WORKING PAPER

ITLS-WP-20-18

Comparing the usefulness of a connected environment during mandatory and discretionary lane-changings

By
Yasir Ali[a], Michiel C.J. Bliemer[a], Zuduo Zheng[b], Md. Mazharul Haque[c]

[a] Institute of Transport and Logistic Studies (ITLS), The University of Sydney Business School, Sydney Australia
[b] School of Civil Engineering, The University of Queensland, Australia.
[c] Queensland University of Technology, Science and Engineering Faculty, School of Civil & Environmental Engineering, Brisbane, Queensland, Australia

September 2020

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES
The Australian Key Centre in
Transport and Logistics Management

The University of Sydney
Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-20-18

TITLE: **Comparing the usefulness of a connected environment during mandatory and discretionary lane-changings**

ABSTRACT: Lane-changing manoeuvre is one of the risky manoeuvres performed by drivers either to reach the planned destination (i.e., mandatory lane-changing; MLC) or to achieve better driving conditions (i.e., discretionary lane-changing; DLC). Essentially both lane-changing types require the driver to acquire surrounding traffic information for efficient and safe lane-changing decisions. However, this does not discount the fact that both these lane-changings are fundamentally different from each other as the urgency of lane-changing is much higher during MLC compared to DLC. A connected environment promises to assist during the lane-changing decision-making process, but the differential effectiveness (or usefulness) of a connected environment for these two lane-changing types remains unexplored due to the novelty of a connected environment and the consequent scarcity of data. To fill this research gap, this study collected lane-changing data from 78 participants who performed MLC and DLC in the CARRS-Q Advanced Driving Simulator. Participants were asked to drive in three randomised driving conditions: baseline condition (without driving aids), connected environment with perfect communication, and connected environment with communication delay. While surrogate measures of safety are analysed and compared using descriptive statistics, a hybrid framework of data mining and classical statistical modelling is employed to examine the usefulness of the connected environment for two lane-changing types. We find that the crash risk associated with MLC is significantly reduced in the connected environment driving conditions compared to that of DLC. Results also reveal that the probability of engaging in a hard-braking event decreases for both the lane-changing types during the connected environment driving conditions, but a higher decrease in magnitude is found for MLC. Age and gender-related differential impact have been observed where young and male drivers have a higher possibility of engaging in a hard-braking event when driving without driving aids, but the connected environment reduces such risk. This study concludes that the usefulness (or effectiveness) of the connected environment is a function of the urgency of a task, which is evidently higher during MLC, thus providing the maximum advantage during MLC.

KEY WORDS: ***Mandatory lane-changing; discretionary lane-changing; connected environment; advanced driving simulator; braking; crash risk.***

AUTHORS:	**Ali, Bliemer, Zheng, Haque**

Acknowledgements:	This research is partly funded by Australian Research Council grants DE160100449 and LP160101021.

CONTACT:	INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (H73) The Australian Key Centre in Transport and Logistics Management The University of Sydney NSW 2006 Australia Telephone: +612 9114 1824 E-mail: business.itlsinfo@sydney.edu.au Internet: http://sydney.edu.au/business/itls

DATE:	September 2020

1. Introduction

A connected environment promises to improve some of the major negative road externalities such as congestion, hazards, and air pollution (e.g., gas emissions). One possible application of a connected environment emerges for the lane-changing decision-making process, which is among riskier manoeuvres required of drivers. Connected environment's impact on lane-changing, however, is still unexplored due to the novelty of a connected environment and the consequent scarcity of data.

Lane-changing has a significant impact on traffic flow characteristics and traffic safety. For instance, lane-changing is associated with triggering congestion (Arai and Sentinuwo, 2012), linked to creating bottleneck that causes traffic breakdown (Wall and Hounsell, 2005), formation of stop-and-go oscillations (Ahn and Cassidy, 2007), capacity drop with shockwaves (Sasoh and Ohara, 2002), and many others. Zheng et al. (2011b) and Zheng et al. (2013) showed strong empirical evidence of lane changing's negative impact on surrounding traffic and driver characteristics in particular. Similarly, lane-changing is also one of the reasons for collisions (e.g., rear-end and sideswipes) (Sen et al., 2003). During 2018, about 2,530 and 827 rear-end and sideswipe collisions were reported, respectively, in New South Wales, Australia (TfNSW, 2019). These daunting statistics confirm the importance of lane-changing in traffic flow efficiency and traffic safety, and thus analysing and modelling lane-changing behaviour become critical and have motivated a large body of literature (Ahmed, 1999, Hidas, 2002, Toledo et al., 2003, Hidas, 2005, Choudhury et al., 2006, Choudhury et al., 2007, Choudhury et al., 2009, Zheng, 2014, Ali et al., 2019b).

In traffic flow theory, lane-changing is classified as mandatory lane-changing (compulsory, MLC) and discretionary lane-changing (voluntary, DLC). MLC is often carried out to reach a specific (or planned) destination, e.g., merging into motorway traffic from the acceleration lane, or changing lanes to diverge from the motorway. On the other hand, DLC is performed to achieve desired driving conditions, e.g., gaining speed advantage by avoiding a slow-moving (or heavy) lead vehicle. This study focusses on both types of lane-changing on motorways in a connected environment.

Compared with other daily routine driving tasks, lane-changing is more complex, requiring lane-changers (or decision-makers) to recognise surrounding traffic conditions (e.g., speed of and distance to the head vehicle in the current driving lane, and the gaps of the lead and lag vehicles in the target lane), revealing its lane-changing intentions to others, and a lane-changing execution. A successful lane-changing decision elevates mental workload and stress, resulting in increased uncertainty and decision errors during the lane-changing decision-making process; this also makes driving more error-prone and dangerous. To this end, driving aids provided by a connected environment are anticipated to increase safety associated with the lane-changing decision-making. However, whether drivers accept these driving aids (and consequently take advantage of a connected environment) and change their driving behaviours accordingly, are some of the questions that have direct implications on the success of a connected environment because the anticipated benefits of a connected environment are a function of human factors (Sharma et al., 2017).

A sound understanding of different lane-changing decisions and the capability to model it under different conditions is the absolute minimum (and the prerequisite) for the success of a connected environment, which may revolutionise the current transportation systems. There is a large body of literature that has analysed and modelled MLC and DLC separately in a traditional environment because of the distinct decision-making mechanisms involved in both types of lane-changing (Ahmed et al., 1996, Ahmed, 1999, Hidas, 2002, Toledo et al., 2003, Hidas, 2005, Choudhury et al., 2006, Bham and Goswami, 2007, Choudhury et al., 2007, Bham, 2009, Choudhury et al., 2009, Marczak et al., 2013, Balal et al., 2014, Balal et al., 2016, Vechione et al., 2018, Ali et al., 2018, Ali et al., 2019a, Ali et al., 2019b, Ali et al., 2020d, Hess et al., 2020); however, only a few studies (Ali et al., 2018, Ali et al., 2019a, Ali et al., 2019b, Ali et al., 2020c,) have focused on the connected environment primarily because connected vehicles (or vehicles operating under a connected environment) are not deployed in the field at a large scale, which restricts much of our operational research. Although some prior research has documented the impact of a connected environment (or connected vehicles) on macroscopic benefits of traffic (Gandhi et al., 2014, Guériau et al., 2016, Guler et al., 2014, Rios-Torres and Malikopoulos, 2017), its impact on microscopic behaviour and a comparison of both lane-changings are yet to be analysed. Albeit these past studies describe a positive impact of connected vehicle technology on lane-changing using numerical simulations, which is a reasonable compromise to real data when unavailable, findings from past research are oversimplified because a critical component—human factors that play a central role in the lane-changing decision-making process—is often ignored (or not accounted for). Furthermore, MLC has an urgency factor (in terms of remaining distance in the acceleration lane) due to which drivers may select a risky gap to avoid a complete halt at the end of the acceleration lane and lead to safety-critical events. On the contrary, DLC has endogenous (e.g., unsatisfaction with the driving lane conditions) as well as exogenous factors (e.g., speed advantage in the adjacent lane causing temptation to change lanes) that are mainly driven by personality traits (Ali et al., 2020c). Given these differences, it would be interesting to examine whether a connected environment has the same benefits for these distinct decision-making processes or whether there exists a differential impact of a connected environment on both types of lane-changing. In addition, it is also worth investigating how the safety benefits (or margins) vary when there is an impairment in the functioning of a connected environment because communication can be compromised by a variety of factors similar to impairments in the functioning of mobile phone technology.

As such, the objective of this study is to compare and quantify the safety impacts of the connected environment on MLC and DLC using real trajectory data obtained from the Advanced Driving Simulator experiment, which is designed to mimic driving conditions in a connected environment. To this end, the remainder of the paper is organised as follows. Section 2 explains the experimental plan, including driving simulator, scenario design, participant details, and data collection procedure. Section 3 describes data processing, surrogate measures of safety, and data mining and statistical modelling techniques used. Section 4 presents descriptive analysis and modelling results, whereas Section 5 compares the benefits of the connected environment for MLC and DLC. Finally, Section 6 concludes the study and provides an outlook for future research.

2. Experimental plan and details

In this study, an innovative driving simulator experiment was designed to collect data related to MLC and DLC manoeuvres. Each participant was asked to perform these manoeuvres in three randomised driving conditions. These conditions are (a) baseline driving (without driving aids), (2) connected environment driving with perfect communication (CE_PC), and (3) connected environment driving with communication delay (CE_CD). Whilst the first condition serves as the '*default*' driving condition to which driving performance and the data quality is compared, the second condition allows the assessment of the lane-changing decision-making in a fully functioning connected environment (i.e., CE_PC) and the third condition enables the evaluation of the impact of a poorly functioning connected environment (i.e., CE_CD) on the lane-changing decision-making.

For the data collection purpose, the Centre for Accident Research and Road Safety-Queensland (CARRS-Q) high fidelity Advanced Driving Simulator was used. More details on the driving simulator can be found in our previous works (Ali et al., 2018, Ali et al., 2019a, Ali et al., 2020c).

In this study, 78 participants were recruited from the general public with a diverse background. The mean age of the participant was 30.8 years (standard deviation [SD] 11.7 years), while the mean age for male and female participants was respectively 34.1 years (SD 12.6 years) and 24.9 years (SD 6.7 years).

It is worth mentioning here that the experiment design has been carefully designed, resulting in generating high-quality data, which have been used in our previous publications addressing various research gaps (Ali et al., 2018, Ali et al., 2019a, Ali et al., 2020b, Ali et al., 2020c, Ali et al., 2020a). While ensuing subsections briefly summarise the experimental setup and design of vehicular interactions, details related to these have been omitted to avoid overlapping with our previous works.

2.1 Experimental setup

The designed driving route in the experiment consists of a 3.2 km motorway segment with two lanes in each direction. The posted speed limit on the motorway was 100 km/h. To minimise driving sequence bias (i.e., learning effect), the driving conditions were randomised for each participant. Ensuing subsections detail the three conditions.

Baseline driving condition

During this condition, each driver drives the simulator car without any driving aids provided by a connected environment and performs MLC as well as DLC. The entire motorway is segmented into three sections (Figure 1(a)), namely MLC scenario, stopped vehicle road portion after driving in the work zone, and DLC scenario. Ensuing paragraphs detail vehicular interactions in each of these sections.

In Section 1, participants face a lane closure (either due to work zone or broken vehicle) at about 500 m from the start of the scenario where they perform an MLC (Figure 1(b)). Participants recognise the lane closure when the programmed leading vehicle (LV_1) in current lane 2 changes lane to adjacent lane 1. The programmed following vehicles (FVs) in lane 1 are scripted to drive at the same speed as the speed of the subject vehicle (SV, driven by the participant) to ensure that all participants face similar gap sizes at the same position. During the MLC event, each participant has five merging opportunities for lane-changing: 45 m, 15 m, 30 m, 60 m, and 90 m.

Assuming that SV selects the first gap (although participants can select any of these available gaps) and moves to the adjacent lane and enters into a lane closure (Figure 1(c)), all other vehicles will follow SV with predefined speed. After travelling about 200 m from the start of lane closure, the lane closure ends and there is a stopped truck and SV is required to move to lane 2 where a DLC event may occur.

Once the participant moves to lane 2, LV_1 is scripted to maintain a spacing of 30 m from SV on lane 2 while the distance (or available gap) between the minivan (shown by yellow vehicle on lane 2) and FV_1 in lane 1 is 60 m. Since the spacing on lane 2 is smaller and the available gap size on lane 1 is larger, this triggers a DLC opportunity as the participants are likely to gain the speed advantage on lane 1 (overtaking in the left lane is allowed in Australia). If SV decides to remain in lane 2, minivan moves to SV's lane (lane 2) and starts moving at 50 km/h (Figure 1(d)). Meanwhile, FVs in lane 1 are moving fast and creating several DLC opportunities in lane 2 for SV. Similar to MLC, five gaps are presented to the participants: 60 m, 30 m, 45 m, 15 m, and 90 m. Once participants perform a DLC manoeuvre, they continue travelling on the motorway and take an exit to the city where the scenario ends.

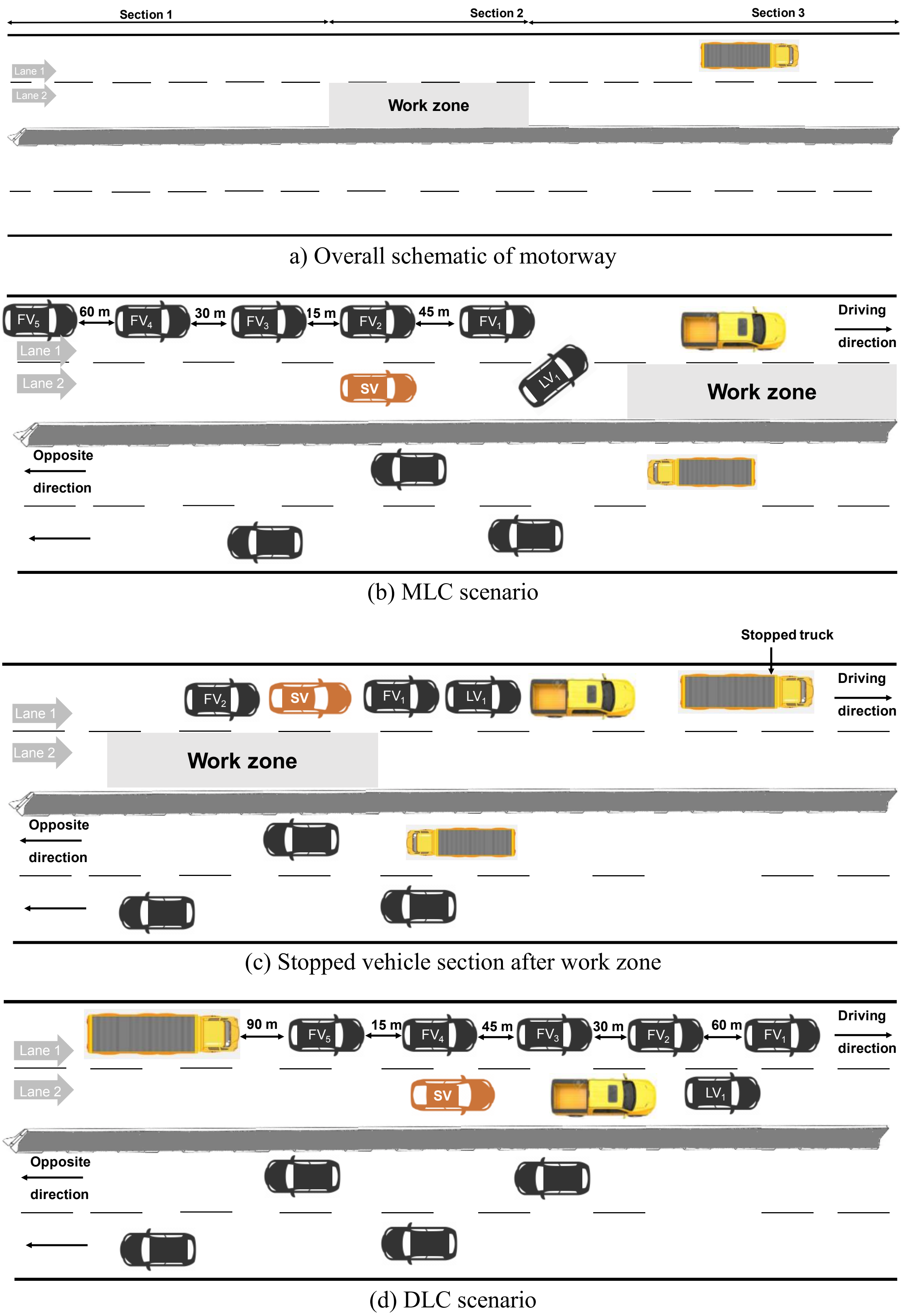


Fig. 1. Design of vehicular interactions in the experiment (not to scale)

Connected environment with perfect communication (CE_PC)

In this scenario, vehicular interactions and roadway design were the same as in the baseline driving condition. However, the participants were assisted with an uninterrupted supply of driving aids provided by the connected environment, representing vehicle-to-vehicle and vehicle-to-infrastructure communications. Driving aids are designed based on a thorough literature review on in-vehicle information systems (or advanced driving assistance systems) and the latest vehicle models fitted with information assistance systems. As such, two forms of driving aids are adopted in this experiment: auditory (beep sound) and imagery messages. This dissemination of information closely resembles the heads-up display mimicking the latest vehicle design.

Figure 2 shows some representative driving aids disseminated during a lane-changing scenario. The temporary advisory message appears at the bottom of the windscreen with a beep sound, displaying upcoming situations, such as a broken vehicle ahead or congestion ahead, which a driver cannot foresee. The lane-changing message is displayed on the left side of the windscreen with a beep sound when a lane-changing opportunity is available in the adjacent lane. Note that there were several other driving aids during the lane-changing scenario, which are not presented here due to brevity (see (Ali et al., 2018, Ali et al., 2020c) for details on these driving aids).

Connected environment with communication delay (CE_CD)

The vehicular interactions, roadway design, and design of driving aids in this scenario remained the same as in the case of perfect communication scenario (CE_PC). The only difference is the time delay (i.e., 1.5 s) in providing driving aids. This delay was selected after a series of pilot studies where different delays in providing driving aids (0.5, 1, 1.5, and 2 s) were tested, and the minimum delay was selected when the participants started to react to the delayed information. This delay is also reported to affect traffic safety negatively in a past study (Talebpour et al., 2015).

All the participants performed a practice drive prior to the start of the actual experiment to become familiar with the driving environment, simulator car, and some representative designed vehicular interactions. Once they felt confident about their driving, they were allowed to participate in the actual experiment.

A considerable amount of time and effort was dedicated to designing and implementing the experiment to manage the workload of the participants and minimise learning effects caused by repeated driving. Each participant took about 10-12 mins on average to complete a scenario, and the entire experiment finished in about 50 mins. Furthermore, the order of scenarios (as mentioned previously) was randomised except for communication delay, which came once the participants have driven in the perfect communication scenario. Moreover, although the scope of this paper is limited to lane-changing, the experiment consists of other driving tasks such as car-following and city events, which are presented and analysed elsewhere (Ali et al., 2020a). By having breaks after each drive, multiple driving tasks, and randomised driving scenario, the effects of learning are likely to be minimal across all the driving conditions.

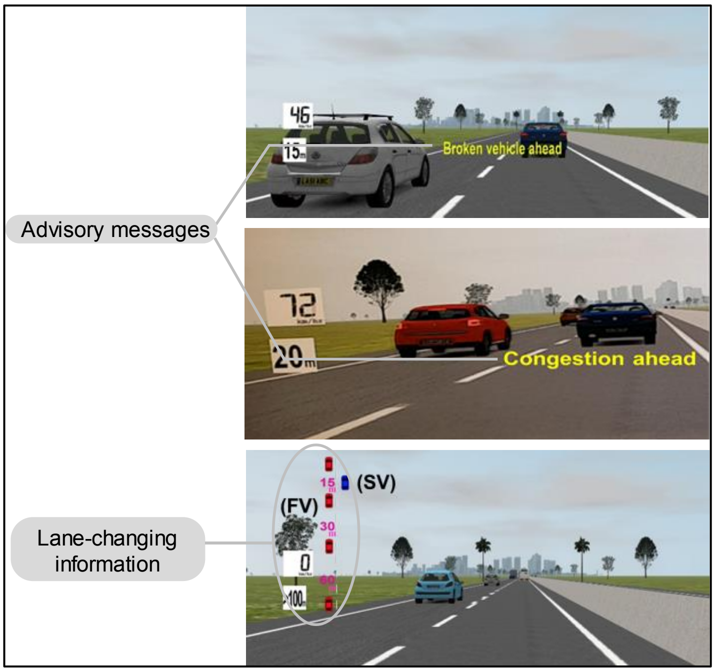


Fig. 2. Design of driving aids in the experiment

3. Data collection and processing

3.1 Dataset

Seventy-eight participants performed the driving simulator experiment across three drives and two lane-changing types, while two DLCs, which were performed close to the exit ramp, were excluded, resulting in 466 trajectories. In addition, four participants were unable to perform the third drive due to motion sickness, forming an imbalanced dataset of 458[1] observations. Recall the simulator software automatically collected vehicle trajectory data in the form of speeds, acceleration, and spacings. Driver demographic information including age, gender, driving experience, education, and previous experience with the driving assistance system, was collected using a pre-driving questionnaire.

3.1.1 Surrogate measures of safety

To study and compare the impact of the connected environment on two different types of lane-changing, two surrogate measures of safety are adopted in this study. As reported previously, an improper lane-changing decision and a poorly executed lane-changing manoeuvre are likely to result in either a rear-end collision or sideswipe. Thus, surrogate measures related to these have been considered and are presented in Table 1.

3.1.2 Data processing

In order to make proper inferences about the connected environment's impact and comparing the usefulness of the connected environment between two lane-changing types, it is of the utmost importance that data related to both these sections (i.e., MLC and DLC) should be the same. Thus, a proper methodology is adopted in this study to extract the relevant data.

[1] 78 (participants) × 3 (driving conditions) × 2 (two lane-changing types) – 4 × 2 (missing data for third drive) – 2 (obscure DLC)

Table 1. Surrogate measures of traffic safety[2]

Variable	Description
PICUD	Possibility index for collision with urgent deceleration is used to analyse the possibility of rear-end collisions assuming that the lead vehicle on the target lane applies sudden (or hard) braking. It is defined as the distance between the lead vehicle on the target lane and the lane-changing vehicle on the current lane when the lead vehicle suddenly braked, and the lane-changer has to decelerate with reaction delay to avoid a collision (Uno et al., 2002). PICUD is calculated during the LC execution period.
SSCR	Sideswipe collision risk is the sum of two sideswipes risks, namely sideswipe red and sideswipe yellow. Red and yellow risks are a function of time-to-collision and post-encroachment time, respectively (Behbahani and Nadimi, 2015). SSCR is calculated during the LC execution period.
Longitudinal acceleration	It is defined as the rate of change of speed measured during the lane-changing event (Wu and Jovanis, 2013)
Speed variation	The standard deviation of speed measured during the lane-changing event (Boonsiripant, 2009)

During a lane-changing event (as mentioned in Table 1), a driver notices the type of lane-changing (either MLC or DLC), finds a suitable gap in the adjacent lane, and prepares for lane-changing execution. Note that the lane-changing event does not include lane-changing execution (often called as lane-changing duration).

The start of the lane-changing event (either MLC or DLC) in the connected environment scenarios is obtained as the time when the first advisory driving aid (that is, 'Broken vehicle ahead' in case of MLC and 'Congestion ahead' in case of DLC) was disseminated to participants. Similarly, the spatial location where the first gap was created in the adjacent lane during the baseline condition (that is, 45 m in case of MLC and 60 m in case of DLC) was taken as the start of a lane-changing event. As a matter of fact, the creation of the first gap was considered as the criterion for providing the messages in the connected environment scenarios.

The start of lane-changing execution (or manoeuvre) is obtained by an algorithm developed in Ali et al. (2018). This algorithm is based on vehicle lane lateral shift profile, providing information on how far SV is from the lane centre. In general, lane lateral shift values remain fairly constant during car-following while they change drastically during lane-changing execution. The proposed algorithm initially finds the maximum point of the lane lateral shift profile and tends to move in a backward direction until the minimum point is located on the profile. The lowest point is called the lane-changing execution point because after this point, lane lateral shift values change significantly. The lane-changing duration is obtained as the time difference between the start of lane-changing execution (obtained from the adopted algorithm) and the time when the lane-changing vehicle was in the adjacent lane (obtained through lane numbering from the trajectory data source). It is worth mentioning here that the efficacy of the adopted algorithm is also tested and confirmed herein as well as in the past studies (Ali et al., 2018, Ali et al., 2020d).

[2] Note that longitudinal acceleration and speed variation are frequently used driving indicators and fairly simple to calculate while PICUD and SSCR measures require basic calculations, which are not presented herein for brevity. We refer the interested readers to the original sources mentioned in Table 1.

3.2 Statistical modelling

To compare the usefulness (or effectiveness) of the connected environment during two types of lane-changing and how an impairment in the provision of driving aids affects the lane-changing decision-making, two statistical models are developed and described below.

First, a deterministic model is developed for analysing speed variations during the lane-changing event. Speed variation, a measure important for reducing congestion and improving traffic safety (i.e., to avoid rear-end collisions) (FHWA, 2014), is the dependant variable in the first model measured as the standard deviation of the speed during the lane-changing event. Generalised estimation equations (GEE) approach is adopted to model speed variations, which can account for the panel nature of our data, i.e., we have multiple observations of the same participant over time. GEEs, an extended form of generalised linear models, are frequently employed to capture correlation arising from panel data because of its flexibility in accommodating non-normal and non-linear relationships within the modelling framework (Haque et al., 2016).

Second, a probabilistic model is developed for evaluating the probability of engaging in a hard-braking event when a driver's deceleration exceeds a certain threshold. The maximum longitudinal deceleration greater or lower than 0.4g (where g is the acceleration due to gravity) during the entire lane-changing event is considered as the binary dependent variable in this model. Wu and Jovanis (2013) concluded that vehicle longitudinal deceleration rate has a direct relationship with safety-critical events, which they found using naturalistic data, therefore we consider this deceleration rate as a surrogate measure of safety. A repeated measure logistic GEE model is applied to determine the probability of engaging in a hard-braking event. Let Y_{njt} be an indicator that is 1 if driver n has a deceleration rate above 0.4g in scenario j at time t, and is 0 otherwise. Applying a logit model, the probability of a deceleration rate higher than 0.4g can be obtained as

$$\Pr(Y_{njt}=1)=\frac{\exp(\mathbf{X}'_{njt}\boldsymbol{\beta})}{1+\exp\left(\mathbf{X}'_{njt}\boldsymbol{\beta}\right)}, \tag{1}$$

where, $\mathbf{X}_{njt}=[x_{njt1},\ldots,x_{njtK}]'$ is a vector of K explanatory variables, and $\boldsymbol{\beta}$ is a vector of K estimable parameters. Following Haque et al. (2016), we assume a constant correlation among multiple observations of the same participant. More details about the repeated measure logistic GEE model and estimation procedure can be found in Haque et al. (2016) and Liang and Zeger (1986). Furthermore, robust variance estimates are considered as they do not impose any restriction on the nature of correlation (either positive or negative) and can provide better parameter estimates (Zorn, 2006, Haque et al., 2016).

It is often challenging in nature to specify the best subset of explanatory variables that not only include main effects, but capture potential interactions among them, which may have a prominent impact on the response variable because of little prior knowledge of the underlying relationships. In general, the modeller provides a priori second- and higher-order interaction effects and non-linearities associated with main effects in conventional approaches before the model estimation. As reported in Haque et al. (2016), the possible combination of main effects and potential higher-order interaction effects grows geometrically and exponentially, respectively, with the number of ordinal and nominal variables. Therefore, it becomes difficult to judiciously decide the inclusion and omission of a variable from the model.

Table 2. Summary statistics of explanatory variables considered in this study

Variable	Description of variables	Count	Percentage	Mean (SD)
Driving condition				
Baseline	Driving without driving aids (reference)	78	100	—
CE_PC	Driving with driving aids (dummy)	78	100	—
CD_CD	Driving with delayed driving aids (dummy)	74	95	—
Type of lane-changing				
MLC	Performing MLC (reference)	230	100	—
DLC	Performing DLC (dummy)	228	99	—
Operational variables				
Spacing	Spacing between the subject vehicle and lead vehicle in m	—	—	24.64 (16.61)
Lag gap	Distance between the rear bumper of the subject vehicle to front bumper of the following vehicle on the adjacent lane in m	—	—	30.45 (20.74)
Demographic variables				
Age groups				
Young	Participant is 18 – 26 years old (dummy)	38	48.72	—
Middle-aged	Participant is 27 – 50 years old (reference)	32	41.02	—
Older	Participant is 51+ years old (dummy)	8	10.26	—
Gender				
Male	Participant is male (reference)	50	64.10	—
Female	Participant is female (dummy)	28	35.90	—

To circumvent this problem, this study employed a hybrid framework of data mining (i.e., decision tree) and statistical modelling (i.e., logistic GEE model) in an iterative process. At the first level, a decision tree classification is used. This is a non-parametric method to obtain possible interactions by classifying the observations in the predictor space in an iterative process. During this classification, various potential predictors exist, and each predictor receives various cut-off values (Choudhary and Velaga, 2019). However, decision trees are associated with type I error due to this multiplicity and are hard to make inferences about the underlying relationship. Despite these shortcomings, the decision tree can be employed to obtain *a priori* knowledge obtained from tree branches and can be used for determining which interaction effects to include in the logistic regression model. At the second level, the logistic model is estimated by considering the interactions from the decision tree. This combined approach allows the consideration of higher-order interaction effects (using decision tree) and makes inferences about model output (using the logistic regression model) (Washington et al., 2011, Haque et al., 2016).

Table 2 provides a summary of explanatory variables considered as input for the decision tree and the logistic GEE model. First, the model is developed by considering the main effects (from Table 2) and significant interaction effects from the decision tree and subsequently pruning the tree to identify the possible interactions again. The model is again estimated based on new interactions from the pruned tree, and this process is iterated until a logically sound and theoretically justified parsimonious model is obtained.

For easy interpretation of coefficient estimates, odds ratios (OR) are calculated as $\exp(\beta_k)$ and provide the magnitude of the relationship of the considered explanatory variable and the probability of the deceleration rate exceeding 0.4*g* (i.e., engaging in a hard-braking event). For continuous explanatory variables, an OR greater (lower) than one reveals an increased (decreased) probability of engaging in a hard-braking event. For categorical variables, the OR represents the change in the categorical variable from zero to one.

4. Results

This section presents descriptive analysis as well as modelling results. Ensuing subsections detail each of these.

4.1 Descriptive analysis of surrogate measures of safety

4.1.1 PICUD

PICUD—potential index for collision with urgent deceleration—is calculated for both MLC and DLC during the lane-changing duration (or execution period), and the minimum PICUD is selected during this period. If PICUD is less than zero, then there is a possibility of a rear-end collision between the lead vehicle in the target lane and the lane-changer (Uno et al., 2002). The frequency of participants (expressed as a percentage of the total number of participants driving in each scenario) having PICUD less than zero is calculated and presented in Table 3. For MLC, about 64%, 26%, and 46% of drivers have PICUD lower than zero during the baseline, CE_PC, and CE_CD driving conditions, respectively. This suggests a 38% and 18% decrease in PICUD in a connected environment with perfect communication (CE_PC) and communication delay (CE_CD), respectively, compared to no driving aids. The corresponding percentages of PICUD for DLC are respectively 47%, 29% and 41%, implying that the CE_PC and CE_CD driving conditions decrease PICUD by 18% and 6%, respectively, compared to the baseline condition. The differences between these frequencies are found to be significantly different as tested by chi-square tests at a 5% significance level.

Figure 3 reveals a change in PICUD frequency for different scenario comparisons (e.g., baseline versus CE_PC, baseline versus CE_CD, and CE_PC versus CE_CD). It can be observed that change in PICUD frequency between baseline and CE_PC is higher for MLC compared to DLC, suggesting that the usefulness of perfect communication is significantly higher for MLC (see the red arrow in Figure 3) compared to DLC (χ^2 = 5.11, p-value = 0.023).

Table 3. Summary of PICUD results

Lane-changing	Scenario	Frequency	Comparison	Significance by a chi-square test
MLC	Baseline	64.1%	Baseline versus CE_PC	χ^2 = 23.23 (p-value < 0.001)
	CE_PC	25.6%	Baseline versus CE_CD	χ^2 = 6.60 (p-value = 0.010)
	CE_CD	45.9%	CE_PC versus CE_CD	χ^2 = 5.55 (p-value = 0.019)
DLC	Baseline	47.4%	Baseline versus CE_PC	χ^2 = 5.96 (p-value = 0.014)
	CE_PC	29.4%	Baseline versus CE_CD	χ^2 = 1.63 (p-value = 0.20)
	CE_CD	40.5%	CE_PC versus CE_CD	χ^2 = 1.40 (p-value = 0.23)

The change in PICUD frequencies between CE_CD and baseline during MLC and DLC are respectively 18.2% and 6.9% (see the purple arrow in Figure 3). A chi-square test indicates that the communication delay has a prominent impact on MLC compared to DLC (χ^2 = 4.08, p-value = 0.04), implying that communication impairment impacts MLC more compared to DLC. By comparing the differential impact of the connected environment on driving conditions (CE_PC versus CE_CD), we find that a delay in the supply of driving aids tends to increase the frequency of PICUD for MLC compared to DLC. We elaborate on this finding further in the next section.

Similar results have been found when comparing the differential effects of connected environment scenarios on PICUD frequency as indicated by the blue arrow in Figure 3 (χ^2 = 2.51, p-value = 0.11).

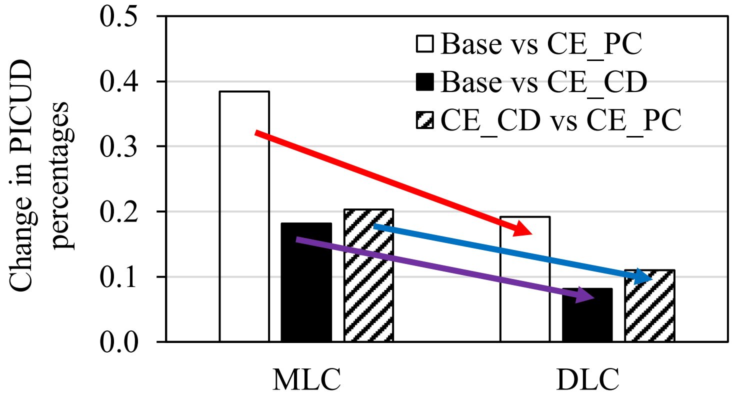


Fig. 3. Comparison of PICUD frequency percentage for different types of lane changing

4.1.2 SSCR

SSCR—sideswipe collision risk—is computed for MLC as well as DLC during the entire lane-changing duration period, and the maximum value of collision risk is selected and presented in Figure 4(a), where drivers are sorted in ascending order with respect to SSCR in the baseline condition. A series of linear mixed models reveal statistically significant difference in SSCR across driving conditions during MLC ($F_{2, 150} = 172$; p-value < 0.001) and DLC ($F_{2, 150} = 112$; p-value < 0.001). Paired t-tests also indicate that SSCRs are significantly different between a pair of drives during each lane-changing type.

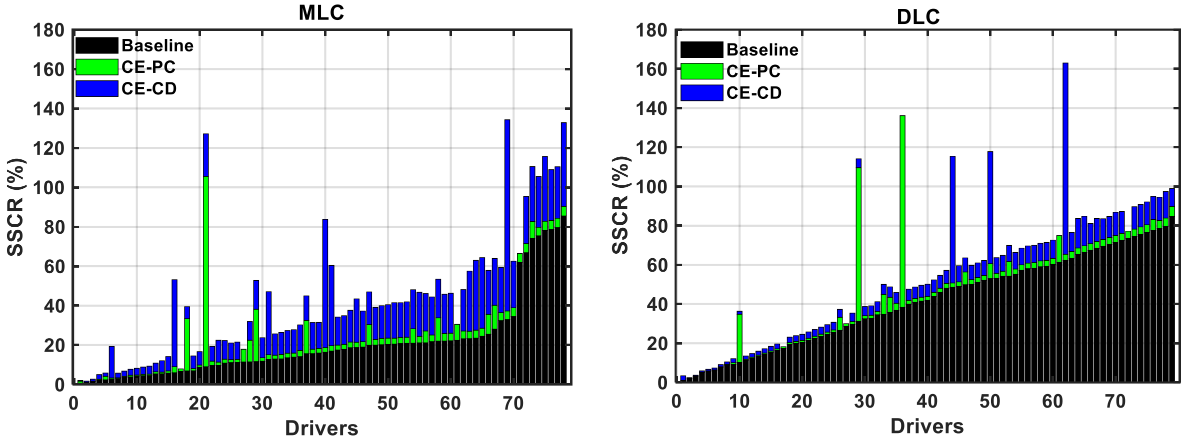
Figure 4(b) shows the difference in SSCR (i.e., SSCR in CE_PC – SSCR in the baseline condition for MLC or DLC), and it can be observed that difference in SSCR is consistently lower during MLC compared to that of DLC. A paired t-test indicates that the average differences in SSCR for MLC and DLC are respectively 36.34% and 15.64%, suggesting that CE_PC reduces sideswipe risk more during MLC compared to DLC; this change is also statistically significant as measured by a paired t-test ($t = -5.59$; p-value < 0.001).

However, contrasting results have been found when comparing the impact of communication delay on the two lane-changing types. More specifically, SSCR risk increases during MLC (compared to perfect communication) relative to DLC (Figure 4(b)). By comparing SSCR across other drivers such as baseline versus CE_PC, etc. similar results like PICUD are found.

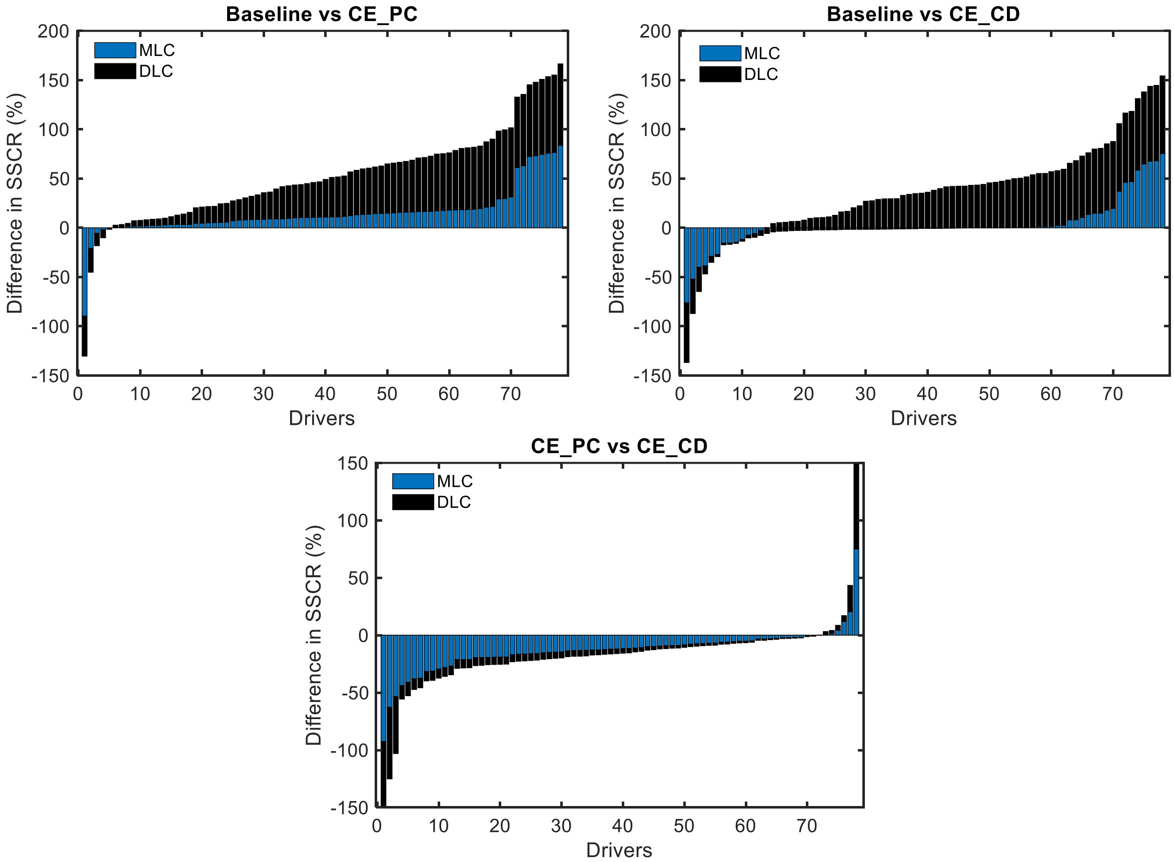
To summarise, the connected environment appears to reduce the SSCR risk during both MLC and DLC. However, the amount of reduction in SSCR is higher during MLC compared to that of DLC.

4.1.3 Speed variation

Speed variation (or fluctuation) has been frequently reported as one of the causes of traffic crashes (Pande and Abdel-Aty, 2006) and creating disturbance in a traffic stream (FHWA, 2014). The standard deviation of speed is measured for each participant during the lane-changing event for each driving condition and lane-changing type, and cumulative distribution curves are displayed in Figure 5. Using a Kolmogorov-Smirnov test for examining whether the speed variations during MLC follow the same distribution or not, we find significant differences between baseline and CE_PC ($p < 0.05$), and baseline and CE_CD ($p < 0.05$), but not between CE_PC and CE_CD ($p > 0.05$). Unlike MLC, no significant difference is found for speed variation during DLC ($p > 0.05$).



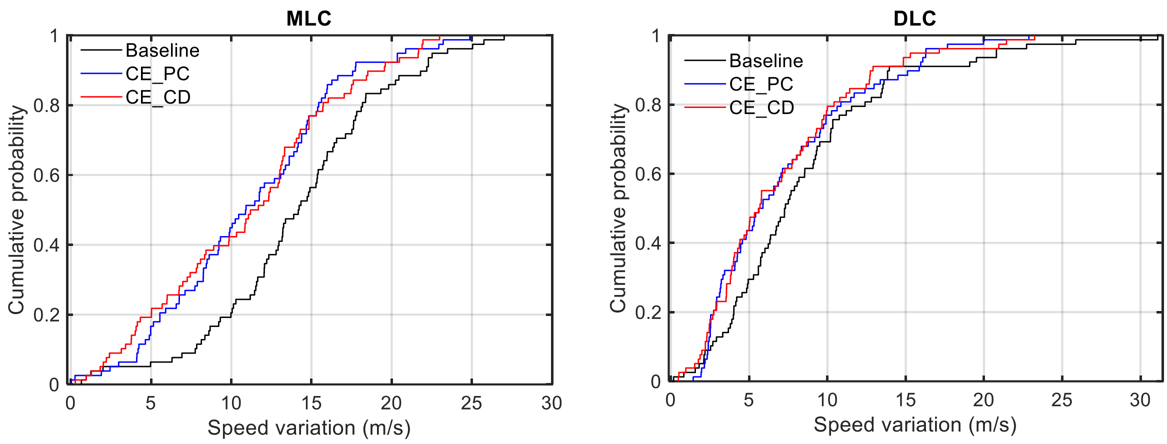
(a)



(b)

Fig. 4. Summary of SSCR results; (a) SSCR for each driving condition in each lane-changing type; (b) Comparison of SSCR between MLC and DLC

From the same cumulative probability curves (Figure 5), the speed variation during MLC for the baseline condition is clearly larger than that for the connected environment scenarios. This indicates that the connected environment driving conditions can make traffic flow smoother, which leads to lower risk. Although the same phenomenon is observed for DLC, it is to a less degree.



(a) (b)

Fig. 5. Cumulative distribution function of speed variations

Alike PICUD and SSCR, the difference in speed variation is calculated to compare the effects of the connected environment on each type of lane-changing. The average difference in speed variations for baseline and CE_PC during MLC and DLC is 3 m/s and 1.34 m/s, respectively, suggesting that speed fluctuations are decreased considerably in the perfect communication driving condition during MLC compared to that during DLC. Again, similar and consistent results, as in the case of PICUD and SSCR, are observed for comparing differences between other driving conditions.

4.1.4 Rate of deceleration

Sudden braking (in the form of hard deceleration) is reported to cause rear-end crashes (Ali et al., 2019a, Haque et al., 2016). As such, the frequency of deceleration rate exceeding a certain threshold is calculated for each participant during the lane-changing event for each driving condition and each lane-changing type, and the results are presented in Table 4. The threshold considered in this study is 0.4*g*, which has been found as a contributory factor in crashes using naturalistic data, and a detailed discussion on this can be found in Wu and Jovanis (2013).

Table 4 indicates that the difference in frequencies (expressed as a percentage of the total number of participants in each scenario) of exceeding deceleration rate above the threshold for the baseline and CE_PC during MLC and DLC is respectively 24% and 12%; this difference in frequencies is statistically significant as measured by the chi-square test (χ^2 = 4.33; *p*-value = 0.037). This suggests that CE_PC is more effective in decreasing the instances of exceeding deceleration rate during MLC compared to DLC, and similar results are found for comparisons between other driving conditions.

Table 5 also implies that the number of instances exceeding the threshold during MLC for the baseline condition are clearly higher than that for the connected environment. This reveals a positive impact of the connected environment leading to a lower propensity of engaging in safety-critical events. Although the same reduction is observed for DLC, the magnitude of reduction is much lower than that of MLC.

4.1.5 Comparing effectiveness of a connected environment between MLC and DLC

To compare usefulness (or effectiveness) of a connected environment for the two types of lane-changing, an effectiveness ratio (ER) is calculated as follows:

$$ER_{ij} = \frac{I_j^{\mathrm{MLC}} - I_i^{\mathrm{MLC}}}{I_j^{\mathrm{DLC}} - I_i^{\mathrm{DLC}}}, \qquad i, j \in \{\text{Base, CE\_PC, CE\_CD}\}, i \neq j, \tag{2}$$

Table 4. Frequency of deceleration rate exceeding the threshold

LC type	Scenario	Frequency	Comparison	Significance by a chi-square test
MLC	Baseline	45%	Baseline versus CE_PC	χ^2 = 13.75 (p-value < 0.001)
	CE_PC	21%	Baseline versus CE_CD	χ^2 = 3.82 (p-value = 0.050)
	CE_CD	36%	CE_PC versus CE_CD	χ^2 = 3.88 (p-value = 0.040)
DLC	Baseline	38%	Baseline versus CE_PC	χ^2 = 2.94 (p-value = 0.080)
	CE_PC	26%	Baseline versus CE_CD	χ^2 = 0.44 (p-value = 0.50)
	CE_CD	35%	CE_PC versus CE_CD	χ^2 = 1.11 (p-value = 0.29)

where ER_{ij} expresses the relative level of improvement in MLC versus DLC of a given surrogate safety measure *I* listed in Table 1 in (connected environment) scenario *j* compared to (connected environment or baseline) scenario *i*. If $ER_{ij} > 1$ (< 1) then scenario *j* has a higher (lower) usefulness for MLC compared to DLC than scenario *i*.

ERs for PICUD are calculated for a different combination of scenarios, and the results are displayed in Figure 6. An ER of 2.11 suggests that the CE_PC driving condition (relative to baseline condition) is more effective in lowering PICUD frequency during MLC compared to DLC. Results also reveal that the differential impact of communication impairment is also higher during MLC compared to DLC (Figure 6).

Alike PICUD, ERs are also calculated for SSCR for different possible pairs of scenarios, and the results are depicted in Figure 6. It can be observed that sideswipe risk (denoted by SSCR) during MLC is reduced in the CE_PC driving condition (relative to the baseline condition) compared to DLC. An impairment in the communication is likely to impact MLC more relative to DLC (see Figure 6).

ERs for differences in speed variation are also presented in Figure 6. Results reveal that speed variations during MLC when using driving aids (compared to no driving aids) are significantly reduced relative to DLC. When driving aids are delayed, its impact is more pronounced during MLC compared to DLC.

ER values for the rate of deceleration are respectively 2, 3, and 1.67 for baseline vs. CE_PC, baseline vs. CE_CD, and CE_CD vs. CE_PC, implying that the effectiveness of connected environment is higher (greater than one) during MLC compared to DLC in reducing the frequency of deceleration rate exceeding the threshold. The findings also suggest that the safety benefit of communication delay is not as great as perfect communication during MLC compared to DLC, which complements our previous finding.

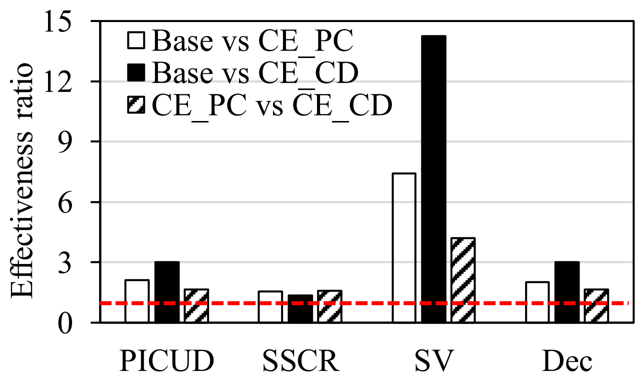


Fig. 6. Effectiveness ratio for comparing the effectiveness of a connected environment between MLC and DLC; *red dotted line shows ER = 1; Base = Baseline*; *SV = speed variation; Dec = deceleration rate*

4.2 Modelling results

4.2.1 GEE model for speed variations

The GEE model for speed variations is estimated using the '*statsmodel'* library in Python (Seabold and Perktold, 2010) and reveals potential factors that affect speed variations. The developed model is a function of main effects variables including indicator variables for the connected environment with perfect communication and communication delay, an indicator for MLC, indicators for age group, and gender, and interaction effects including MLC during the connected environment with perfect communication, spacings and lag gaps during MLC in the connected environment with perfect communication. Table 5 presents the model estimates. Note that the model presented in Table 5 is the parsimonious model selected from a series of models based on goodness-of-fit measures reported in the literature such as marginal R_2, QIC, and Quasi-likelihood (Zheng, 2000, Pan, 2001).

The indicator for driving condition (i.e., connected environment with perfect communication, CE_PC) is significant at a 95% confidence level and found to be negatively associated with speed variations, suggesting that the speed variations in perfect communication driving condition are lower with approximately 2.33 m/s less speed variation compared to when driving without driving aids. Similarly, when the communication is impaired (i.e., with communication delay), the speed variations decrease by about 1.55 m/s compared to the baseline condition (i.e., no driving aids). Although speed variations reduce in both driving conditions of the connected environment, implying a more stable and smoother driving, the speed variations in the perfect communication scenario are about 1.5 times lower than those with communication delay, suggesting the detrimental impact of communication delay compared to the connected environment with perfect communication.

To compare the differential effect of lane-changing on speed variations, the indicator for mandatory lane-changing (i.e., MLC) is found to be positive and significant. Compared to DLC, the speed variations are approximately 2.37 m/s higher during MLC. This suggests that drivers fluctuate their speeds more in the acceleration lane to look for available gaps compared to DLC whereby drivers reduce their speeds for achieving better driving conditions.

Table 5. Summary statistics of the GEE model for speed variations

Parameter	*Coefficient*	*s.e*	*Wald statistics*	*p-value*
Constant	9.61	0.79	147.81	<0.001
Main effects				
CE_PC	-2.33	0.46	22.12	<0.001
CE_CD	-1.55	0.63	6.09	0.013
MLC	2.37	0.77	19.12	<0.001
Young drivers	1.33	0.65	4.41	0.039
Older drivers	-1.75	0.74	5.55	0.018
Male	1.03	0.51	4.13	0.042
Interaction effects				
MLC × CE_PC	-4.51	0.89	25.45	<0.001
MLC × CE_PC × Spacing	-0.035	0.009	12.69	<0.001
MLC × CE_PC × Lag gaps	-0.028	0.012	4.83	0.028
Estimated correlation parameter (alpha)	26.5	2.2		
Marginal R_2 = 0.31; QIC = 1147; Quasi-likelihood = -572; Number of observations = 458; Number of clusters = 156; Max: cluster size = 3				

Baseline is the reference category

Driver age is classified into three classes, as mentioned in Table 2. Compared to middle-aged drivers, young drivers are associated with higher speed variations, with about 1.3 m/s higher speed variation than that of middle-aged drivers. Furthermore, the model reveals a more stable speed variation for older drivers with a reduction in speed variations of approximately 1.75 m/s compared to that of middle-aged drivers.

The variable for male drivers is found to be significant and has a positive impact on the speed variations. More specifically, male drivers' speed variations are about 1 m/s higher than that of female drivers.

The interaction effect for the connected environment with perfect communication and MLC has a significant impact on speed variations. The model suggests that the speed variations are reduced when MLC is performed with a fully functioning connected environment (i.e., CE_PC) with a lower speed variation of approximately 4.5 m/s.

The interaction effects for spacing and lag gaps (separately) during MLC in the connected environment with perfect communication is significant and negatively associated with speed variation in the GEE model. These interaction effects imply that when spacing and lag gaps increase and drivers are assisted with surrounding traffic information during MLC, the speed variations decrease by about 0.4 m/s and 0.3 m/s, respectively.

4.2.2 Logistic GEE model for exceeding decelerating rate

(a) Decision tree

Recall that it is important to determine higher-order interaction effects for the model development, as mentioned in Section 4.2. To this end, a decision tree is constructed using the Chi-Squared Automatic Interaction Detection (CHAID) algorithm by using the '*CHAID'* library of Python (Ramotowski and Fitzgerald, 2020). As the name suggests, this tree is constructed from various possible combinations and divisions using a Chi-square test (Choudhary and Velaga, 2019). The dependent variable is a binary variable (i.e., a driver exceeding the deceleration threshold or not), whereas the input variables are driving conditions, type of lane-changing, and driver demographics, as shown in Table 2. A k-fold cross-validation is conducted to construct the tree, where k is considered as 10. This process classifies the data into 10 unique portions, and each one of them is used to assess the tree structure. As such, nine-tenths of the data on each cycle is used to train the tree. The developed tree correctly classified 75% of cases using 30 leaves for a total tree of the size of 57 nodes. Driving conditions reveal the highest information gain and thus is located at the top of the tree (Figure 7).

Each terminal node, indicated by a number in square brackets after the node in Figure 6, represents a possible interaction term. Table 6 presents all potential interaction terms obtained from the decision tree. Note that each classification is tested using a chi-square value and the corresponding p-value less than 0.05.

The decision tree divides the deceleration rates above/below the threshold by classifying the data into 30 smaller and homogeneous groups, and the corresponding statistics are presented below in the parenthesis (see Figure 7). The two numbers in the parenthesis represent how many cases reach the node with deceleration rate above and below the threshold, respectively. For instance, the statistics of terminal node 1 indicate that 100% of drivers (i.e., 2 out of 2) in the baseline condition with lag gaps less than or equal to 24.6 m have a higher deceleration rate. Similarly, terminal node 12 suggests that about 19% of middle-aged male drivers (i.e., 10 out of 52) in the CE_PC driving condition have higher deceleration rates while the rest have lower deceleration rates. Terminal node 16 implies that about 14% of older drivers (i.e., 1 out of 7) in the CE_PC driving condition performing DLC with spacing < 27.5 m have higher deceleration rates.

Table 6. Interaction effects obtained from the decision tree

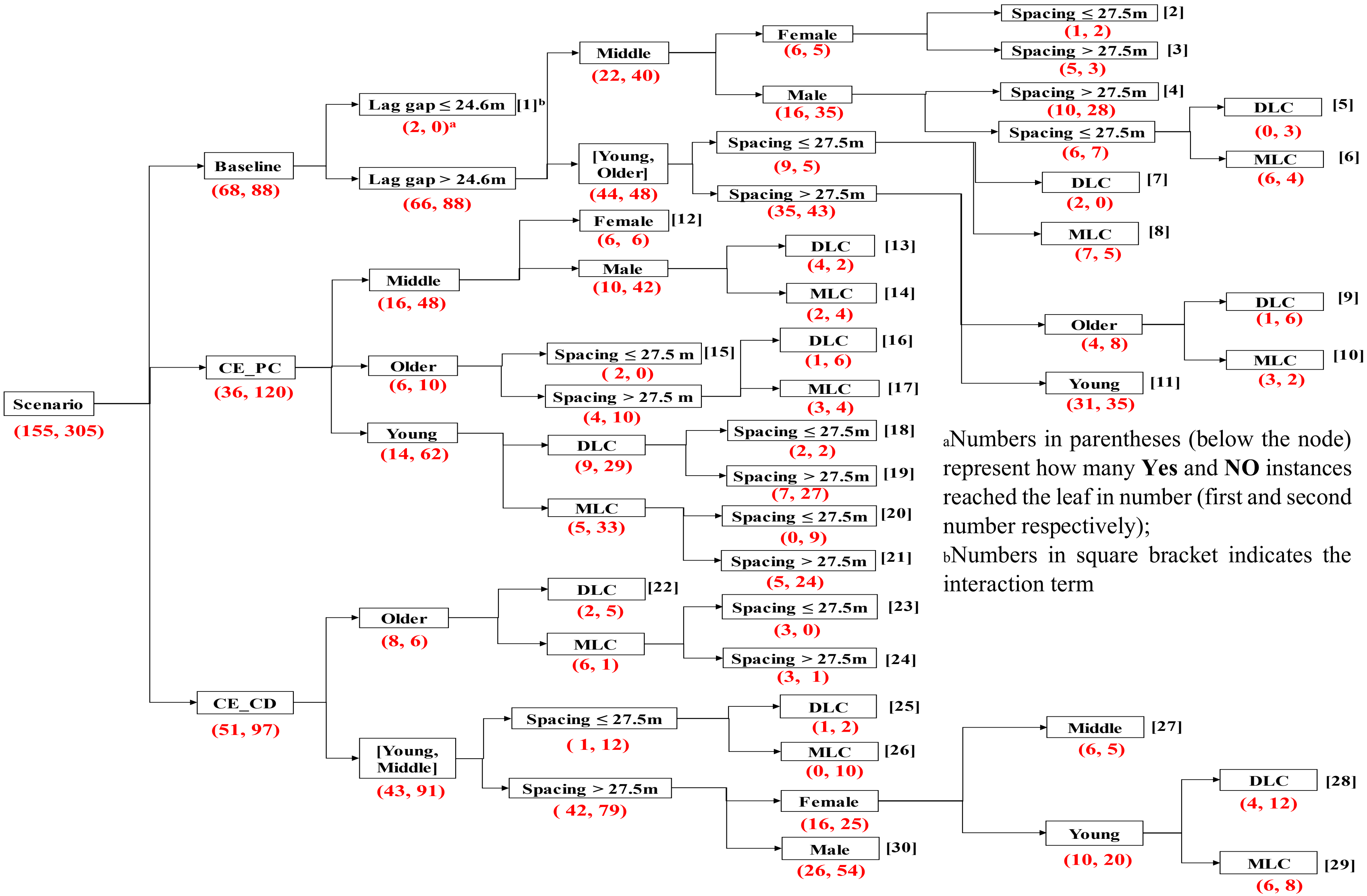
No	Description
1	Drivers in baseline condition with lag gaps ≤ 24.6 m
2	Middle-aged female drivers in baseline condition with lag gaps > 24.6m and spacing ≤ 27.5 m
3	Middle-aged female drivers in baseline condition with lag gaps >24.6m and spacing > 27.5 m
4	Middle-aged male drivers in baseline condition with lag gaps > 24.6m and spacing > 27.5 m
5	Middle-aged male drivers in baseline condition performing DLC with lag gaps > 24.6m and spacing ≤ 27.5 m
6	Middle-aged male drivers in baseline condition performing MLC with lag gaps > 24.6m and spacing ≤ 27.5 m
7	Young (or older) drivers in baseline condition performing DLC with lag gaps > 24.6m and spacing ≤ 27.5 m
8	Young (or older) drivers in baseline condition performing MLC with lag gaps > 24.6m and spacing ≤ 27.5 m
9	Older drivers in baseline condition performing DLC with lag gaps > 24.6m and spacing > 27.5 m
10	Older drivers in baseline condition performing MLC with lag gaps > 24.6m and spacing > 27.5 m
11	Young drivers in baseline condition with lag gaps > 24.6m and spacing > 27.5 m
12	Middle-aged female drivers in CE_PC
13	Middle-aged male drivers in CE_PC performing DLC
14	Middle-aged male drivers in CE_PC performing MLC
15	Older drivers in CE_PC with spacing ≤ 27.5 m
16	Older drivers in CE_PC performing DLC with spacing > 27.5 m
17	Older drivers in CE_PC performing MLC with spacing > 27.5 m
18	Young drivers in CE_PC performing DLC with spacing ≤ 27.5 m
19	Young drivers in CE_PC performing DLC with spacing > 27.5 m
20	Young drivers in CE_PC performing MLC with spacing ≤ 27.5 m
21	Young drivers in CE_PC performing MLC with spacing > 27.5 m
22	Older drivers in CE_CD performing DLC
23	Older drivers in CE_CD performing MLC with spacing ≤ 27.5 m
24	Older drivers in CE_CD performing MLC with spacing > 27.5 m
25	Young (or older) drivers in CE_CD performing DLC with spacing ≤ 27.5 m
26	Young (or older) drivers in CE_CD performing MLC with spacing ≤ 27.5 m
27	Middle-aged female drives in CE_CD with spacing > 27.5 m
28	Young female drivers in CE_CD with performing DLC with spacing > 27.5 m
29	Young female drivers in CE_CD with performing MLC with spacing > 27.5 m
30	Young (or older) male drivers in CE_CD with spacing > 27.5 m

(b) Model interpretation

The significant variables estimated by the repeated measure GEE logistic model along with the probabilities of drivers' engaging in a hard-braking event during the lane-changing event are presented in Table 7. The parsimonious model (presented in Table 7) is compared to a model without interaction terms. The Wald χ^2 statistics for the model with and without interaction variables are respectively 66.4 and 43.2, implying that both models possess a reasonable explanatory power. However, goodness-of-fit measures (e.g., AIC, Wald χ^2, QIC, and Quasi-likelihood) suggest that the model with interaction terms outperforms the model without interaction terms. Thus, the model with interaction terms is selected in this study.

The selected model has an exchangeable correlation coefficient, ρ, of 0.38, suggesting that there exists a significant correlation among repeated observations of each driver, which is accounted for in our repeated measures GEE logistic model.

The parsimonious model contains eight main effects parameters: spacing, lag gaps, dummy variables for CE_PC, CE_CD, MLC, young and older drivers, male, and four higher-order interaction effects obtained from the decision tree: *interaction terms* 13, 15, 18, and 28. Note that these variables are explained in Table 6. The odds ratio as mentioned in Section 4.2 is calculated for each parameter estimate that suggests the influence of a particular variable on the odds of engaging in a hard-braking event while keeping all other effects constant in the model. It is important to mention here that the probability of engaging in a hard-braking event is also a function of other factors, which are explained in the next section.



aNumbers in parentheses (below the node) represent how many **Yes** and **NO** instances reached the leaf in number (first and second number respectively);
bNumbers in square bracket indicates the interaction term

Fig. 7. Decision tree schematic for the deceleration exceeding the threshold

Table 7. Repeated measure logistic GEE model predicting the likelihood of engaging in a hard-braking event

Parameter	*Coefficient*	*s.e*	*Wald statistics*	*p-value*	*OR*	*95% CI of OR*	
						Lower	*Upper*
Constant	-1.211	0.314	14.83	<0.001			
Main effects							
Perfect communication	-1.690	0.227	37.17	<0.001	0.185	0.260	0.629
Communication delay	-0.506	0.252	4.02	0.044	0.603	0.109	1.097
DLC	0.566	0.258	4.79	0.008	1.761	1.256	2.267
Young drivers	0.647	0.272	5.59	0.018	1.911	1.378	2.444
Older drivers	-0.789	0.373	4.47	0.034	0.454	0.277	1.185
Male	0.476	0.234	4.11	0.042	1.610	1.151	2.068
Spacing	-0.014	0.003	3.91	0.048	0.986	0.980	0.992
Lag gap	-0.012	0.004	5.45	0.019	0.988	0.980	0.996
Interaction effects							
Interaction 13	-2.180	0.649	11.26	<0.001	0.113	1.159	1.385
Interaction 15	-1.661	0.741	5.03	0.024	0.189	1.263	1.641
Interaction 18	-1.068	0.506	4.45	0.034	0.343	0.648	1.334
Interaction 28	-1.733	0.810	4.57	0.032	0.177	1.411	1.764
Wald Chi-square = 66.4; DF = 12; *p*-value < 0.001; QIC = 529; Quasi-likelihood = -253; Number of observations = 458; Number of clusters = 156; Max: cluster size = 3; Exchangeable correlation (ρ) = 0.38							

Interaction term 13: Middle-aged male drivers in CE_PC performing DLC
Interaction term 15: Older drivers in CE_PC with spacing ≤ 27.5 m
Interaction term 18: Young drivers in CE_PC performing DLC with spacing ≤ 27.5 m
Interaction term 28: Young female drivers in CE_CD with performing DLC with spacing > 27.5 m

The dummy variable for CE_PC is significant at a 95% confidence level in explaining the possibility of drivers engaging in a hard-braking event during a lane-changing manoeuvre. The model suggests that drivers are less likely to engage in a hard-braking event when they are assisted by continuous and on-time driving aids. More specifically, the probability of engaging in a hard-braking event decreased by about 81.5% (i.e., $\exp(-1.69) = 0.185$).

The dummy variable for CE_CD is also significant and negatively associated with the likelihood of drivers engaging in a hard-braking event during the lane-changing event. Compared to no driving aids, drivers are less likely to engage in a hard-braking event when driving aids are delayed, as the probability is decreased by 39.7. This suggests that the risk is reduced, but at a lower magnitude, when the delayed information is offered to drivers.

The indicator for MLC suggests that overall, the odds of engaging in a hard-braking event is increased by 76.1% when performing an MLC manoeuvre compared to a DLC manoeuvre. It is worth mentioning here that this odds ratio estimate does not include interactions related to the connected environment driving conditions such as *interaction term* 13 and *interaction term* 18. As we have incorporated several interaction effects in this study, the full spectrum of connectivity's impact on the type of lane-changing (i.e., MLC or DLC) would be identifiable along with driver demographics.

The GEE model suggests that young drivers are more likely to engage in a hard-braking event than middle-aged drivers, namely 1.91 times according to the odds ratio. On the other hand, older drivers have a lower propensity of engaging in a hard-braking event compared to middle-aged drivers with an odds ratio of 0.454. We elaborate these findings further in the next section. Furthermore, the model shows that male drivers are 1.61 times more likely to engage in a hard-braking event compared to female drivers.

Looking at spacing, which is the distance between the lead vehicle and the subject vehicle in the current driving lane, we find that a one metre increase in spacing is associated with a 1.4% decrease in the possibility of a hard-braking event during lane-changing, keeping all other variables constant. With respect to lag gap, which is the distance between the subject vehicle and the immediate follower on the adjacent lane, we find that a one metre increase in lag gap decreases the probability of drivers engaging in a hard-braking event by 1.2% by keeping all other variables constant.

Apart from the main effects, the developed GEE model contains four interaction terms. *Interaction term* 13 shows that a middle-aged male driver, when performing DLC, is less likely to engage in a hard-braking event with the corresponding odds about 88% lower. Similarly, *Interaction term* 15 indicate that older drivers in the fully functioning connected environment are about 81% less likely to engage in a hard-braking event during the lane-changing event when the spacing between the lead vehicle and subject vehicle is less than 27.5 m. *Interaction terms* 18 and 28 can be interpreted in a similar manner.

5. Discussion

5.1 Speed variations and lane-changing

Speed variation is one of the factors associated with crash risk (Aarts and Van Schagen, 2006, Zheng et al., 2010, Roshandel et al., 2015) and affects traffic flow characteristics (Svenson, 2009). For instance, speed variations during lane-changing are associated with a bottleneck (Bertini and Cassidy, 2002) and traffic oscillations (Zheng et al., 2011a). To this end, driving aids provided by the connected environment can assist in a lane-changing driving task and thereby minimise speed variations. For this purpose, speed variations during different lane-changing types are analysed using the developed model. More specifically, using the model estimates presented in Table 5, and the mean values presented in Table 4, the effectiveness of the connected environment during MLC and DLC can be determined and computed as follows:

$$
\begin{aligned}
\text{mean speed variation} = \exp\big(& \ \ 9.61 \\
& -2.33\times \mathrm{CE\_PC} \\
& -1.55\times \mathrm{CE\_CD} \\
& +2.37\times \mathrm{MLC} \\
& +1.33\times \mathrm{YoungDriver} \\
& -1.75\times \mathrm{OlderDriver} \\
& +1.03\times \mathrm{MaleDriver} \\
& -4.51\times \mathrm{CE\_PC}\times \mathrm{MLC} \\
& -0.028\times \mathrm{MLC}\times \mathrm{CE\_PC}\times \mathrm{lagGap} \\
& -0.035\times \mathrm{CE\_PC}\times \mathrm{MLC}\times \mathrm{Spacing}\big).
\end{aligned}
\tag{3}
$$

Keeping all other variables constant (i.e., mean), the speed variations in the CE_PC and baseline driving conditions during MLC are respectively 3.43 m/s and 11.98 m/s, while the corresponding the speed variations for DLC are 7.28 m/s and 9.61 m/s, respectively, implying that the relative speed variation (baseline – CE_PC during MLC vs baseline – CE_PC during DLC) reduces by 3.67 times in case of MLC compared to DLC. In the literature, due to forceful nature of MLC (often performed at merging or weaving sections) and when performed without a driving assistance system, it disrupts traffic more (Ali et al., 2019b) and results in stop-and-go oscillations (Ahn and Cassidy, 2007), traffic breakdown, and capacity drops (Cassidy and Rudjanakanoknad, 2005). Nevertheless, although the connected environment appears to reduce the speed variations during both MLC and DLC, its impact is more prominent and higher during MLC. This finding suggests that the connected environment provides the highest benefit when drivers require assistance the most (during MLC). This is because a driver's workload and stress are likely to increase during an MLC manoeuvre as they continuously need to find a gap in the target lane and monitor the remaining distance in the acceleration lane simultaneously. Most drivers would want to merge before reaching the end of the acceleration lane to avoid coming to a complete halt, which leads to increased impatience and elevates risky gap selection and crash risk (Ahmed, 1999, Ali et al., 2019a). However, when they are assisted by the connected environment, their speed variations drastically decrease because drivers can select a proper gap size (as subsequent gap sizes are provided by the connected environment) and receive surrounding traffic information that can assist in minimising speed variations. On the other hand, the urgency factor associated with DLC is minimum as it depends on the driver's discretion to change lanes or persist with the prevailing conditions in the current driving lane. Nevertheless, the connected environment seems to be assisting during DLC as well as speed variations are reduced, but not as much as during MLC.

Keeping all other variables constant, the differences in the speed variations (CE_CD vs CE_PC) predicted by the model for MLC and DLC are respectively 7.01 m/s and 2.33 m/s, suggesting that delayed information supply during MLC increases speed variations threefold compared to that of DLC.

5.2 Hard-braking probabilities during lane-changing

To examine whether the connected environment has any differential impact on MLC and DLC manoeuvres, a complex model that accounts for such interactions is needed. As such, we have estimated a repeated measure logistic GEE model for determining the possibility of drivers engaging in a hard-braking event, which has a significant impact on traffic flow efficiency and safety as indicated in several studies (Simons-Morton et al., 2009, Van Driel and Van Arem, 2010, Wu et al., 2010, Haque and Washington, 2015). The developed model can provide probabilities of engaging in a hard-braking event for each lane-changing type, gender, age group, and varying spacing and lag gap values. The probabilities can be calculated using the parameter estimates reported in Table 7 together with mean values of the explanatory variables (except for lane-changing type and driving condition). Using Equation 1, the predicted probabilities for drivers' engaging in a hard-braking event during MLC and DLC in the baseline (without driving aids) and CE_PC condition can be computed as

$$P_{\text{Base,MLC}} = \frac{\exp[-1.211 + 0.566 \times 1 - 0.014 \times 24.64 - 0.012 \times 30.45]}{1 + \exp[-1.211 + 0.566 \times 1 - 0.014 \times 24.64 - 0.012 \times 30.45]} = 0.20, \tag{4}$$

$$P_{\text{Base,DLC}} = \frac{\exp[-1.211 + 0.566 \times 0 - 0.014 \times 24.64 - 0.012 \times 30.45]}{1 + \exp[-1.211 + 0.566 \times 0 - 0.014 \times 24.64 - 0.012 \times 30.45]} = 0.13, \tag{5}$$

$$P_{\text{CE\_PC,MLC}} = \frac{\exp[-1.211 - 1.69 \times 1 + 0.566 \times 1 - 0.014 \times 24.64 - 0.012 \times 30.45]}{1 + \exp[-1.211 - 1.69 \times 1 + 0.566 \times 1 - 0.014 \times 24.64 - 0.012 \times 30.45]} = 0.04, \tag{6}$$

$$P_{\text{CE\_PC,DLC}} = \frac{\exp[-1.211 - 1.69 \times 1 + 0.566 \times 0 - 0.014 \times 24.64 - 0.012 \times 30.45]}{1 + \exp[-1.211 - 1.69 \times 1 + 0.566 \times 0 - 0.014 \times 24.64 - 0.012 \times 30.45]} = 0.02. \tag{7}$$

The probabilities of engaging in a hard-braking event during MLC and DLC when drivers are not assisted by driving aids are respectively 20% and 13% (Eqs. 4 and 5) while the corresponding probabilities in the CE_PC driving condition are 4% and 2% (Equations 6 and 7), respectively. This suggests that, although probabilities decrease for both types of lane-changing, a 5% higher decrease is observed during MLC in the CE_PC driving condition, further highlighting the superior benefits of driving aids during MLC.

To examine the effects of the connected environment driving conditions on MLC and DLC by varying spacing and lag gaps along with driver demographic, probabilities are calculated and presented in Figure 8. Clearly, the probabilities of engaging in a hard-braking event decrease with an increase in spacing for both types of lane-changing and driving conditions (Figure 8(a)). This finding is intuitive and can be explained by the fact that when the distance between the lane-changer and its leader is large, there is a minimal chance for drivers to exhibit higher deceleration and subsequently engage in a hard-braking event. Notably, a higher decrease in the probability is observed during MLC compared to its counterpart. A similar trend but with different magnitude of impact has been observed when the lag gap is varied (see Figure 8(b)).

A differential comparison of the connected environment driving conditions reveals that drivers in CE_PC have the lowest probability of engaging in hard-braking events compared to CE_CD. However, the communication delay is still performing better than the baseline condition. This also confirms that the safety benefits of CE_CD are not as much as CE_PC.

Figure 9 displays how the probabilities of engaging in a hard-braking event vary with driver age and gender. It is evident that probabilities of engaging in a hard-braking event are higher for young drivers compared to that of middle-aged and older drivers. For instance, the probabilities of engaging in a hard-braking event for young, middle-aged, and older male drivers performing DLC at a 5 m spacing in the CE_PC (baseline) are respectively 16% (51%), 1% (12%), and 4% (19%) while the corresponding probabilities for MLC are 9% (37%), 0.6% (3%), and 2% (12%), respectively. Two noteworthy observations are: (a) young drivers, although their probabilities decrease in the connected environment compared to the baseline condition, tend to decelerate more sharply and thus are more likely to be involved in safety-critical events compared to other groups of drivers. This finding aligns with Andersen et al. (2000) who reported age-related differences in deceleration rates and found that young drivers have a higher propensity of engaging in a collision due to their novice driving skills compared to other drivers; (b) the benefit of the connected environment is higher for DLC compared to its counterpart in the same driving condition. This contrasting finding may be explained by the fact that drivers during DLC appear to take the maximum advantage of advanced information about congestion and adjust their speeds accordingly and thereby decelerate more gradually and exhibit stable driving. These findings also advocate further in-depth analysis of the microscopic (or individual level) assessment of the usefulness of the connected environment.

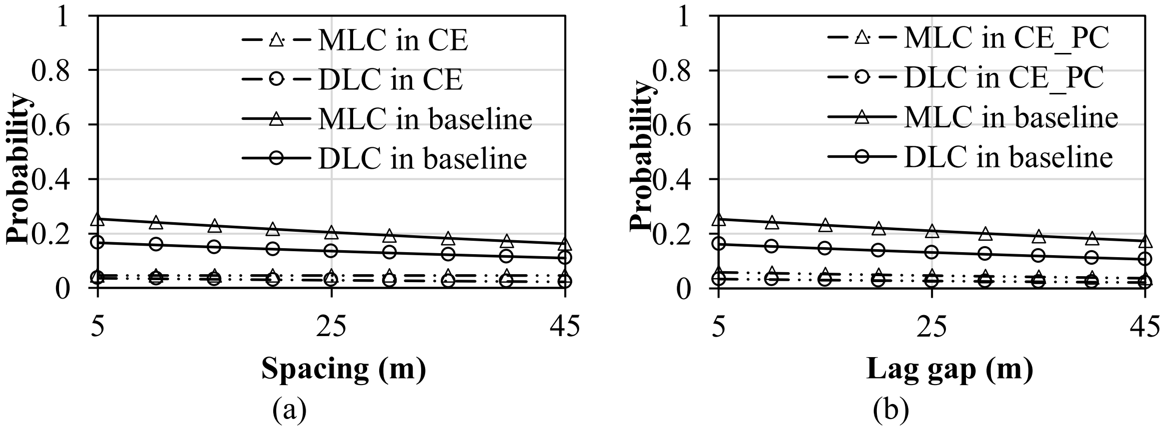


Fig. 8. Effects of spacing and lag gap on the probability of engaging in a hard-braking event during the connected environment with perfect communication

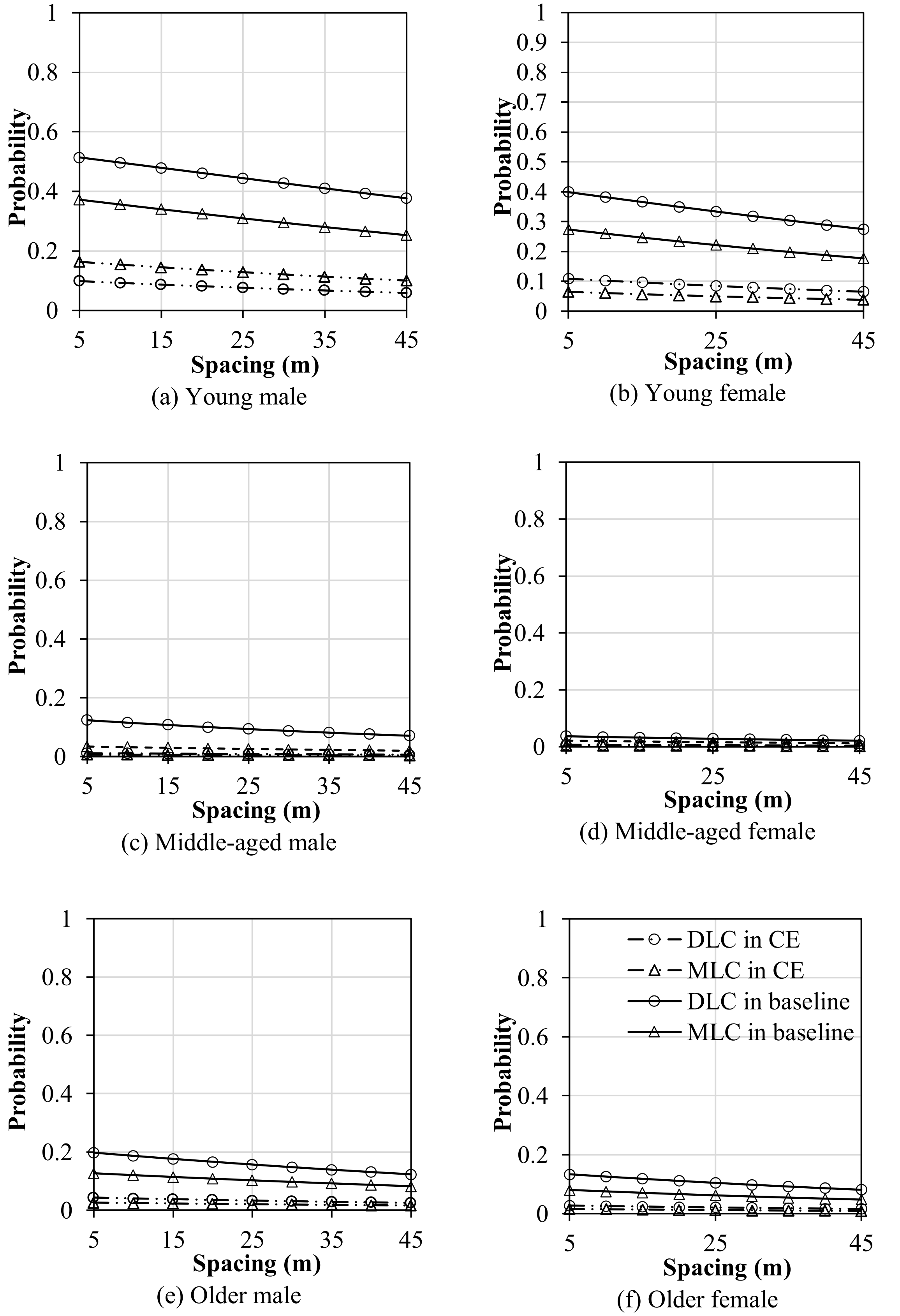


Fig. 9. Probability of engaging in a hard-braking event for different driver demographics

Compared to male drivers, female drivers have a lower probability of engaging in a hard-braking event (Figure 9). On average, a 10% decrease in the probability of engaging in a hard-braking event is observed for female drivers compared to male drivers. Similar findings are reported by some other studies (Li et al., 2015, Li et al., 2016), where female drivers' deceleration rates were lower than male drivers as these (male) drivers tend to be more aggressive and exhibit risky driving behaviour.

6. Conclusions and future research

This study examined the usefulness of the connected environment on two types of lane-changing, namely mandatory lane-changing (MLC) and discretionary lane-changing (DLC). Because of the distinct nature of these two lane-changing types and different mechanisms involved in performing these lane-changings, it is expected that there could be a differential impact of the connected environment. To address this research gap, this study performed descriptive analyses as well as developed Generalised Estimation Equations (GEE) models to examine the differential benefits of the connected environment. Seventy-eight participants performed MLC and DLC on a simulated motorway in the CARRS-Q Advanced Driving Simulator, where they received surrounding traffic information in a fully functioning as well as impaired communication systems.

Two surrogate measures of safety were adopted for evaluating the possibility of rear-end collisions and sideswipes during a lane-changing manoeuvre. Descriptive analyses revealed that the safety margin in the connected environment driving conditions during MLC is higher compared to its counterpart. Similarly, the reduction of speed variations in MLC, thanks to the connected environment, is larger than that in DLC. Moreover, impairment in driving aids is likely to reduce safety margin more during MLC compared to DLC.

To develop further insights on speed variations during lane-changing, a GEE model is developed that suggests that although speed variations were higher during MLC when driving aids are not available, the amount of reduction was considerably lower considerably when drivers performed MLC in the presence of driving aids. Furthermore, the degree of reduction in speed variation was higher during MLC than that of DLC.

To study whether the connected environment lowers the probability of engaging in a hard-braking event for the two lane-changing types, a hybrid framework of data mining and classical statistical modelling has been employed. Using a decision tree analysis, all possible interaction effects were obtained, which were used as input into a repeated measure logistic GEE model. The developed model reveals that although the possibility of engaging in a hard-braking event was much higher during MLC in a traditional driving environment (without driving aids) compared to its counterpart, these events drastically decreased when drivers were assisted with driving aids to perform MLC. The model also suggests that young and male drivers have a higher probability of engaging in hard-braking events. While examining the interaction effect of gender with age group on lane-changing type, we find that the connected environment was more effective during DLC compared to its counterpart, which needs further investigation.

Overall, this study concludes that where the urgency of lane-changing is higher, the connected environment is expected to benefit more. Because these situations (e.g., merging, weaving, etc.) are prone to high crash risk and drivers, in general, tend to avoid these crash risks, they take the maximum advantage of the available information. On the contrary, drivers during DLC, although utilising driving aids, have lower safety benefits compared to MLC because of less urgency of lane-changing. This aligns with a recent study where the authors found that the usefulness of information is higher when the headway is small, which is an indicator of crash risk (Sharma et al., 2019).

This study provides a better understanding of the impact of the connected environment on the two lane-changing types and uncovers that although the connected environment provides similar surrounding traffic information in both these situations, their impact is different and thus warrants separate model formulation for each lane-changing type in a connected environment. Furthermore, our findings can also assist in improving the design of driving aids related to DLC to increase its effectiveness in the decision-making process.

This study was conducted in a simulated environment using a single high-fidelity driving simulator, and lane-changing behaviour was compared in a relative manner. In this study, we programmed all surrounding vehicles in such a way as to mimic real driving conditions. However, it would be interesting to use multiple connected driving simulators with human drivers (as lane-changer as well as an immediate follower) or replicate the same study in field driving conditions where drivers' decisions are more volatile and can provide more insights into the differential impact of a connected environment. Lastly, this study tested only one design of driving aids in a connected environment; different designs may lead to different conclusions, and it would be worthwhile to explore the impact of different types of driving aids.

References

Aarts, L. & Van Schagen, I. 2006. Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention,* 38**,** 215-224.

Ahmed, K., Ben-Akiva, M., Koutsopoulos, H. & Mishalani, R. 1996. Models of freeway lane changing and gap acceptance behavior. *Transportation and traffic theory,* 13**,** 501-515.

Ahmed, K. I. 1999. *Modeling drivers' acceleration and lane changing behavior.* Massachusetts Institute of Technology.

Ahn, S. & Cassidy, M. J. Freeway traffic oscillations and vehicle lane-change maneuvers. Proceedings of the 17th International Symposium on Transportation and Traffic Theory, Amsterdam, pp. 691–710. , 2007.

Ali, Y., Haque, M. M., Zheng, Z., Washington, S. & Yildirimoglu, M. 2019a. A hazard-based duration model to quantify the impact of connected driving environment on safety during mandatory lane-changing. *Transportation research part C: emerging technologies,* 106**,** 113-131.

Ali, Y., Sharma, A., Haque, M. M., Zheng, Z. & Saifuzzaman, M. 2020a. The impact of the connected environment on driving behavior and safety: A driving simulator study. *Submitted*.

Ali, Y., Zheng, Z. & Haque, M. M. 2018. Connectivity's impact on mandatory lane-changing behaviour: evidences from a driving simulator study. *Transportation Research Part C: Emerging Technologies,* 93**,** 292-309.

Ali, Y., Zheng, Z., Haque, M. M. & Wang, M. 2019b. A game theory-based approach for modelling mandatory lane-changing behaviour in a connected environment. *Transportation Research Part C: Emerging Technologies,* 106**,** 220-242.

Ali, Y., Zheng, Z., Haque, M. M., Yildirimoglu, M. & Washington, S. 2020b. CLACD: a Complete LAne-Changing Decision model for traditional and connected environments. *Submitted*.

Ali, Y., Zheng, Z., Haque, M. M., Yildirimoglu, M. & Washington, S. 2020c. Understanding the discretionary lane-changing behaviour in the connected environment. *Accident Analysis & Prevention,* 137**,** 105463.

Ali, Y., Zheng, Z., Haque, M. M., Yildirimoglu, M. & Washington, S. 2020d. Understanding the discretionary lane-changing behaviour in the connected environment. *Accident Analysis and Prevention (2nd round of review)*.

Andersen, G. J., Cisneros, J., Saidpour, A. & Atchley, P. 2000. Age-related differences in collision detection during deceleration. *Psychology and aging,* 15**,** 241.

Arai, K. & Sentinuwo, S. R. 2012. Spontaneous-braking and lane-changing effect on traffic congestion using cellular automata model applied to the two-lane traffic. *IJACSA) International Journal of Advanced Computer Science and Applications,* 3.

Balal, E., Cheu, R. L., Gyan-Sarkodie, T. & Miramontes, J. 2014. Analysis of discretionary lane changing parameters on freeways. *International Journal of Transportation Science and Technology,* 3**,** 277-296.

Balal, E., Cheu, R. L. & Sarkodie-Gyan, T. 2016. A binary decision model for discretionary lane changing move based on fuzzy inference system. *Transportation Research Part C: Emerging Technologies,* 67**,** 47-61.

Behbahani, H. & Nadimi, N. 2015. A framework for applying surrogate safety measures for sideswipe conflicts. *International Journal for Traffic & Transport Engineering,* 5**,** 371-383.

Bertini, R. L. & Cassidy, M. J. 2002. Some observed queue discharge features at a freeway bottleneck downstream of a merge. *Transportation Research Part A: Policy and Practice,* 36**,** 683-697.

Bham, G. & Goswami, V. 2007. Gap acceptance behavior in mandatory lane changes under congested and uncongested traffic on a multi lane freeway.

Bham, G. H. Estimating driver mandatory lane change behavior on a multi lane freeway. Proc. 88th Annu. Meet. Transp. Res. Board, 2009. 1-22.

Boonsiripant, S. 2009. *Speed profile variation as a surrogate measure of road safety based on GPS-equipped vehicle data.* Georgia Institute of Technology.

Cassidy, M. J. & Rudjanakanoknad, J. 2005. Increasing the capacity of an isolated merge by metering its on-ramp. *Transportation Research Part B: Methodological,* 39**,** 896-913.

Choudhary, P. & Velaga, N. R. 2019. Driver behaviour at the onset of yellow signal: a comparative study of distraction caused by use of a phone and a music player. *Transportation research part F: traffic psychology and behaviour,* 62**,** 135-148.

Choudhury, C., Akiva, M. B., Toledo, T., Rao, A. & Lee, G. 2006. Cooperative lane changing and forced merging model: NGSIM final report. *Federal Highway Administration*.

Choudhury, C., Ramanujam, V. & Ben-Akiva, M. 2009. Modeling acceleration decisions for freeway merges. *Transportation Research Record: Journal of the Transportation Research Board***,** 45-57.

Choudhury, C. F., Ben-Akiva, M. E., Toledo, T., Lee, G. & Rao, A. Modeling cooperative lane changing and forced merging behavior. 86th Annual Meeting of the Transportation Research Board, Washington, DC, 2007.

Fhwa 2014. Speed harmonization. *FHWA-HRT-15-012, HRDO-01/11-14(750)E***,** Office of Operations Research and Development, U.S Department of Transportation, Federal Highway Authority.

Gandhi, U. D., Singh, A., Mukherjee, A. & Chandak, A. Smart vehicle connectivity for safety applications. 2014 International Conference on Reliability Optimization and Information Technology (ICROIT), 2014. IEEE, 262-266.

Guériau, M., Billot, R., El Faouzi, N.-E., Monteil, J., Armetta, F. & Hassas, S. 2016. How to assess the benefits of connected vehicles? A simulation framework for the design of cooperative traffic management strategies. *Transportation research part C: emerging technologies,* 67**,** 266-279.

Guler, S. I., Menendez, M. & Meier, L. 2014. Using connected vehicle technology to improve the efficiency of intersections. *Transportation Research Part C: Emerging Technologies,* 46**,** 121-131.

Haque, M. M., Ohlhauser, A. D., Washington, S. & Boyle, L. N. 2016. Decisions and actions of distracted drivers at the onset of yellow lights. *Accident Analysis & Prevention,* 96**,** 290-299.

Haque, M. M. & Washington, S. 2015. The impact of mobile phone distraction on the braking behaviour of young drivers: a hazard-based duration model. *Transportation research part C: emerging technologies,* 50**,** 13-27.

Hess, S., Choudhury, C. F., Bliemer, M. C. & Hibberd, D. 2020. Modelling lane changing behaviour in approaches to roadworks: contrasting and combining driving simulator data with stated choice data. *Transportation research part C: emerging technologies,* 112**,** 282-294.

Hidas, P. 2002. Modelling lane changing and merging in microscopic traffic simulation. *Transportation Research Part C: Emerging Technologies,* 10**,** 351-371.

Hidas, P. 2005. Modelling vehicle interactions in microscopic simulation of merging and weaving. *Transportation Research Part C: Emerging Technologies,* 13**,** 37-62.

Li, X., Yan, X. & Wong, S. 2015. Effects of fog, driver experience and gender on driving behavior on S-curved road segments. *Accident Analysis & Prevention,* 77**,** 91-104.

Li, X., Yan, X., Wu, J., Radwan, E. & Zhang, Y. 2016. A rear-end collision risk assessment model based on drivers' collision avoidance process under influences of cell phone use and gender—A driving simulator based study. *Accident Analysis & Prevention,* 97**,** 1-18.

Liang, K.-Y. & Zeger, S. L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika,* 73**,** 13-22.

Marczak, F., Daamen, W. & Buisson, C. 2013. Merging behaviour: Empirical comparison between two sites and new theory development. *Transportation Research Part C: Emerging Technologies,* 36**,** 530-546.

Pan, W. 2001. Akaike's information criterion in generalized estimating equations. *Biometrics,* 57**,** 120-125.

Pande, A. & Abdel-Aty, M. 2006. Assessment of freeway traffic parameters leading to lane-change related collisions. *Accident Analysis & Prevention,* 38**,** 936-948.

Ramotowski, M. & Fitzgerald, R. 2020. Chi-Squared Automatic Inference Detection (CHAID) decision tree. *Apache Software License,* Version 5.3.0.

Rios-Torres, J. & Malikopoulos, A. A. 2017. A survey on the coordination of connected and automated vehicles at intersections and merging at highway on-ramps. *IEEE Transactions on Intelligent Transportation Systems,* 18**,** 1066-1077.

Roshandel, S., Zheng, Z. & Washington, S. 2015. Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. *Accident Analysis & Prevention,* 79**,** 198-211.

Sasoh, A. & Ohara, T. 2002. Shock wave relation containing lane change source term for two-lane traffic flow. *Journal of the Physical Society of Japan,* 71**,** 2339-2347.

Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference, 2010. Scipy, 61.

Sen, B., Smith, J. D. & Najm, W. G. 2003. Analysis of lane change crashes. United States. National Highway Traffic Safety Administration.

Sharma, A., Ali, Y., Saifuzzaman, M., Zheng, Z. & Haque, M. M. Human Factors in Modelling Mixed Traffic of Traditional, Connected, and Automated Vehicles. International Conference on Applied Human Factors and Ergonomics, 2017. Springer, 262-273.

Sharma, A., Zheng, Z., Bhaskar, A. & Haque, M. M. 2019. Modelling car-following behaviour of connected vehicles with a focus on driver compliance. *Transportation research part B: methodological,* 126**,** 256-279.
Simons-Morton, B. G., Ouimet, M. C., Wang, J., Klauer, S. G., Lee, S. E. & Dingus, T. A. Hard braking events among novice teenage drivers by passenger characteristics. Proceedings of the... International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, 2009. NIH Public Access, 236.
Svenson, O. 2009. Driving speed changes and subjective estimates of time savings, accident risks and braking. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition,* 23**,** 543-560.
Talebpour, A., Mahmassani, H. S. & Hamdar, S. H. 2015. Modeling lane-changing behavior in a connected environment: A game theory approach. *Transportation Research Procedia,* 7**,** 420-440.
Tfnsw. 2019. *Centre for Road Safety* [Online]. Transport for New South Wales, Australia https://roadsafety.transport.nsw.gov.au/statistics/interactivecrashstats/nsw.html?tabnsw=2 [Accessed 26 November 2019].
Toledo, T., Koutsopoulos, H. & Ben-Akiva, M. 2003. Modeling integrated lane-changing behavior. *Transportation Research Record: Journal of the Transportation Research Board***,** 30-38.
Uno, N., Iida, Y., Itsubo, S. & Yasuhara, S. A microscopic analysis of traffic conflict caused by lane-changing vehicle at weaving section. Proc. 9th meeting of Euro Working Group on Transportation (EWGT), Bari, Italy, 2002.
Van Driel, C. J. & Van Arem, B. 2010. The impact of a congestion assistant on traffic flow efficiency and safety in congested traffic caused by a lane drop. *Journal of Intelligent Transportation Systems,* 14**,** 197-208.
Vechione, M., Balal, E. & Cheu, R. L. 2018. Comparisons of Mandatory and Discretionary Lane-Changing Behavior on Freeways. *International Journal of Transportation Science and Technology,* 7**,** 124-136.
Wall, G. & Hounsell, N. 2005. Microscopic modelling of motorway diverges. *European Journal of Transport and Infrastructure Research,* 5**,** 139-158.
Washington, S., Karlaftis, M. G. & Mannering, F. L. 2011. *Statistical and econometric methods for transportation data analysis*, Second ed, Chapman & Hall/CRC Boca Raton, FL.
Wu, G., Boriboonsomsin, K., Zhang, W.-B., Li, M. & Barth, M. 2010. Energy and emission benefit comparison of stationary and in-vehicle advanced driving alert systems. *Transportation research record,* 2189**,** 98-106.
Wu, K.-F. & Jovanis, P. P. 2013. Defining and screening crash surrogate events using naturalistic driving data. *Accident Analysis & Prevention,* 61**,** 10-22.
Zheng, B. 2000. Summarizing the goodness of fit of generalized linear models for longitudinal data. *Statistics in medicine,* 19**,** 1265-1275.
Zheng, Z. 2014. Recent developments and research needs in modeling lane changing. *Transportation research part B: methodological,* 60**,** 16-32.
Zheng, Z., Ahn, S., Chen, D. & Laval, J. 2011a. Applications of wavelet transform for analysis of freeway traffic: bottlenecks, transient traffic, and traffic oscillations. *Transportation Research Part B: Methodological,* 45**,** 372-384.
Zheng, Z., Ahn, S., Chen, D. & Laval, J. 2011b. Freeway traffic oscillations: microscopic analysis of formations and propagations using wavelet transform. *Procedia-Social and Behavioral Sciences,* 17**,** 702-716.
Zheng, Z., Ahn, S., Chen, D. & Laval, J. 2013. The effects of lane-changing on the immediate follower: Anticipation, relaxation, and change in driver characteristics. *Transportation research part C: emerging technologies,* 26**,** 367-379.

Zheng, Z., Ahn, S. & Monsere, C. M. 2010. Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis & Prevention,* 42**,** 626-636.
Zorn, C. 2006. Comparing GEE and robust standard errors for conditionally dependent data. *Political Research Quarterly,* 59**,** 329-341.