	1.32 (0.39)	2.17 (1.51)
	Zebra crossing and traffic lights	12 (17.4%)	0.91 (0.42)	1.20 (1.15)
	Zebra crossing and warning lights	4 (5.8%)	1.46 (0.18)	2.03 (0.81)

Table 7
Descriptive statistics of driver characteristics and vehicle characteristics.

Variable	Levels	Observations (percentage per event)	M (SD) of minimum time to zebra	M (SD) of maximum deceleration
Age	18–29	11 (15.9%)	1.33 (0.39)	1.76 (1.25)
	30–39	11 (15.9%)	1.09 (0.50)	2.69 (1.26)
	40–49	34 (49.3%)	1.21 (0.45)	1.62 (1.39)
	50–70	13 (18.8%)	1.26 (0.35)	1.22 (0.71)
Gender	Female	37 (53.6%)	1.22 (0.47)	1.75 (1.38)
	Male	32 (46.4%)	1.23 (0.39)	1.72 (1.23)
Driver looked towards pedestrian	No	1 (1.4%)	1.03 (NA)	6.49 (NA)
	Yes	68 (98.6%)	1.23 (0.43)	1.67 (1.17)
Engagement in non-driving tasks	No	67 (97.1%)	1.22 (0.43)	1.75 (1.32)
	Yes	2 (2.9%)	1.33 (0.02)	1.38 (0.25)
Vehicle segment	Small	56 (81.2%)	1.27 (0.42)	1.86 (1.37)
	Medium	13 (18.8%)	1.03 (0.41)	1.21 (0.84)

had zebras with traffic signs or warning lights ($r_P = 0.26$, $p_P = 0.030$; $r_S = 0.26$, $p_S = 0.028$). The short minimum time to zebra at intersections with zebra crossings and traffic lights can be explained by the fact that all observed interactions at these crossings occurred when drivers had a green traffic light and pedestrians crossed. The correlations between the minimum time to zebra and the other explanatory variables available were weak ($|r| < 0.20$).

A regression model was developed to simultaneously capture the impact of multiple explanatory variables on the minimum time to zebra. Table 8 presents the goodness-of-fit measures, and Table 9 shows the estimation results. The minimum time to zebra $\mathrm{MinTTZ_n(t)}$ during the event at time *t* for driver *n* is presented in equation (3):

$$\begin{aligned}\mathrm{MinTTZ_n(t)} = \mu^{\mathrm{TTZ}} + \lambda^{\mathrm{TTZ}}_{\mathrm{Speed}} \bullet \mathrm{Speed(t)} + \lambda^{\mathrm{TTZ}}_{\mathrm{DrivGreenLight}} \\ \bullet \mathrm{DrivGreenLight(t)} + \lambda^{\mathrm{TTZ}}_{\mathrm{OthPedBef}} \\ \bullet \mathrm{OthPedBef(t)} + \lambda^{\mathrm{TTZ}}_{\mathrm{RenMeg}} \bullet \mathrm{RenMeg_n} + \omega^{\mathrm{TTZ}} \bullet \upsilon^{\mathrm{TTZ}}_{n}(t)\end{aligned} \tag{3}$$

where μ^{TTZ} is the mean value, λ^{TTZ} is a vector of parameters related to the explanatory variables in Table 9, and ω^{TTZ} is the parameter associated with the observation-specific error term $\upsilon^{\mathrm{TTZ}}_{n}(t) \sim N(0,1)$. The null hypothesis that the mean and the standard deviation of the residuals come from a normal distribution could not be rejected at the 5% significance level (one-sample Kolmogorov-Smirnov test: test statistic = 0.068, p-value = 0.886).

Most parameters associated with the explanatory variables in equation (3) are statistically significant at the 5% level. One parameter which was statistically significant at the 10% level was included due to the small sample size (69 events). The correlation matrix and variance inflation factors did not indicate multicollinearity among the variables in equation (3). The R^2 indicates that the explanatory variables capture 29.8% of the variability in the minimum time to zebra. Drivers had a shorter minimum time to zebra when the speed at the beginning of the interaction was high. The acceleration, the longitudinal and the lateral distance to the closest pedestrian did not have a significant impact and were therefore excluded from the model. Drivers had a shorter minimum time to zebra when they had a green light at the intersection and a pedestrian crossed (i.e., unprotected pedestrian crossing or illegal crossing). The number, age, gender and glance direction of the pedestrians did not have a significant impact. Drivers had a shorter minimum time to zebra when other pedestrians were present before the beginning of the event (p-value = 0.0741). The country, weather conditions, type of locality, and intersection type did not influence the minimum time to zebra. Controlled for the other variables, priority rules with traffic signs and road markings and pedestrian crossings with zebras did not have a significant impact. The driver had a shorter minimum time to zebra when the vehicle segment was medium instead of small. The driver's age and gender did not have a significant impact. Finally, we tested a driver-specific error term that did not have a significant effect and was therefore not included in the model. The regression model in equation (3) was also estimated separately for each country. In this dataset, situations in which drivers had a green light at the intersection and the pedestrians crossed were observed only in France. This variable was omitted from the specification when the model was estimated using UK data. The results showed that the parameters associated with the other variables did not differ significantly between countries.

We investigated the impact of changes in the explanatory variables on the minimum time to zebra by calculating the minimum time to zebra for observations in which only one variable was changed while keeping

Table 8
Statistics of the regression model predicting the minimum time to zebra. The log-likelihood measures are computed using ML estimation methods.

Statistics	
Number of parameters associated with the explanatory variables	4
Number of drivers	28
Number of observations	69
Constant log likelihood	−38.86
Constant Akaike information criterion	81.72
Constant Bayesian information criterion	86.19
Final log likelihood	−26.63
Final Akaike information criterion	65.27
Final Bayesian information criterion	78.67
R^2	0.298

Table 9
Estimation results of the regression model predicting the minimum time to collision. The model parameters are estimated based on REML methods using the R package 'nlme', which calculates the test statistics of the fixed effects only.

Variable	Description	Parameters	Estimate	t-stat.	p-value
–	Mean minimum time to zebra	μ^{TTZ}	1.40	23.6	<0.0005
Speed	Speed of the subject vehicle in km/h at the beginning of the event	λ^{TTZ}_{Speed}	−0.0129	−2.33	0.0229
DrivGreenLight	Binary variable equal to one when the pedestrian crossed while the driver had a green light	$\lambda^{TTZ}_{DrivGreenLight}$	−0.399	−3.33	0.0014
OthPedBef	Binary variable equal to one when other pedestrians were present before the beginning of the event	$\lambda^{TTZ}_{OthPedBef}$	−0.208	−1.82	0.0741
MedSeg	Binary variable equal to one when the vehicle segment was medium	λ^{TTZ}_{MedSeg}	−0.340	−2.94	0.0046
υ^{TTZ}_n	Observation-specific error term	ω^{TTZ}	0.370	–	–

all the other variables fixed. In the baseline observation, the speed was equal to 23.13 km/h, the pedestrian did not cross while the driver had a green traffic light, no other pedestrians crossed before the interaction, and the vehicle segment was small. The characteristics were selected based on the mean conditions in the sample. The minimum time to zebra associated was equal to 1.40 s (i.e., the mean in Table 9). Fig. 3 shows the results for the continuous and the nominal explanatory variables. When we compare all the different variables, we note that pedestrians crossing when the driver had the green traffic light has the largest impact on the minimum time to zebra. Appendix A includes an out-of-sample validation analysis of the model presented in Table 9. In most validation samples, the final regression models have lower mean forecasting errors than those that incorporate only constants. The findings indicate that the final model is useful for predicting the response of drivers who were not part of the estimation sample and who lived in a different country.

3.2.2. Maximum deceleration

The relationships between maximum deceleration, traffic conditions, infrastructure characteristics, driver state and characteristics, and pedestrian behaviour, state and characteristics were explored using correlation analysis. A high maximum deceleration was observed when the acceleration at the beginning of the interaction was low (r_P = -0.49, p_P = < 0.0005; r_S = -0.57, p_S = < 0.0005), the pedestrian glanced towards the vehicle (r_P = 0.17, p_P = 0.173; r_S = 0.23, p_S = 0.067), the pedestrian was a child, a teenager or an elderly (r_P = 0.18, p_P = 0.175; r_S = 0.24, p_S = 0.068), other pedestrians had already crossed before (r_P = 0.23, p_P = 0.054; r_S = 0.14; p_S = 0.25), the intersection priority rules were based on traffic signs and road markings (r_P = 0.22, p_P = 0.078; r_S = 0.22, p_S = 0.070), the pedestrian crossing had zebras with traffic signs or warning lights (r_P = 0.30, p_P = 0.012; r_S = 0.31, p_S = 0.009), and the driver was young (r_P = -0.21, p_P = 0.089; r_S = -0.25, p_S = 0.040). A low maximum deceleration was observed when the pedestrian crossed and the driver had a green traffic light (r_P = -0.19, p_P = 0.116; r_S = -0.24, p_S = 0.044). The correlations between the maximum deceleration and the other explanatory variables available were weak ($|r| <$ 0.20).

A regression model was developed to simultaneously capture the impact of multiple explanatory variables on the maximum deceleration. Table 10 presents the goodness-of-fit measures, and Table 11 shows the estimation results. The maximum deceleration $MaxDec_n(t)$ during the event at time *t* for driver *n* is presented in equation (4):

$$\log(MaxDec_n(t)) = \mu^D + \lambda^D_{Speed} \bullet Speed(t) + \lambda^D_{Acc} \bullet Acc(t)$$
$$+ \lambda^D_{DrivGreenLight} \bullet DrivGreenLight(t) + \lambda^D_{OthPedBef} \bullet OthPedBef(t)$$
$$+ \lambda^D_{PedChildTeenEld} \bullet PedChildTeenEld(t) + \lambda^D_{MissPedAge} \bullet MissPedAge(t)$$

Table 10
Statistics of the regression model predicting the maximum deceleration. The log-likelihood measures are computed using ML estimation methods.

Statistics	
Number of parameters associated with the explanatory variables	8
Number of drivers	28
Number of observations	69
Constant log likelihood	−87.76
Constant Akaike information criterion	179.52
Constant Bayesian information criterion	183.98
Final log likelihood	−54.08
Final Akaike information criterion	128.16
Final Bayesian information criterion	150.50
R^2	0.623

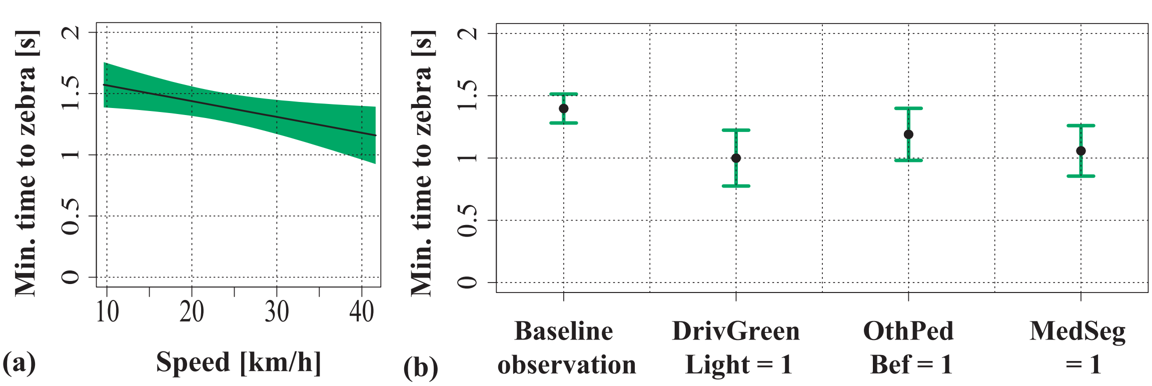

Fig. 3. Impact of the (a) continuous and (b) nominal explanatory variables on the minimum time to zebra during longitudinal interactions. *DrivGreenLight* is a binary variable equal to one when the pedestrian crossed while the driver had a green light. *OthPedBef* is a binary variable equal to one when other pedestrians were present before the beginning of the event. *MedSeg* is a binary variable equal to one when the vehicle segment was medium. Black lines indicate the marginal means and green ribbons indicate the 95% confidence intervals. The axis scales were selected depending on the minimum and maximum values in the data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 11
Estimation results of the regression model predicting the maximum deceleration. The model parameters are estimated using REML methods in the R package 'nlme', which calculates test statistics for the fixed effects only.

Variable	Description	Parameters	Estimate	t-stat.	p-value
–	Mean maximum deceleration	μ^D	0.103	1.08	0.283
Speed	Speed of the subject vehicle in km/h at the beginning of the event	λ^D_{Speed}	−0.021	−2.40	0.020
Acc	Acceleration of the subject vehicle in m/s^2 at the beginning of the event	λ^D_{Acc}	−0.507	−7.20	<0.0005
DrivGreenLight	Binary variable equal to one when the pedestrian crossed while the driver had a green light	$\lambda^D_{DrivGreenLight}$	−0.619	−3.19	0.002
OthPedBef	Binary variable equal to one when other pedestrians were present before the beginning of the event	$\lambda^D_{OthPedBef}$	0.776	4.15	0.001
PedChildTeenEld	Binary variable equal to one when the pedestrian was a child, teenager or elderly	$\lambda^D_{PedChildTeenEld}$	0.517	2.11	0.039
MissPedAge	Binary variable equal to one when the pedestrian age was missing	$\lambda^D_{MissPedAge}$	−0.608	−2.58	0.012
PedGlanceVeh	Binary variable equal to one when the pedestrian glanced towards the vehicle	$\lambda^D_{PedGlanceVeh}$	0.592	2.48	0.016
MissPedGlance	Binary variable equal to one when the pedestrian glance direction was missing	$\lambda^D_{MissPedGlance}$	0.577	2.02	0.047
υ^D_n	Observation-specific error term	ω^D	0.568	–	–

$$+ \lambda^D_{PedGlanceVeh} \bullet \mathrm{PedGlanceVeh}(t) + \lambda^D_{MissPedGlance} \bullet \mathrm{MissPedGlance}(t) + \omega^D \bullet \upsilon^D_n(t) \quad (4)$$

where μ^D is the mean value, λ^D is a vector of parameters related to the explanatory variables in Table 11, and ω^D is the parameter associated with the observation-specific error term $\upsilon^D_n(t) \sim N(0,1)$. A logarithmic transformation for the maximum deceleration was chosen because it more closely aligned with the empirical results and showed a significant improvement in goodness of fit compared to a linear specification. The null hypothesis that the mean and the standard deviation of the residuals come from a normal distribution could not be rejected at the 5% significance level (one-sample Kolmogorov-Smirnov test: test statistic = 0.077, p-value = 0.781).

All parameters associated with the explanatory variables in equation (4) are statistically significant at the 5% level. The correlation matrix and variance inflation factors did not detect multicollinearity between the variables in equation (4). The R^2 indicates that the explanatory variables capture 62.3% of the variability in the logarithm of the maximum deceleration. Drivers decelerated more when the speed and the acceleration at the beginning of the event were low. The longitudinal and lateral distances to the closest pedestrian did not have a significant impact and were therefore excluded from the model. Drivers decelerated less when they had a green light at the intersection and the pedestrian crossed (i.e., unprotected pedestrian crossing or illegal crossing). Drivers decelerated more when other pedestrians were present before the beginning of the event. Drivers decelerated more when the pedestrians were identified as children, teenagers, or elderly and the pedestrian glanced towards the vehicle. Neither the number nor the gender of the pedestrians had a significant impact. The country, weather conditions, type of locality, and intersection type did not influence the maximum deceleration. Controlled for the other variables, priority rules with traffic signs and road markings and pedestrian crossings with zebras did not have a significant impact. The vehicle segment, the age and gender of the driver did not have a significant effect. Finally, we tested a driver-specific error term that did not have a significant effect and was therefore not included in the model. The regression model in equation (4) was also estimated separately for each country. In this dataset, situations in which drivers had a green light at the intersection, the pedestrians were children, teenagers, or elderly, and the pedestrians glanced toward the vehicle were observed only in France. These variables were omitted from the specification when the model was estimated using UK data. The results showed that the parameters associated with the other variables did not differ significantly between countries.

We investigated the impact of changes in the explanatory variables on the maximum deceleration by calculating the maximum deceleration for observations in which only one variable was changed while keeping all the other variables fixed. In the baseline observation, the speed was equal to 23.13 km/h, the acceleration was equal to −0.97 m/s^2, the pedestrian was an adult, was looking towards the crossing and did not cross while the driver had a green traffic light, and no other pedestrians crossed before the interaction. The characteristics were selected based on the mean conditions in the sample. The maximum deceleration associated was equal to 1.08 m/s^2. Fig. 4 shows the results for the continuous and the nominal explanatory variables. When we compare all the variables, we note that the acceleration at the beginning of the event and other pedestrians crossing before the event have the largest impact on the maximum deceleration. Appendix A includes an out-of-sample validation analysis of the model presented in Table 11. Across all validation samples, the final regression models exhibit lower mean forecasting errors than those that incorporate only constants. The findings indicate that the final model is useful for predicting the response of drivers who were not part of the estimation sample and who lived in a different country.

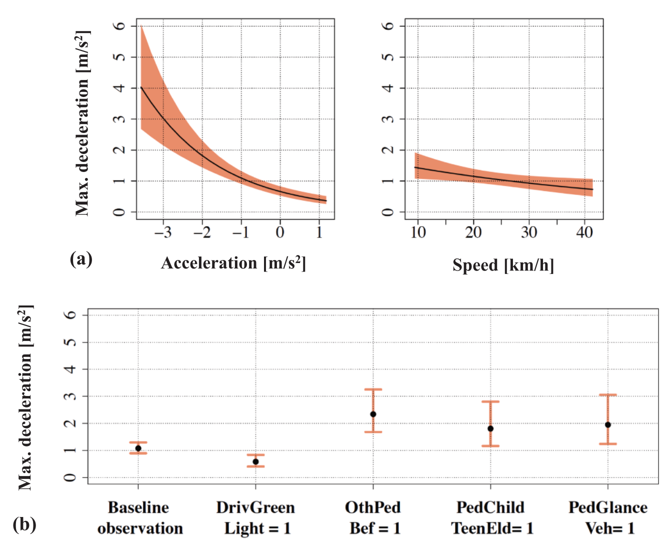

Fig. 4. Impact of the (a) continuous and (b) nominal explanatory variables on the maximum deceleration during longitudinal interactions. *DrivGreenLight* is a binary variable equal to one when the pedestrian crossed while the driver had a green light. *OthPedBef* is a binary variable equal to one when other pedestrians were present before the beginning of the event. *PedChildTeenEld* is a binary variable equal to one when the pedestrian was a child, teenager or elderly. *PedGlanceVeh* is a binary variable equal to one when the pedestrian glanced towards the vehicle. Black lines indicate the marginal means, and orange ribbons indicate the 95% confidence intervals. The axis scales were selected depending on the minimum and maximum values in the data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion and conclusions

Driver behaviour data from the UDRIVE study, collected in the UK and France as part of a European project examining driver behaviour across countries, were analysed using correlation and regression analysis to identify the factors affecting driver behaviour characteristics when pedestrians cross in front of the vehicle. The results showed that the minimum time to zebra was shorter when the speed at the beginning of the interaction was high, the pedestrian crossed while the driver had a green traffic light, the vehicle segment was medium instead of small, and other pedestrians had already crossed. The maximum deceleration was lower when the acceleration at the beginning of the interaction was high, the pedestrian crossed while the driver had a green traffic light, no other pedestrians had already crossed, the pedestrian was not a child, teenager or elderly, and the pedestrian did not glance toward the vehicle. Model estimation for each country and cross-country validation indicated that the parameters of the shared variables were consistent, while differences mainly arose in traffic situations observed in only one country. Caution is needed in interpreting the results due to the small sample of interactions and the limitations of the data collection method. The main conclusion is that accounting for the factors identified in this study could improve the realism and prediction accuracy of driver support systems and microscopic traffic simulations.

4.1. Main findings

4.1.1. Minimum time to zebra

The results of the regression model indicated that the minimum time to zebra was significantly influenced by multiple factors, even after controlling for the driver behaviour characteristics at the beginning of the interaction. The minimum time to zebra was short when the driver's speed at the beginning of the interaction was high. This finding shows that it is harder for drivers to control the driving task and maintain a safe gap at high speeds. The result is consistent with previous findings in a driver simulator experiment (Dozza et al., 2020). The time to zebra was short when the pedestrian crossed while the driver had a green traffic light. In these situations, pedestrians might cross due to unprotected pedestrian crossings (i.e., pedestrians and drivers have the green light simultaneously) or illegally (i.e., pedestrians have the red traffic light). The finding may be explained by the fact that drivers with a green traffic light do not expect pedestrians to cross at the zebras and, therefore, are less likely to maintain a safe gap. Previous studies did not analyse the impact of pedestrian crossing while the driver had a green traffic light. The time to zebra was short when other pedestrians were on the road and had already crossed before the beginning of the event. A possible explanation could be that it was harder for the drivers to maintain a safe gap because the pedestrians crossed accepting a shorter gap than in the previous interaction. Previous studies did not analyse the impact of other pedestrians crossing before the interaction. The time to zebra was shorter when the vehicle segment was medium instead of small. The data collection systems were similar in the two vehicle segments. A possible interpretation could be that the two vehicle segments had different braking responses or, considering that the participants used their private vehicles, the drivers of the two vehicle segments differed in driving styles. Previous studies did not investigate the effect of vehicle segments and driving styles on driver behaviour characteristics.

4.1.2. Maximum deceleration

The results of the regression model indicated that the maximum deceleration was significantly influenced by multiple factors, even after controlling for the driver behaviour characteristics at the beginning of the interaction. The maximum deceleration was high when the driver's speed and acceleration at the beginning of the interaction were low. These findings show that drivers often started to decelerate before the beginning of the interaction defined based on the onset of the pedestrian collision warning. The result is consistent with previous findings on time to collision at braking (Dozza et al., 2020; Lubbe & Rosén, 2014; Sun et al., 2022). When the initial acceleration was low, the maximum deceleration was high, while the minimum time to zebra remained unaffected, suggesting that drivers began decelerating early to maintain a safe gap. When the initial speed was high, a low maximum deceleration and a short minimum time to zebra were observed, indicating that drivers were not able to respond appropriately. The maximum deceleration was low when the pedestrian crossed while the driver had a green traffic light (i.e., unprotected pedestrian crossings or illegal pedestrian crossing). The finding may be explained by the fact that drivers with a green traffic light did not expect pedestrians to cross at the zebras and, therefore, were less likely to decelerate. In these conditions, a low maximum deceleration was observed together with a short minimum time to zebra, indicating that drivers were unable to respond appropriately. Previous studies did not analyse the impact of pedestrian crossing while the driver had a green traffic light. The maximum deceleration was high when other pedestrians were on the road and had already crossed before the beginning of the event. A possible explanation is that drivers decelerated more because they expected other pedestrians to cross based on prior interactions. In these situations, a high maximum deceleration and a short minimum time to zebra were observed, suggesting that drivers responded promptly. Previous studies did not analyse the impact of other pedestrians crossing before the interaction. Drivers decelerated more when the pedestrians were identified as children, teenagers, or the elderly. An explanation could be that drivers decelerated more because they expected slower crossing speeds or unpredictable behaviour from children, teenagers, and the elderly. In these conditions, there was no impact on the minimum time to zebra, suggesting that drivers responded promptly to maintain a safe gap. Previous studies did not analyse the effect of pedestrian age. Drivers decelerated more when the pedestrian glanced towards the vehicle. An interpretation could be that drivers recognised pedestrians' intention to cross more clearly. In these situations, there was no effect on the minimum time to zebra, indicating that drivers responded promptly to maintain a safe gap. Previous studies did not analyse the impact of pedestrian glance direction.

4.2. Recommendations for future research

This study investigated the main factors influencing driver behaviour characteristics when pedestrians cross during events detected by the pedestrian collision warning of the MobilEye® smart camera, which was triggered when the time to collision dropped below two seconds under sufficient lighting conditions. The current findings cannot be directly generalised to other events. Future studies are suggested to analyse driver behaviour characteristics using a larger sample of observations and to explore different thresholds. In this study, the dependent variable was a single value of the interaction characteristic (e.g., minimum time to zebra) during a specific time interval (i.e., the interaction). In future studies, the time dimension could be accounted for by using an aggregated measure of interaction characteristics (e.g., the mean time to zebra during the interaction or the percentage of time when the time to zebra is shorter than 1.5 s).

The number of events (69) and participants (28) in the current analysis was relatively small because only participants who experienced one valid longitudinal interaction were included. Further analysis based on a larger number of events is needed to assess the impact of the factors that did not have a significant effect in the regression analysis but showed a correlation in the correlation analysis. The sample characteristics (driver age, driving experience, gender, and vehicle model) did not reflect the general driving population. Caution is needed when generalising the findings towards the general driving population. For example, the results showed that the medium-vehicle segment (i.e., Renault Megane) had a shorter minimum time to zebra than the small-vehicle segment (i.e., Renault Clio) in the sample. The result cannot be directly generalised to other vehicle segments that were not included in the sample. Further investigations are required to generalise the findings to other models and makes belonging to the medium and the small-vehicle segments in the market. In this study, each participant drove only her/his private vehicle. Further analysis is needed to understand the link between driver characteristics (e.g., personality traits and driving styles) and the vehicle segment owned. Further analysis is also needed to generalise the findings to age groups not included in the sample (e.g., drivers older than 75 years). The validation analysis suggested that, to enhance the model forecasting accuracy, future research should investigate in greater depth the effect of driver characteristics on interaction characteristics.

The data were collected in the UK and France. While the models estimated for each country indicated that the parameters associated with shared variables were generally consistent, certain traffic situations (e.g., pedestrians crossing when drivers had a green light) were observed only in France. Model accuracy was lower in the cross-country validation, which may be explained by differences in road characteristics, traffic regulations and behavioural norms. The small number of interactions per country limits the generalisability of the results, and further analysis with a larger sample is required to identify country-specific effects.

Further research is also required to generalise the findings in this study to vehicles equipped with pedestrian collision warning systems and advanced driver assistance systems. In the literature, there are few naturalistic driving studies with adaptive cruise control and lane-keeping systems (Varotto et al., 2022), and the datasets are often too small to investigate interactions with pedestrians.

To assess the impact of driver distraction and non-driving task engagement on driver behaviour characteristics when pedestrians cross, future studies with larger samples are needed. In the current study, a very small percentage of drivers did not look towards the pedestrian or were involved in non-driving tasks in the 5 s before the interaction. Previous findings showing that driver inattention is a major contributing factor to near-crashes and safety-relevant incidents with pedestrians could not be confirmed (Dingus et al., 2006). The result might be explained by differences in selection criteria and thresholds for identifying interactions (e.g., short time to collision instead of large decelerations). More detailed annotation procedures can be developed to capture driver inattention (e.g., glance behaviour) and situational awareness during the interaction. This research direction should be further investigated as driver assistance systems, such as adaptive cruise control and lane-keeping systems, can potentially increase driver inattention and engagement in distracting activities. Future studies could include questionnaires and interviews to analyse factors such as familiarity with the route and time pressure that can influence driver responses but cannot be captured using video data.

4.3. Recommendations for practice

The data collection techniques (naturalistic driving data and manual annotation) and the statistical analysis methods (regression models) presented in this study are suitable for investigating interactions between drivers and pedestrians in real traffic situations. These methods allow us to observe driver behaviour with greater detail and in a broader range of traffic situations than fixed videography methods. In addition, they allow us to explicitly measure the impact of multiple factors on the minimum time to zebra and the maximum deceleration while approaching pedestrians. The results are expected to have higher external validity than findings from test-track or driving simulator studies, as they reflect real-world driving behaviour. However, participants are aware they are being observed in the instrumented vehicle, and the study design provides only limited experimental control. The findings are of interest to legislators, corporations, and academics designing intelligent transport systems and evaluating the effect of these systems on traffic operations.

Pedestrian collision warning and driver assistance systems may increase their forecasting accuracy and control performances by incorporating the factors identified in this study. Based on the results, pedestrian collision warning systems and infrastructure-to-vehicle communication systems should be developed to track pedestrian movements at intersections with traffic lights and warn drivers in case of potential unprotected or illegal crossings. A warning could be generated when pedestrians have crossed in non-safety relevant situations, as other pedestrians might cross afterwards. An early warning could be generated when the pedestrians are children, teenagers and the elderly. Drivers could be warned using in-vehicle messages and specific traffic signs on the infrastructure. From an infrastructure point of view, paying closer attention to pedestrian flows and waiting times in traffic light design could reduce the chances of illegal pedestrian crossing at intersections. Information campaigns could raise pedestrian awareness of the risks of illegal crossing at intersections and the importance of glancing towards the vehicle before crossing. Although further investigations based on a larger sample are needed, the findings suggest that differences between vehicle segments, countries and individual drivers should be accounted for in designing systems supporting drivers. Systems incorporating these results should be accepted by drivers in a wider range of traffic conditions and have a beneficial impact on traffic operations.

The realism and forecasting accuracy of microscopic traffic simulations investigating the effect of interactions with pedestrians on traffic operations could be improved by including the factors identified in this study. The findings have shown a large variability between and within drivers in the responses to pedestrians that can be explained by the traffic conditions, the pedestrian behaviour and the vehicle characteristics. Few traffic simulations have implemented empirical findings to describe driver interactions with pedestrians (Chen et al., 2019; Lu et al., 2016). Based on the current results, future research should focus on developing mathematical models that describe the longitudinal movement of drivers while approaching pedestrians as a function of the distance to the pedestrians and of the instantaneous behaviour characteristics. This advanced mathematical model should explicitly capture adaptations to the characteristics and behaviour of the pedestrian (e.g., crossing when the driver has a green traffic light). The advanced driver behaviour model can be implemented into a microscopic traffic simulation to forecast the effect of interactions with pedestrians on traffic operations with a higher level of prediction accuracy.

CRediT authorship contribution statement

Silvia F. Varotto: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Girish Kumaar Srinivasan Ravi Kumar:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization. **Ruomei Liu:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization. **Matheus H.W. Stratermans:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization. **Reyzha Wikki Zaninshi:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization. **Eleonora Papadimitriou:** Writing – review & editing, Supervision. **Meng Wang:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was partially funded by the Dutch Ministry of

Infrastructure and Water Management. The authors are grateful to Tsippy Lotan, Oren Musicant and Alexandra Harel at Or Yarok for sharing the annotation conducted within the UDRIVE project, which represented the starting point for the analysis presented in this study. The authors thank Willem Vlakveld at SWOV Institute for Road Safety Research for his valuable comments on the original manuscript.

Appendix A. Validation analysis

A validation analysis was executed in which the final regression models in Table 9 and Table 11 were compared to a regression model that included only a constant. The objective is to assess the ability of the final regression models to forecast the behaviour of drivers not included in the estimation sample and of drivers in a different country. To assess the forecasting abilities of the final models, we should apply them to other empirical databases. We implemented two out-of-sample cross-validation approaches because, currently, we do not have access to other similar databases, and the number of drivers and countries available is small (Hastie et al., 2009). In the first approach, the drivers were divided into five groups randomly. We estimated the models using observations from four groups (80% of the drivers) and validated them using observations from the group not included in the estimation sample (20% of the drivers). The procedure was repeated five times (five-fold cross-validation). In the second approach, we estimated the models using observations from drivers in one country and validated them using observations from drivers in the other country. The procedure was repeated for each country (cross-validation). We have chosen the Root Mean Square Error (RMSE) as an evaluation metric to assess the model performance on the validation samples. This metric indicates the forecasting errors between the minimum time to zebra (maximum acceleration) forecasted by the model and the minimum time to zebra (maximum acceleration) observed in the data. The accuracy of the model is higher when the RMSE is smaller. The improvement in accuracy was calculated as the percentage improvement in the RMSE of the final model in comparison with the RMSE of the model with only a constant.

Table A1, and Table A2, present the RMSE of the models predicting the minimum time to zebra on the validation samples. The improvement in accuracy is shown in the last column of each table. The findings show that the final regression models across most validation samples have smaller mean forecasting errors than the models with only a constant. Comparing the different groups of drivers in Table A1, we note that the final model predicting the minimum time to zebra shows a similar accuracy when validated in group (2) and a reduction in accuracy when validated in group 5. The findings indicate that some drivers in these groups responded differently from others in maintaining a minimum time to zebra. Personality traits or driving styles might explain these differences between drivers, and further research is needed to explain their origin. In Table A2, the final model showed a smaller accuracy improvement when validated based on the data in France. The prediction error was large when the pedestrian crossed while the driver had a green traffic light. The number of interactions in the UK was small, and these traffic situations were not observed. This dissimilarity between countries might be explained by road characteristics, traffic regulations (e.g., right- or left-hand traffic), or country-specific behaviour. Further investigation is needed based on a larger sample.

Table A3, and Table A4, present the RMSE of the models predicting the maximum deceleration on the validation samples. The final regression models across all validation samples have smaller mean forecasting errors than the models with only a constant. Comparing the different groups of drivers in Table A3, we note that the final model predicting the maximum deceleration shows a smaller accuracy improvement when validated in groups 3 and 5. The result indicates that some drivers in these groups decelerated differently from others when pedestrians crossed. In Table A4, the final model showed a smaller accuracy improvement when validated based on the data in France. The interactions in the UK did not include traffic situations in which pedestrians crossed while the driver had a green traffic light, the pedestrians were children, teenagers or elderly, and the pedestrian glanced toward the vehicle. Further investigation is needed based on a larger sample. The main conclusion from this analysis is that, in most situations, the final regression models proposed are useful to forecast the responses of drivers not included in the estimation sample and of drivers in a different country.

Table A1

Five-fold cross-validation of the regression model forecasting the minimum time to zebra for different driver groups. $\hat{\beta}$ indicates the regression model, and C denotes the model that includes only the constant.

Validation subsample	Drivers	Observations	RMSE (c)	RMSE $(\hat{\beta})$	$\frac{\text{RMSE(c)-RMSE}(\hat{\beta})}{\text{RMSE(c)}}$
1	5	15	0.356	0.248	0.3030
2	5	13	0.396	0.399	−0.0090
3	6	14	0.449	0.414	0.0783
4	6	13	0.489	0.387	0.2088
5	6	14	0.442	0.477	−0.0790
M	5.60	13.8	0.426	0.385	0.1004
SD	0.49	0.75	0.046	0.084	0.1561

Table A2

Cross-validation of the regression model forecasting the minimum time to zebra for a different country. $\hat{\beta}$ indicates the regression model, and C denotes the model that includes only the constant.

Validation subsample	Drivers	Observations	RMSE (c)	RMSE $(\hat{\beta})$	$\frac{\text{RMSE(c)-RMSE}(\hat{\beta})}{\text{RMSE(c)}}$
France	18	57	0.451	0.427	0.0520
UK	10	12	0.425	0.409	0.0380

Table A3
Five-fold cross-validation of the regression model forecasting the maximum deceleration for different driver groups. $\widehat{\beta}$ indicates the regression model, and C denotes the model that includes only the constant.

Validation subsample	Drivers	Observations	RMSE (c)	RMSE $(\widehat{\beta})$	$\frac{\text{RMSE}(c)-\text{RMSE}(\widehat{\beta})}{\text{RMSE}(c)}$
1	5	15	0.948	0.686	0.2763
2	5	13	0.788	0.509	0.3545
3	6	14	0.900	0.732	0.1864
4	6	13	0.876	0.383	0.5631
5	6	14	0.838	0.704	0.1599
M	5.60	13.8	0.870	0.603	0.3081
SD	0.49	0.75	0.061	0.150	0.1620

Table A4
Cross-validation of the regression model forecasting the maximum deceleration for a different country. $\widehat{\beta}$ indicates the regression model, and C denotes the model that includes only the constant.

Validation subsample	Drivers	Observations	RMSE (c)	RMSE $(\widehat{\beta})$	$\frac{\text{RMSE}(c)-\text{RMSE}(\widehat{\beta})}{\text{RMSE}(c)}$
France	18	57	0.888	0.756	0.1495
UK	10	12	0.877	0.638	0.2730

Data availability

The authors do not have permission to share data.

References

Angioi, F., Bassani, M., 2022. The implications of situation and route familiarity for driver-pedestrian interaction at uncontrolled mid-block crosswalks. Transport. Res. F: Traffic Psychol. Behav. 90, 287–299. https://doi.org/10.1016/j.trf.2022.09.003.
Bärgman, J., van Nes, N., Christoph, M., Jansen, R., Heijne, V., Carsten, O., Dotzauer, M., Utech, F., Svanberg, E., Pereira, M., Forcolin, F., Kovaceva, J., Guyonvarch, L., Hibberd, D., Lotan, T., Winkelbauer, M., Sagberg, F., Stemmler, E., Gellerman, H., … Fox, C. (2017). *UDrive deliverable D41.1: The UDrive dataset and key analysis results.* (1st edn). UDRIVE Consortium. https://doi.org/10.26323/UDRIVE_D41.1.
Bella, F., Silvestri, M., 2015. Effects of safety measures on driver's speed behavior at pedestrian crossings. Accid. Anal. Prev. 83, 111–124. https://doi.org/10.1016/j.aap.2015.07.016.
Bella, F., Silvestri, M., 2021. Vehicle–pedestrian interactions into and outside of crosswalks: effects of driver assistance systems. Transport 36 (2), 98–109. https://doi.org/10.3846/transport.2021.14739.
Carsten, O., Kircher, K., Jamson, S., 2013. Vehicle-based studies of driving in the real world: the hard truth? Accid. Anal. Prev. 58, 162–174. https://doi.org/10.1016/j.aap.2013.06.006.
Chen, P., Zeng, W., Yu, G., 2019. Assessing right-turning vehicle-pedestrian conflicts at intersections using an integrated microscopic simulation model. Accid. Anal. Prev. 129, 211–224. https://doi.org/10.1016/j.aap.2019.05.018.
Dingus, T., Klauer, S., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., & Gupta, S. (2006). *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment.* Department of Transportation. National Highway Traffic Safety.
Dozza, M., Boda, C.-N., Jaber, L., Thalya, P., Lubbe, N., 2020. How do drivers negotiate intersections with pedestrians? the importance of pedestrian time-to-arrival and visibility. Accid. Anal. Prev. 141, 105524. https://doi.org/10.1016/j.aap.2020.105524.
El-Basyouny, K., Sayed, T., 2013. Safety performance functions using traffic conflicts. Saf. Sci. 51 (1), 160–164. https://doi.org/10.1016/j.ssci.2012.04.015.
Habibovic, A., Tivesten, E., Uchida, N., Bärgman, J., Ljung Aust, M., 2013. Driver behavior in car-to-pedestrian incidents: an application of the driving reliability and error analysis method (DREAM). Accid. Anal. Prev. 50, 554–565. https://doi.org/10.1016/j.aap.2012.05.034.
Hastie, T., Tibshirani, R., Friedman, J.H., 2009. *The elements of statistical learning* (2nd ed). Springer.
Jansen, R., Lotan, T., Winkelbauer, M., Bärgman, J., Kovaceva, J., Donabauer, M., Pommer, A., Musicant, O., Harel, A., Wesseling, S., Christoph, M., van Nes, N., & de Goede, M. (2017). *UDRIVE deliverable 44.1 Interactions with vulnerable road users, of the EU FP7 Project UDRIVE* (1st edn). UDRIVE Consortium. https://doi.org/10.26323/UDRIVE_D44.1.
Johnsson, C., Laureshyn, A., De Ceunynck, T., 2018. In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. Transp. Rev. 38 (6), 765–785. https://doi.org/10.1080/01441647.2018.1442888.
Lu, L., Ren, G., Wang, W., Chan, C.-Y., Wang, J., 2016. A cellular automaton simulation model for pedestrian and vehicle interaction behaviors at unsignalized mid-block crosswalks. Accid. Anal. Prev. 95, 425–437. https://doi.org/10.1016/j.aap.2016.04.014.
Lubbe, N., Davidsson, J., 2015. Drivers' comfort boundaries in pedestrian crossings: a study in driver braking characteristics as a function of pedestrian walking speed. Saf. Sci. 75, 100–106. https://doi.org/10.1016/j.ssci.2015.01.019.
Lubbe, N., Rosén, E., 2014. Pedestrian crossing situations: quantification of comfort boundaries to guide intervention timing. Accid. Anal. Prev. 71, 261–266. https://doi.org/10.1016/j.aap.2014.05.029.
Lüdecke, D., 2018. ggeffects: tidy data frames of marginal effects from regression models. Journal of Open Source Software 3 (26), 772. https://doi.org/10.21105/joss.00772.
Markkula, G., Romano, R., Madigan, R., Fox, C.W., Giles, O.T., Merat, N., 2018. Models of human decision-making as tools for estimating and optimizing impacts of vehicle automation. Transportation Research Record: Journal of the Transportation Research Board 2672 (37), 153–163. https://doi.org/10.1177/0361198118792131.
Perez, M.A., Sudweeks, J.D., Sears, E., Antin, J., Lee, S., Hankey, J.M., Dingus, T.A., 2017. Performance of basic kinematic thresholds in the identification of crash and near-crash events within naturalistic driving data. Accid. Anal. Prev. 103, 10–19. https://doi.org/10.1016/j.aap.2017.03.005.
Pinheiro, J., Bates, D., & R Core Team. (2024). *nlme: Linear and Nonlinear Mixed Effects Models* (p. 3.1-165) [Data set]. https://doi.org/10.32614/CRAN.package.nlme.
Portera, A., Angioi, F., Di Stasi, L.L., Bassani, M., 2024. Effectiveness of smart LED strips at mid-block crosswalks under distracted driving conditions. Transp. Eng. 16, 100253. https://doi.org/10.1016/j.treng.2024.100253.
Sheykhfard, A., Haghighi, F., 2018. Behavioral analysis of vehicle-pedestrian interactions in Iran. Sci. Iran. 25 (4), 1968–1976. https://doi.org/10.24200/sci.2017.4201.
Sheykhfard, A., Haghighi, F., Bakhtiari, S., Pariota, L., 2023. Safety margin evaluation of pedestrian crossing through critical thresholds of surrogate measures of safety: area with zebra crossing versus area without zebra crossing. Transportation Research Record: Journal of the Transportation Research Board 2677 (1), 396–408. https://doi.org/10.1177/03611981221099510.
Sheykhfard, A., Haghighi, F., Papadimitriou, E., Van Gelder, P., 2021. Review and assessment of different perspectives of vehicle-pedestrian conflicts and crashes: passive and active analysis approaches. Journal of Traffic and Transportation Engineering (english Edition) 8 (5), 681–702. https://doi.org/10.1016/j.jtte.2021.08.001.
Sun, S., Zhang, Z., Zhang, Z., Deng, P., Tian, K., Wei, C., 2022. How do human-driven vehicles avoid pedestrians in interactive environments? A Naturalistic Driving Study. Sensors 22 (20), 7860. https://doi.org/10.3390/s22207860.
Tian, R., Chien, S., Chen, Y., Sherony, R., 2019. Pedestrian moving patterns during potential conflicts with 110 on-road driving vehicles. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 63 (1), 2036–2040. https://doi.org/10.1177/1071181319631434.
Tian, R., Li, L., Yang, K., Chien, S., Chen, Y., Sherony, R., 2014. Estimation of the vehicle-pedestrian encounter/conflict risk on the road based on TASI 110-car naturalistic driving data collection. IEEE Intelligent Vehicles Symposium Proceedings 2014, 623–629. https://doi.org/10.1109/IVS.2014.6856599.

Tian, R., Li, L., Yang, K., Jiang, F., Chen, Y., Sherony, R., 2015. Single-variable scenario analysis of vehicle-pedestrian potential crash based on video analysis results of large-scale naturalistic driving data. In: Duffy, V.G. (Ed.), Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management: Ergonomics and Health. Springer International Publishing, pp. 295–304, 10.1007/978-3-319-21070-4_30.

Van Lint, J.W.C., Calvert, S.C., 2018. A generic multi-level framework for microscopic traffic simulation—Theory and an example case in modelling driver distraction. Transp. Res. BMethodol. 117, 63–86. https://doi.org/10.1016/j.trb.2018.08.009.

Van Nes, N., Bärgman, J., Christoph, M., Van Schagen, I., 2019. The potential of naturalistic driving for in-depth understanding of driver behavior: UDRIVE results and beyond. Saf. Sci. 119, 11–20. https://doi.org/10.1016/j.ssci.2018.12.029.

Várhelyi, A., 1998. Drivers' speed behaviour at a zebra crossing: a case study. Accid. Anal. Prev. 30 (6), 731–743. https://doi.org/10.1016/S0001-4575(98)00026-8.

Varotto, S.F., Mons, C., Hogema, J.H., Christoph, M., Van Nes, N., Martens, M.H., 2022. Do adaptive cruise control and lane keeping systems make the longitudinal vehicle control safer? Insights into speeding and time gaps shorter than one second from a naturalistic driving study with SAE Level 2 automation. Transp. Res. Part C Emerging Technol. 141, 103756. https://doi.org/10.1016/j.trc.2022.103756.

Wu, J., Radwan, E., Abou-Senna, H., 2018. Assessment of pedestrian-vehicle conflicts with different potential risk factors at midblock crossings based on driving simulator experiment. Adv. Transp. Stud. 44, 33–46. https://doi.org/10.4399/97888255143463.

Yang, Y., Lee, Y.M., Kalantari, A.H., De Pedro, J.G., Horrobin, A., Daly, M., Solernou, A., Holmes, C., Markkula, G., Merat, N., 2024. Using distributed simulations to investigate driver-pedestrian interactions and kinematic cues: implications for automated vehicle behaviour and communication. Transport. Res. F: Traffic Psychol. Behav. 107, 84–97. https://doi.org/10.1016/j.trf.2024.08.027.