Journal of Safety Research 84 (2023) 353–363

Contents lists available at ScienceDirect

Journal of Safety Research

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

Ability of youth operators to reach agricultural all-terrain vehicles controls

Guilherme De Moura Araujo, Farzaneh Khorsandi Kouhanestani *, Fadi A. Fathallah

Department of Biological and Agricultural Engineering, University of California, Davis, Davis, CA, USA

ARTICLE INFO

Article history:
Received 7 September 2022
Received in revised form 12 September 2022
Accepted 21 November 2022
Available online 30 November 2022

Keywords:
Anthropometry
ATV
Children
Farm
Quadbike

ABSTRACT

Problem: Utility All-Terrain Vehicles (ATVs) are one major cause of youth injuries and fatalities on farms. Utility ATVs have heavy weights and fast speeds that require complex maneuvering. Youth's physical capabilities may not be sufficient to perform those complex maneuvers correctly. Therefore, it is hypothesized that most youth engage in ATV-related incidents because they ride vehicles unfit for them. There is a need to assess ATV-youth fit based on youth anthropometry. *Method:* This study focused on evaluating potential inconsistencies between the operational requirements of utility ATVs and the anthropometric measures of youth through virtual simulations. Virtual simulations were performed to assess 11 youth-ATV fit guidelines proposed by several ATV safety advocacy organizations (National 4-H council, CPSC, IPCH, and FReSH). In total, 17 utility ATVs along with male-and-female-youth of nine ages (8 to 16 years old) and three height percentiles (5th, 50th, and 95th) were evaluated. *Results:* The results demonstrated a physical mismatch between ATVs' operational requirements and youth's anthropometry. For example, male-youth aged 16 of the 95th height percentile failed to pass at least 1 out of the 11 fit guidelines for 35 % of all vehicles evaluated. The results were even more concerning for females. Female youth 10 years old and younger (from all height percentiles) failed to pass at least one fit guideline for all ATVs evaluated. *Discussion:* Youth are not recommended to ride utility ATVs. *Practical Applications:* This study provides quantitative and systematic evidence to modify current ATV safety guidelines. Furthermore, youth occupational health professionals could use the present findings to prevent ATV-related incidents in agricultural settings.

1. Problem

The use of utility All-Terrain Vehicles (ATVs) as working machines adds a heavy burden to the American public health system (Helmkamp, Marsh, & Aitken, 2011). According to data from the 2019 National Electronic Injury Surveillance System, over 95,000 emergency department (ED) visits were due to an ATV-related incident. Around 36.8 % of those ED visits involved youth younger than 18 years old, and 15.3 % of the incidents happened on farms or ranches (Wiener, Waters, Harper, Shockey, & Bhandari, 2022). Indeed, using utility ATVs in the farm setting is extremely dangerous for youth; ATVs are one of the most frequently cited causes of incidents among farm youth (Hendricks & Hard, 2014; Weichelt & Gorucu, 2018).

ATVs have three or four low-pressure tires, narrow wheelbase, and high center of gravity (Ayers, Conger, Comer, & Troutt, 2018; Chou, Khorsandi, Vougioukas, & Fathallah, 2022; House, Schwebel, Mullins, Sutton, Swearingen, Bai, & Aitken, 2016). Due to safety concerns, the production of three-wheelers ceased in the United States in 1987 (Voreacos, 1987). Three-wheelers were known to be even more prone to rollovers than four-wheeled ATVs (David, 1998).

Utility ATVs and sport models (which include youth ATV models) have several design differences. Utility models have higher ground clearance, stronger torque for hauling and towing, rear and front racks for carrying loads or mounting equipment, a hitch to pull implements, and heavyer weights (Khorsandi et al., 2021). Accordingly, utility ATVs are more suitable and more commonly used for tasks in agricultural settings. Therefore, in this study, agricultural ATVs are defined as utility ATVs used on farms and ranches.

Agricultural ATVs have heavy weights and fast speeds that require complex maneuvering. Youth's physical capabilities may not be sufficient to perform those complex maneuvers correctly. In fact, many studies have shown that youth are more vulnerable to injuries than adults because of their less developed physical

* Corresponding author.
E-mail addresses: gdemoura@ucdavis.edu (G. De Moura Araujo), fkhorsandi@ucdavis.edu (F. Khorsandi Kouhanestani), fathallah@ucdavis.edu (F.A. Fathallah).

https://doi.org/10.1016/j.jsr.2022.11.010
0022-4375/

Nomenclature

Name	Abbreviation
4-Wheel-Drive	4WD
All-terrain vehicle	ATV
American Academy of Pediatrics	AAP
American National Standards Institute	ANSI
ATV Safety Institute	ASI
Cohen's Kappa Coefficient	K
Computer-Aided Design	CAD
Crush Protection Device	CPD
Cubic Capacity	cc
Department of Trade Industry	DTI
Electric Power Steering	EPS
Emergency Department	ED
Farm and Ranch eXtension in Safety and Health	FReSH
General Accounting Office	GAO
Intermountain Primary Children's Hospital	IPCH
Loss of Control Event	LCE
National 4-H Council	N4-HC
National Children's Center for Rural and Agricultural Health and Safety	NCCRAHS
National Safe Tractor and Machinery Operation Program	NSTMOP
Seat Reference Point	SRP
Specialty Vehicle Institute of America	SVIA
Three-dimensional	3-D
U.S. Consumer Product Safety Commission	CPSC
Virtual Reality	VR

capabilities and psychological and behavioral characteristics (Brison et al., 2006; Hard & Myers, 2006; Hendricks, Myers, Layne, & Goldcamp, 2005; Marlenga, Pickett, & Berg, 2001; Pollack-Nelson, Vredenburgh, Zackowitz, Kalsher, & Miller, 2017; Reed, Ebert-Hamilton, Manary, Klinich, & Schneider, 2005; Serre et al., 2010; Towner & Mytton, 2009), which likely affect their ability to safely operate agricultural vehicles (Bernard et al., 2010; Chang, Fathallah, Pickett, Miller, & Marlenga, 2010; Fathallah, Chang, Berg, Pickett, & Marlenga, 2008; Fathallah, Chang, Pickett, & Marlenga, 2009). Furthermore, previous studies have shown that ATV-rider misfit is another important risk factor (Bernard et al., 2010; Jennissen, Miller, Tang, & Denning, 2014).

Despite compelling evidence showing that utility ATVs are unsuitable for youth, the most popular guidelines for ATV-youth fit disregard the rider's physical capabilities. Instead, those recommendations are based on the rider's age (Academy, 2018), vehicle's maximum speed (ANSI/SVIA, 2017), vehicle's engine size (CPSC, 2006), or farm machinery training certificate (Garvey, Murphy, Yoder, & Hilton, 2008). For instance, youth as young as 14 can operate utility ATVs while employed on non-family-owned farms if they receive training through an accredited farm machinery safety program, such as the National Safe Tractor and Machinery Operation Program (NSTMOP) (Garvey et al., 2008). The NSTMOP training includes tractor and ATV education, where students must pass a written knowledge exam and a functional skills test to receive a certificate (Murphy, 2020). Nevertheless, programs such as the NSTMOP lack appropriate coverage of specific ATV-related subjects, such as active riding and physical matches of ATVs and youth.

If the ATV is not fit to the rider, they will likely be unable to properly operate the ATV's controls, which increases their chance of incidents and consequently may lead to injuries and fatalities. In addition, the traditional guidelines adopted to fit ATVs for youth are inconsistent in evaluating their preparedness to ride. The suggested fitting criteria are subject to variances in state law and lack scientifically-based evidence. While some recommendations based upon the riders' physical capabilities exist (CPSC, 2006; FReSH, 2012; IPCH, 2018; National 4-H Council, 2005), the adoption of these recommendations has not gained attention because they are not comprehensive and lack quantitative and systematic data.

Recommendations based on riders' physical capabilities appear to provide a better foundation to determine if the machine is suitable for the rider (Bernard et al., 2010). Therefore, there is a need to evaluate youth-ATV fit based on the riders' physical capabilities (e.g., anthropometry, strength, and field of vision).

Since 95 % of all ATV-related fatalities involving youth between 1985 and 2009 included agricultural ATVs (Denning, Harland, & Jennissen, 2014), the purpose of this study is to evaluate the mismatches between the operational requirements of utility ATVs and the anthropometric characteristics of youth.

It has been hypothesized that youth are mainly involved in ATV incidents because they ride vehicles unfit for them. This study evaluated ergonomic inconsistencies between youth's anthropometric measures and utility ATVs' operational requirements. The ability of youth to safely operate ATVs was evaluated through computer simulations that comprised 11 fit criteria and male-and-female youth of varying ages (8–16 years old) and height percentiles (5th, 50th, and 95th) operating 17 utility ATV models.

2. Methods

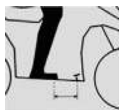
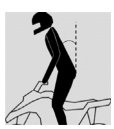
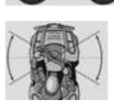
Youth-ATV fit was analyzed through virtual simulations and was carried out in five steps. First, 11 guidelines were identified for the fit of youth and ATVs. The second step consisted of identifying a database containing anthropometric measures of youth of various ages (8–16 years old), genders (males and females), and height percentiles (5th, 50th, and 95th). The third step consisted of collecting the dimensions of 17 ATV models to create a three-dimensional (3-D) representation of them. The fourth step consisted of using SAMMIE CAD (SAMMIE CAD Inc., Leics., UK) and Matlab (Matlab, v2021a; Mathworks, Natick, MA) to evaluate if the youth's anthropometric measures conform to the guidelines identified in step one. Lastly, the results of the virtual simulations were validated in field tests with actual riders and ATVs.

2.1. Fit criteria

The fit criteria provide movement-restraint thresholds that check if the rider can safely reach all controls and perform active riding, which requires the operator to shift their center of gravity to maintain the vehicle's stability, especially when turning or traveling on slopes (Thorbole, Aitken, Graham, Miller, & Mullins, 2012). Maintaining a correct posture is essential because, otherwise, the rider's ability to control the vehicle is compromised, which puts them and potential bystanders at risk.

The reach criteria considered in this study were selected based on the recommendations of the following institutions: (a) National 4-H Council (2005), (b) U.S. Consumer Product Safety Commission (CPSC) (2006), (c) Intermountain Primary Children's Hospital (IPCH) (2018), and (d) Farm and Ranch eXtension in Safety and Health (FReSH) Community of Practice (2012). Disregarding overlaps, these guidelines consisted of 11 anthropometric measures of fit, which are presented in Table 1.

Table 1
ATV-rider fit criteria.

ID	Criterion		Institution(s)	"Fit success" and reasoning for each criterion
1	Handlebar-knee distance		National 4-H Council, CPSC	Handlebar–knee distance > 200 mm. This is necessary to ensure the rider can reach the handlebar and steer around obstacles.
2	Hand size compared to ATV handlebar reach		National 4-H Council, IPCH	With hand placed in the normal operating position and fingers straight out, the first joint from the tip of the middle finger extends beyond the brake lever. This is important to guarantee that the rider can activate the brake lever.
3	Brake-foot position		National 4-H Council	Distance from the "ball" of the foot (at its most rearward position in the ATV's foot well) to the brake pedal divided by the length of the foot < 105 %. A disproportional rate indicates a risk for ineffective foot–brake operation.
4	Standing-seat clearance		National 4-H Council, CPSC, FReSH	Clearance zone between rider's crotch and ATV seat > 150 mm. This is important to guarantee that the rider can rise the torso up from the ATV seat to maintain balance and avoid distracting longitudinal torso impacts that occur while traversing rough terrains.
5	Elbow angle		National 4-H Council, IPCH	A narrow elbow angle (<90°) indicates excessive arm flexion, while an angle too wide (>135°) indicates the arms are excessively straight due to the grips being too far apart, which forces the rider to lean the torso to the outside of the turn to achieve an adequate range of handlebar turning
6	Upper leg		National 4-H Council	Upper leg within 10° of parallel to the ground. An upper leg too far off from parallel to the ground can compromise the rider's ability to activate the foot brake and keep balance.
7	Angle of lean from vertical		CPSC	Angle of lean from vertical < 30°. This is important to guarantee a correct posture while riding the ATV. Leaning forward significantly over the handlebars to steer when raised off the seat, can shift the system's center of gravity, increasing the likelihood of the ATV tipping forward.
8	Control reach		CPSC	Riders must be able to reach all ATV controls while seated upright.
9	Footrest reach		CPSC	Riders must keep their feet firmly on the footrests when not activating the foot-brakes. This is important to ensure the rider can maintain balance and not lose control of the ATV.
10	Knee angle		CPSC	Knee angle at least 45° while sitting and with the feet flat on the footrest. An angle wider than 45° indicates a risk for ineffective foot-brake operation.
11	Control grip		CPSC, FReSH	Riders must keep a grip on the handlebar and maintain throttle and brake control when turning the handlebar from lock to lock position. This is especially important while performing a sharp turn or a swerve.

2.2. Human mockups

Human mockups were developed in SAMMIE CAD. This computer program allows users to create customized virtual humans based on eight anthropometric dimensions, as shown in Fig. 1a. In total, 54 youth mockups were created, a combination of two genders, nine ages (8–16), and three body size percentiles in height (5th, 50th, and 95th). The age range was selected because most youth start operating farm machinery at 8 years old (Marlenga et al., 2001), and most ATV-related crashes occur with riders younger than 16 years old (Denning et al., 2014). Two adult mockups (male and female of the 50th body-size percentile) were also created to establish a baseline for comparisons. The anthropometric measures used as input to SAMMIE CAD were retrieved from the database of Snyder et al. (1977), which includes measurements from 3,900 subjects from 2 to 18 years of age for both genders. The adopted anthropometric measures were based on the mean values of groups of subjects with the same age, gender, and height.

One of the required inputs (seated shoulder height) was not available in the database used for this study. Therefore, the missing input was computed using the available data. The seated shoulder height was calculated by subtracting the head and neck length from the seated height (Fig. 1b).

2.3. ATV mockups

In total, 17 utility ATV models were evaluated. Selected models consisted of vehicles of varying engine sizes (200–700 cc) from the most common ATV manufacturers on U.S. farms (Apollo, Arctic Cat, CF Moto, Honda, Polaris, and Yamaha). General descriptive vari-

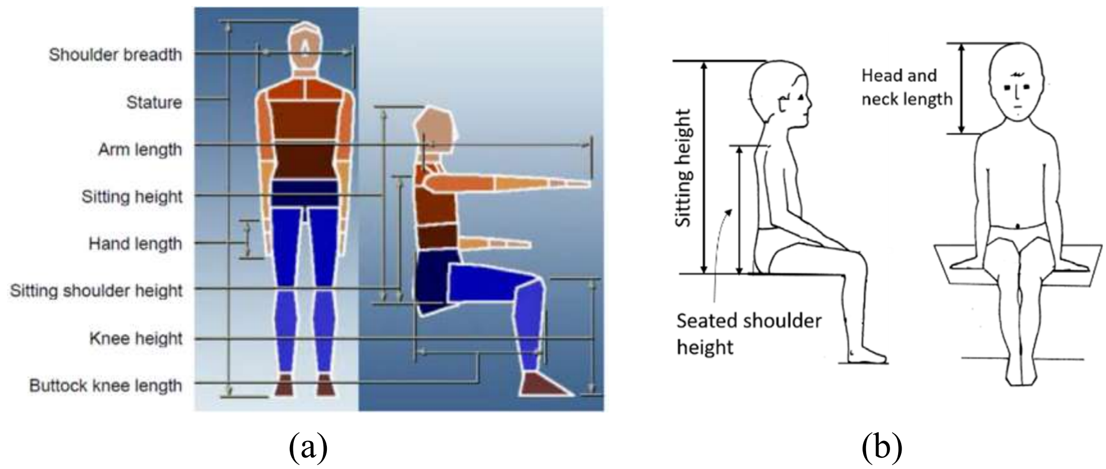


Fig. 1. SAMMIE CAD human creation. (a) Selected input variables (source: SAMMIE CAD Inc.); (b) Interpolation of missing variable (seated shoulder height) - Adapted from Snyder et al. (1977).

ables such as manufacturer, model, series, engine capacity (cc), drive terrain (4 W/2W), transmission, and suspension type were recorded.

ATV mockups were developed based on the spatial coordinates (X, Y, Z) of selected ATV features (e.g., ATV seat, chassis, handlebars, footrests, and controls). An original attempt to record spatial coordinates of ATV features consisted of using Photogrammetry, a technique in which several pictures of an object are taken from various angles and then processed to create a 3-D model. Nevertheless, this technique proved inefficient, as initial trials were time-consuming, and the results had unsatisfactory accuracy. A second attempt consisted of using a virtual reality (VR) tracking system. This alternative proved fast to implement with excellent accuracy (±1 mm); hence, this technique was selected and presented in the following section.

2.3.1. Data acquisition

The VR tracking system (Vive – HTC Corporation, China) utilized in this experiment consisted of two controllers and two infrared laser emitter units (lighthouses). The system allows the user to move in 3-D space and use motion-tracked handheld controllers to interact with the environment. The system uses the lighthouses to shoot horizontal and vertical infrared laser sweeps that are detected by photodiodes positioned in the surrounding of the controller's surface (Niehorster, Li, & Lappe, 2017). The position and orientation of the controllers are calculated by the difference in time at which each photodiode is hit by the laser (Kreylos, 2016). By placing the controller over selected vertices of ATV features, it was possible to record their spatial coordinates, which allowed the development of the 3-D ATV mockups.

A custom program was developed to calibrate the system, log, and manipulate data. This program was initially retrieved from Kreylos (2016) and then modified to meet the specific needs of the present study. The software runs in Linux operating systems and has several functionalities that are useful to the user. Examples of these functionalities are a 3-D grid, which allows for real-time visualization of labeled points, and a measuring tool (to verify the measurement scale).

A probe was custom-manufactured and attached to the controllers to ease the calibration process and data collection. The probe was made of metal and had a rounded tip, which made it wear-resistant and prevented it from damaging the ATVs. The measurements were collected inside a tent covered by a white rooftop that reduces the interference of solar rays in the communication between the lighthouses and the photodiodes in the controllers. In total, 38 points were collected per ATV. The points were selected aiming to get an efficient representation of all selected ATV controls (hand brake lever, foot brake pedal, steering handlebar, throttle lever, hand gearshift lever, and foot gearshift pedal) and additional features that were used to assist the virtual simulations, such as the seat and the footrests. After data filtering, the data were processed in SAMMIE CAD for a 3-D representation of the evaluated vehicle, as shown in Fig. 2.

2.4. Data analysis

ATV-rider fit was evaluated through SAMMIE CAD and Matlab. Fit criteria 4, 5, 6, 7, 8, 9, and 10 (Table 1) were evaluated in SAMMIE CAD because their assessment involved complex interactions between riders and ATVs, such as measuring the angle of the rider's knee while riding. SAMMIE CAD provides a 3-D environment and full control of human mockups, which makes it possible to evaluate those complex interactions. The simulations performed in SAMMIECAD consisted of: (1) creating 3-D human mockups; (2) creating 3-D ATV mockups; and (3) integrating (1) and (2) in the virtual environment to simulate their interaction. For each simulation, the correct reach posture was achieved by positioning the human limbs according to the specific task's requirement. For example, a seated position was adopted when evaluating fit criterion 10 (knee angle), as shown in Fig. 3a. On the other hand, a standing straddling posture was selected when evaluating fit criterion 4 (clearance zone between the rider's crotch and ATV seat), as shown in Fig. 3b.

Some criteria involve the youth reaching a specific control (e.g., criteria 5, 7, 8, and 9). The feature "Reach" under the "Human" menu on SAMMIE CAD was used to evaluate the ability of the youth mockups to reach the selected controls. The "Reach" was set as "Absolute," and "Object Point" was set as "Control." When the selected control could be successfully reached, the software would display an animation of the human limb reaching the desired object (the rider was assigned a score of 1 – meaning that they fulfilled the requirements of that criterion). On the other hand, if the control was out of reach, SAMMIE CAD would show an error window and display the required distance for the human limb to reach the desired control (the rider was assigned a score of 0 – meaning that they failed to pass that specific criterion).

Simulations involving buttons and levers were performed with the fingertip of the index finger or the thumb, accordingly. Simulations involving levers or the handlebars were performed with palm-grip-hand postures. All controls on the right side of the

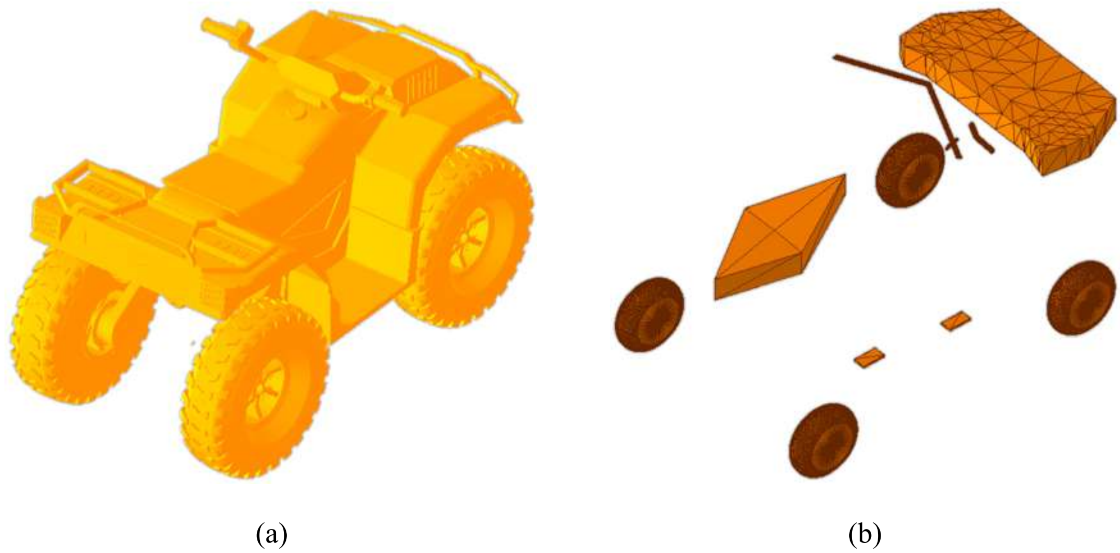


Fig. 2. 3-D representation of ATV mock-ups. (a) Fully assembled model – for visualization purposes only; (b) Example of a 3-D ATV mock-up used for the virtual simulations in SAMMIE CAD.

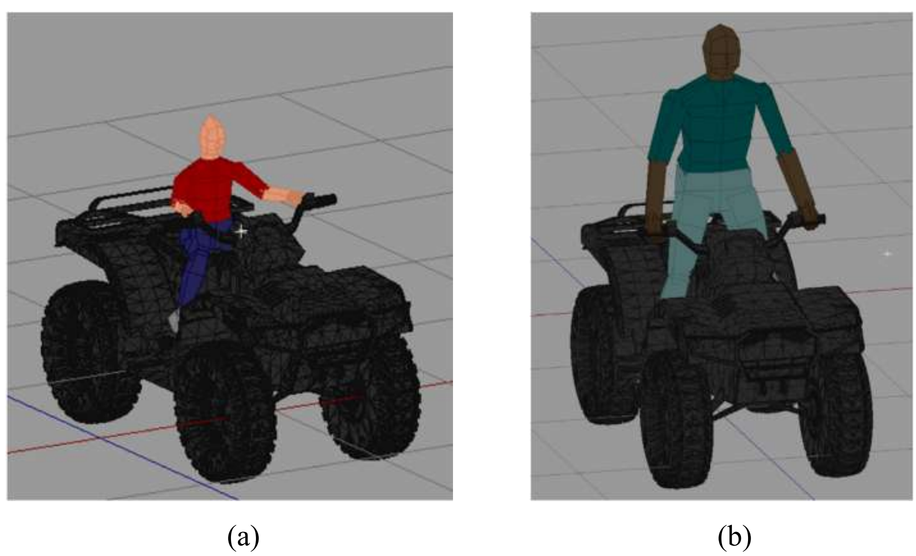


Fig. 3. Different reach postures. (a) Seated posture (9 yr. old – 5th percentile boy); (b) Standing straddling posture (16 yr. old – 95th percentile boy).

ATV were simulated with the right hand/foot, and all controls on the left side of the ATV were simulated with the left hand/foot. Specific controls that required using both hands, such as the handlebars, were simulated with both hands.

Criteria 1, 2, 3, and 11 were evaluated through Matlab because their assessment required the computation of simpler calculations, such as the distance between the rider's knee and the ATV's handlebars. Matlab also provided the ability to automate the calculations for a more efficient data analysis. A code was generated based on conditional statements to assess whether riders' anthropometric measures conformed to the constraints imposed by the ATV design. For instance, when evaluating criterion 1, the distance between the ATV footrests and the handlebars minus the rider's knee height must be greater than 200 mm (Table 1).

For each reach criterion, riders received a binary score (1 if the rider fulfilled the requirements of that criterion; and 0 otherwise). Riders with a total score of 11 (adequate reach for all evaluated criteria) were classified as "capable of riding the ATV." On the other hand, riders with a total score below 11 (inadequate reach of at least one or more criteria) were classified as "not capable of riding the ATV."

2.5. Validation

In order to validate the results of the virtual simulations, an experiment including three adults (two males and one female) and one study ATV (model Yamaha Grizzly EPS – 700) was carried out. Each subject had completed an ATV safety riding course prior to the experiment and was awarded a certificate from the *ATV Safety Institute* (ATV-Safety-Institute, 2009). The capability of the subjects to fulfill each fit criterion was evaluated and recorded. For the field tests, a measuring tape graduated in mm was used to measure distances and a digital angle finder (General Tools & Instruments LLC., New York, NY, USA) to measure angles. To assist in some of the angle measurements, a straight edge 48″ ruler (model J48EM, Johnson level & Tool, Mequon, WI, USA) and a magnetic level (model 7500 M, Johnson level & Tool, Mequon, WI, USA) were used.

The anthropometric measures of the subjects were taken with a body-measuring tape and then used as input in SAMMIE CAD to create 3-D mockups. The results observed in the experimental setting were then compared to those observed in the virtual simulations through the Cohen's Kappa coefficient (K) (Landis & Koch, 1977), which is a statistic widely used to measure inter-rater reliability for qualitative (categorical) items (McHugh, 2012). A Z-test ($\alpha = 0.05$) was performed to evaluate whether the value of K was statistically different than zero, which would imply that the virtual simulations are reasonable.

3. Results

Seventeen ATV models were evaluated from eight different manufacturers. Engine capacity ranged from 174-686 cc, with most vehicles in 100–400 cc (35 %). Moreover, 58 % of the ATVs evaluated included electric power steering (EPS), 4 wheel-drive (58 %), solid suspension (88 %), and manual transmission (48 %).

Findings of individual reach criteria for the ATV models are presented in Tables 2 and 3, for males and females, respectively. The last column of those tables (Total) represents the percent of observations for which riders scored 11 points (i.e., they fulfilled the requirements of all 11 fit guidelines). Criterion 1 (Handlebar-knee distance) seemed difficult for 16-year-old-males of the 95th body-size percentile. This result may be attributed to the height of these subjects, which decreases the gap between their knee and the handlebars (Bernard et al., 2010).

Unlike criterion 1, criterion 2 (hand size compared to ATV handlebar reach) did not present any difficulty for the virtual youth (Tables 2 and 3). Indeed, virtual subjects of all ages, body-size percentiles, and genders succeeded in this criterion for all (100 %) evaluated vehicles.

Criteria 3, 4, 6, 7, 8, 9, 10, and 11 all presented a similar trend where young riders do not conform well to these criteria, but older riders do (Tables 2 and 3). The contrast in success rate among subjects of different ages and height percentiles are likely also attributed to the variations in height among the subjects. For example, virtual 8-year-old-female riders of the 95th percentile did not pass criterion 5 for any of the evaluated ATVs. In contrast, their 16-year-old-counterpart passed the same criterion for 75 % of the evaluated ATVs (Table 3), a surprising difference of 75 %.

The results from Tables 2 and 3 indicate that 8-year-old youth would probably not be able to control utility vehicles when traversing rough or uneven terrains (Criterion 4 – Standing seat clearance). This finding likely explains the fact that youth are more subject to loss of control events (LCEs) than adults (McBain-Rigg, Franklin, McDonald, & Knight, 2014).

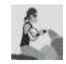
The results of the simulations related to Criterion 7 (Angle of lean from vertical) indicated that youth 9 years old and younger are more likely to lean forward over 30° (safety threshold) when raised off the seat to reach the handlebars of agricultural ATVs. As a result, the center of gravity of the ATV can shift forward, thus increasing the chances of a tip over.

Lastly, some results of the simulations related to Criterion 5 (elbow angle) were concerning. Males up to 11 years old and females up to 13 of the 50th percentile passed this criterion for less than 50 % of the evaluated ATVs.

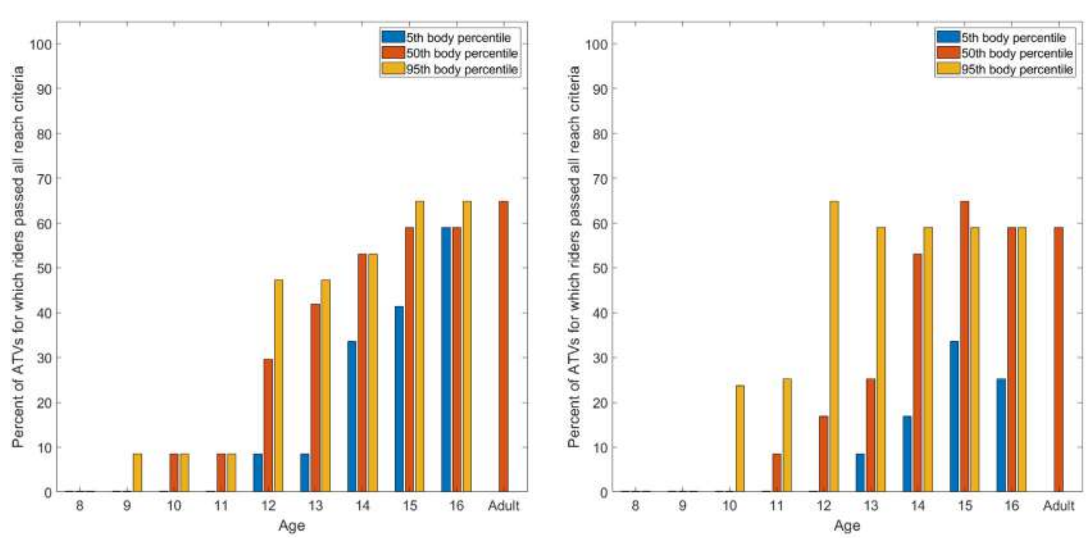
Table 2
Percent of observations (n = 17) for which reach criteria did not limit adult-sized ATV usage by males of various ages and percentiles.

Age	Percentile	Criteria											
													
		1	2	3	4	5	6	7	8	9	10	11	Total
8	5th	94	100	65	25	0	0	42	42	0	0	6	0
	50th	94	100	77	33	0	8	50	58	8	8	12	0
	95th	94	100	94	67	0	8	83	83	8	8	35	0
9	5th	94	100	77	50	0	0	42	42	0	0	12	0
	50th	94	100	94	83	0	8	58	67	8	8	29	0
	95th	94	100	94	92	8	50	83	92	50	50	41	8
10	5th	94	100	77	42	0	8	58	67	8	8	12	0
	50th	94	100	94	92	8	25	92	100	25	25	35	8
	95th	94	100	94	100	8	58	92	100	58	58	65	8
11	5th	94	100	94	92	0	8	92	100	8	8	29	0
	50th	94	100	94	100	8	50	92	100	50	50	41	8
	95th	94	100	94	100	8	58	92	100	58	58	71	8
12	5th	94	100	94	92	8	42	92	100	42	42	41	8
	50th	94	100	94	100	33	58	92	100	58	58	65	29
	95th	88	100	94	100	58	92	92	100	92	92	88	47
13	5th	94	100	94	100	8	50	92	100	50	50	35	8
	50th	94	100	94	100	42	92	92	100	92	92	71	42
	95th	82	100	94	100	67	92	92	100	92	92	88	47
14	5th	94	100	94	100	33	58	92	100	58	58	71	33
	50th	94	100	94	100	58	92	92	100	92	92	88	53
	95th	82	100	94	100	92	92	92	100	92	92	88	53
15	5th	94	100	94	100	42	58	92	100	58	58	71	41
	50th	88	100	94	100	83	92	92	100	92	92	88	59
	95th	82	100	94	100	100	100	92	100	100	100	88	65
16	5th	88	100	94	100	67	92	100	100	92	92	88	59
	50th	82	100	94	100	92	92	100	100	92	92	88	59
	95th	71	100	94	100	92	100	100	100	100	100	88	65
Adult	50th	82	100	94	100	100	92	100	100	92	92	88	65

Table 3
Percent of observations (n = 17) for which reach criteria did not limit adult-sized ATV usage by females of various ages and percentiles.

Age	Percentile	Criteria											
													
		1	2	3	4	5	6	7	8	9	10	11	Total
8	5th	94	100	53	17	0	0	8	8	0	0	6	0
	50th	94	100	77	25	0	0	25	25	0	0	12	0
	95th	94	100	94	75	0	8	83	83	8	8	12	0
9	5th	94	100	77	42	0	0	58	58	0	0	12	0
	50th	94	100	88	75	0	8	83	83	8	8	12	0
	95th	94	100	94	83	0	33	92	92	33	33	41	0
10	5th	94	100	82	58	0	0	67	67	0	0	12	0
	50th	94	100	94	75	0	25	92	92	25	25	35	0
	95th	94	100	94	100	25	67	100	100	67	67	65	24
11	5th	94	100	88	75	0	8	100	100	8	8	18	0
	50th	94	100	94	92	8	42	100	100	42	42	41	8
	95th	94	100	94	100	25	75	100	100	75	75	77	25
12	5th	94	100	94	83	0	33	100	100	33	33	29	0
	50th	94	100	94	100	17	58	100	100	58	58	65	17
	95th	94	100	94	100	67	92	100	100	92	92	88	65
13	5th	94	100	94	92	8	33	100	100	33	33	41	8
	50th	94	100	94	100	25	67	100	100	67	67	77	25
	95th	88	100	94	100	67	92	100	100	92	92	88	59
14	5th	94	100	94	92	17	33	100	100	33	33	65	17
	50th	94	100	94	100	58	83	100	100	83	83	77	53
	95th	88	100	94	100	67	92	100	100	92	92	88	59
15	5th	94	100	94	100	33	58	100	100	58	58	71	33
	50th	94	100	94	100	67	92	100	100	92	92	88	65
	95th	82	100	94	100	75	92	100	100	92	92	88	59
16	5th	94	100	94	100	25	67	100	100	67	67	77	25
	50th	94	100	94	100	67	83	100	100	83	83	88	59
	95th	82	100	94	100	75	92	100	100	92	92	88	59
Adult	50th	94	100	94	100	67	75	100	100	75	75	77	59

The percent of ATVs in which riders passed all criteria is presented in Fig. 4. The main finding is that certain youth should not ride most utility ATVs. For instance, the average (50th percentile) male operator aged 16 passed all 11 safety criteria for less than 60 % of the evaluated vehicles. That number decreases sharply for younger youth or youth of the same age but smaller height percentile. A similar trend was also observed for female operators.



(a) (b)

Fig. 4. Percent of observations for which riders passed all 11 fit criteria. (a) Males and (b) Females.

3.1. Validation

The results of the validation tests are presented in Table 4 and summarized in a confusion matrix (Table 5). In the confusion matrix, the outcome of the test (pass/no pass) is labeled in both horizontal and vertical axes. The horizontal axis represents the number of outcomes predicted by the virtual simulations, and the vertical axis represents the ground truth data (field experiments). The results of the virtual simulations were very close to those of the field tests, with a total accuracy of 88 %.

The Z-test determined that the Cohen's Kappa coefficient (K = 0.45) was significantly greater than zero (p = 0.036), indicating that the virtual simulations are reasonable. This approach to evaluate ergonomic inconsistencies between youth's anthropometry and the operational requirements of ATVs proved to be an effective and accurate technique.

Not all results of the virtual simulations matched those of the field tests. One unexpected result is related to criterion 6 (upper leg angle). It was observed that the mean angle between the riders' upper leg and the horizontal plane (parallel to the ground) was 16.7°, slightly above the recommended threshold (10°). Similarly, two subjects failed to pass criterion 5 (elbow angle) in the actual field tests but passed it in the virtual simulation.

During the field tests, riders were asked to sit comfortably as if they were just about to start riding the ATV. We argue that it would be possible for riders to adjust their way of sitting so they would pass both fit criteria; however, it would not result in the most ergonomic posture from the rider's standpoint. On the other hand, in the virtual simulations, our ultimate goal was to place the 3-D subjects' mockups to physically conform to the proposed fit criteria. Thus, it was impossible to predict whether the final adopted postures in the simulations would match those selected by the riders in the validation tests. Therefore, we argue that despite some outcomes of the virtual simulations did not match those of the field tests, the results of the virtual simulations are still reasonable. One just has to be cognizant that the outcomes of the virtual simulations represent a hypothetical scenario where the rider is able to attain a posture based on their anthropometric measures relative to the ATV, not on their preferences.

Table 5
Confusion matrix based on the validation tests.

Actual outcome (field tests)		Pass	No Pass
	Pass	27	0
	No pass	4	2
		Predicted outcome (virtual simulations)	

5. Discussion

This study evaluated limitations in youth's anthropometric dimensions when riding commonly used ATVs. Using a combination of actual field measurements and a novel digital simulation approach, the present study evaluated 11 ATV fit criteria for youth. The major finding was that youth are not recommended to ride adult-sized ATV models, which is a common practice in the United States (Bernard et al., 2010; Office (GAO), 2010; Jennissen et al., 2014). This finding raises serious concern regarding youth's ability to ride ATVs, especially when unsupervised.

5.1. Limitations of youth

The present findings outlined that some youth are too small, which makes them incapable of properly reaching the vehicle's hand/foot brakes, resting their feet on the footrests, or having to lean forward beyond 30° to reach the handlebars when rising off the seat. Failing to activate the ATV brakes limits the youth's ability to reduce the speed or to stop the vehicle, which likely prevents them from avoiding unexpected hazards, such as obstacles or bystanders (Fathallah et al., 2008). In fact, previous research has shown that a significant number of ATV incidents include hitting a stationary object (Balthrop et al., 2007; Concannon et al., 2012; Helmkamp et al., 2011; Jennissen, Wetjen, Hoogerwerf, O'Donnell, & Denning, 2018; Lower & Herde, 2012).

In addition, the inability to place the feet on the footrests when not breaking the ATV entails a functional loss of control of the vehicle. ATV LCEs occur frequently and are a significant cause of injury and death in agriculture (Carman et al., 2010; Clay, Treharne, Hay-Smith, & Milosavljevic, 2014; Milosavljevic et al., 2011). This finding indicates an opportunity for manufacturers to consider changing the design of their machines, allowing riders to adjust the ATV's seat height, which would likely reduce longitudinal torso impact while traversing rough and uneven terrains. Furthermore, leaning beyond 30° can cause the ATV to tip forward, resulting in a rollover. Most ATV-related crashes on farms and ranches, especially those resulting in deaths, involve rollovers (Cavallo, Gorucu, & Murphy, 2015; Chou et al., 2022; Khorsandi, Ayers, & Fong, 2019; Lower & Herde, 2012; Lower, Monaghan, & Rolfe, 2016; McIntosh, Patton, Rechnitzer, & Grzebieta, 2016).

On the other hand, some youth are too tall, which decreases the clearance zone between their legs and the handlebars. A clearance zone smaller than 200 mm makes it difficult for the rider to properly reach and steer the handlebars (CPSC, 2006; National 4-H Council, 2005). Consequently, riders may lose control of the vehicle (Clay, Hay-Smith, Treharne, & Milosavljevic, 2015; McIntosh et al., 2016) or have difficulty keeping it at a safe speed. As mentioned before, these series of events can lead to injuries and deaths.

Table 4
Validation tests separated by subject and specific fit criterion.

Subject	Subject 1 (male)		Subject 2 (male)		Subject 3 (female)	
Criterion	Real	Virtual	Real	Virtual	Real	Virtual
1	1	1	1	1	1	1
2	1	1	1	1	1	1
3	1	1	1	1	1	1
4	0	0	1	1	1	1
5	1	1	0	1	0	1
6	0	1	0	1	0	0
7	1	1	1	1	1	1
8	1	1	1	1	1	1
9	1	1	1	1	1	1
10	1	1	1	1	1	1
11	1	1	1	1	1	1

Furthermore, despite some results showing that youth are capable of riding many of the ATVs evaluated in this study, other risk factors such as experience, psychological, and cognitive development cannot be overlooked (FReSH, 2012; NCCRAHS, 2018). Youth who are high in thrill-seeking are more likely to engage in risky ATV riding behaviors, regardless of their safety awareness (Jinnah & Stoneman, 2016). Those cases require external interventions, such as changes in legislation, improved ATV design, and use of crush protection devices (Jinnah & Stoneman, 2016).

5.2. Lack of inclusive designs

The results indicated that most utility ATV models are unfit for youth. As such, there is an increased chance of incidents when youth ride these vehicles. There is a need to design ATVs that better accommodate riders of various sizes.

5.3. Assessment of ATV-youth fit guidelines

The results of this validation experiment showed that some riders failed criteria 5 and 6 even though they seemed able to operate the study vehicle comfortably and safely according to our ATV safety research team. Particularly, subjects 1, 2 and 3 presented elbow angles of 129°, 170° and 172.5°, respectively. While fit criterion 5 recommends an elbow angle between 90 ° and 135°, it is not uncommon to see motorcycle riders reporting comfortable elbow angle values up to 168° (Arunachalam, Singh, & Karmakar, 2021).

Moreover, subjects 1, 2, and 3 presented upper leg angles of 14°, 14.7°, and 21.4°, respectively (above the recommended threshold of 10°). A previous survey regarding motorcycle riders' perceived comfortable posture reported optimum upper leg angles as high as 23° (Arunachalam et al., 2021). It is our understanding that fit guidelines 5 and 6 are rather conservative, and their proposed thresholds may rule out riders that are perfectly able to ride utility ATVs safely and comfortably. As such, we propose some modifications to those fit guidelines.

First, we recommend that the rider's elbow angle should be between 90° and 170° as long as the rider feels comfortable steering the handlebars and is able to pass fit criteria 8 (control reach) and 11 (control grip). Moreover, we recommend that the rider's upper leg angle should be within 23° of parallel to the ground as long as the rider is able to pass criteria 3 (brake-foot position), 8 (control reach), and 9 (footrest reach). These new thresholds were selected based on the empirical results of our validation experiments and the angle values reported in the previously mentioned survey (Arunachalam et al., 2021).

Lastly, we stress that the fit guidelines are essential to assess whether the machine is suitable to the rider. We strongly recommend that stakeholders consider the fit criteria when evaluating youth's readiness to ride a utility ATV.

5.4. Changes in guidelines and policies for youth operating ATVs

Current guidelines for ATV-youth fit are mainly based on the rider's age (Academy, 2018) and vehicle's engine size (CPSC, 2006) and maximum speed (ANSI/SVIA, 2017). However, these recommendations are not supported by the present findings, which clearly showed that some fit criteria favor smaller youth while some benefit taller youth, regardless of the rider's age and vehicle's engine size or maximum speed. Furthermore, previous studies have also demonstrated that only rider's age and ATV characteristics are insufficient to evaluate youth-ATV fit (Bernard et al., 2010; De Moura Araujo and Khorsandi, 2020; De Moura Araujo, Khorsandi, Kabakibo, & Kreylos, 2021). As such, we strongly recommend that parents, dealerships, youth occupational health professionals, and policy makers adopt fit guidelines based on the reach ability of youth for the assessment of youth-ATV fit.

5.5. Study limitations

There are noteworthy limitations of this study that need to be considered when interpreting the results. First, one may argue that the database selected for this study (Snyder et al., 1977) is outdated. Nevertheless, to the best of our knowledge, this is the only available source that includes enough parameters to create youth mockups on SAMMIE CAD. In addition, there is no clear evidence of the secular trend in anthropometry over U.S. youth over the past 40 years (Fathallah et al., 2009). For instance, when investigating other sources (CHILDATA – DTI (1995)), we did not observe any significant differences (p-value < 0.05) in the mean values of shoulder breadth and hand length for youth aged 5 or 10 years old. However, it is reasonable to assume that there might be differences in the sizes of the youth population of 2022 and their counterparts of 1977. This potential difference should be considered in the interpretation and generalizability of the present findings.

Second, although we used a systematic approach to identify common ATVs used in the United States, the sample is subject to sampling error and may not be necessarily representative of the models ridden specifically by youth. Moreover, safe and effective riding of utility ATVs involves consideration of factors other than the ability of youth to reach its controls or attain a specific posture. ATVs are rider-active vehicles, which means that riders must be able to shift their body weight to safely perform maneuvers such as turning, negotiating hills, and crossing obstacles (Jennissen et al., 2014; National 4-H Council, 2005). These circumstances warrant further investigation.

Third, we had to determine the absolute location of each control due to feasibility issues. The further-most position was used as the standard position for all controls with gradual adjustment such as the hand gearshift, while pedals were set to resting position.

Fourth, all the human mockups were placed at the ATVs' seat reference point (SRP). This may not be the "best-case" scenario from a reach standpoint since many riders, especially small youth, tend to sit closer to the handlebars (ahead of the SRP) to allow control reaching. However, the SRP is a standardized expected seat position, which allowed for a consistent evaluation approach among the various conditions. The effect of seating adaption to reach controls while riding requires further assessment.

Finally, the reach simulations were performed with static mockups (i.e., we did not evaluate any trunk or hip movement). In real riding situations, riders may shift their hips forward and/or bend their trunks to reach an otherwise unreachable control and perform active riding. However, while active riding can increase the ATV's stability by 10–30 % (Shortland, 2013), there is no clear evidence that active riding and rider separation reduces the risk of rollover for agricultural ATVs specifically (Grzebieta, Rechnitzer, & McIntosh, 2015).This warrants further investigation.

6. Summary

This study evaluated the potential mismatches between youth's anthropometric measures and operational requirements of 17 ATV models. The study's main findings/receommendations were: (1) Most riders failed to pass at least 1 out of the 11 fit criteria for the evaluated vehicles; (2) Youth are not recommended to ride utility ATVs; and (3) Only engine size, maximum speed, and rider's age are insufficient indicators of youth-ATV fit.

The present findings, along with the results of a recent study regarding the forces required for effective ATV operation (De Moura Araujo & Khorsandi, 2020), raise serious questions about the ability of youth to safely operate utility ATVs in common use on U.S. farms and about the validity of current youth-ATV fit guidelines. Therefore, we recommend that the readiness of youth to ride ATVs, especially for occupational purposes, should be carefully evaluated by their parents/guardians. Moreover, we argue that current youth-ATV fit guidelines should be reviewed and updated based on quantitative and systematic data comparing the physical ability of youth and the operational requirements of ATVs.

7. Practical applications

The present study provides such quantitative and systematic data comparing the physical ability of youth and the operational requirements of ATVs. These data support manufacturers in considering design changes or manufacturing new machines and provides critical evidence contributing to the scientific basis for modifying regulatory/advisory guidelines for youth operating utility ATVs.

Acknowledgments

This work was supported by the National Children's Center for Rural and Agricultural Health and Safety (NCCRAHS) under grant No NIOSH 5U54 OH009568. In addition, the Western Center for Agricultural Health and Safety provided a seed grant for conducting preliminary studies.

References

American Academy of Pediatrics. (2018). *ATVs Are Dangerous to Children: New Data Demonstrates Continued Risk of Injury and Death* https://doi.org/https://www.aap.org/en-us/about-the-aap/aap-press-room/Pages/ATVDataCPSC.aspx.

Arunachalam, M., Singh, A. K., & Karmakar, S. (2021). Perceived comfortable posture and optimum riding position of Indian male motorcyclists for short-duration riding of standard motorcycles. *International Journal of Industrial Ergonomics, 83*. https://doi.org/10.1016/j.ergon.2021.103135.

ATV-Safety-Institute. (2009). *ATV E-Course*. ATV Safety Institute Online https://atvsafety.org/atv-ecourse/.

Ayers, P., Conger, J., Comer, R., & Troutt, P. (2018). Stability Analysis of Agricultural Off-Road Vehicles. *Journal of Agricultural Safety and Health, 24*(3), 167–182 https://doi.org/10.13031/jash.12889.

Balthrop, P. M., Nyland, J. A., Roberts, C. S., Wallace, J., Van Zyl, R., & Barber, G. (2007). Orthopedic trauma from recreational all-terrain vehicle use in central Kentucky: A 6-year review. *The Journal of Trauma, 62*(5), 1163–1170. https://doi.org/10.1097/01.ta.0000229814.08289.9a.

Bernard, A. C., Mullineaux, D. R., Auxier, J. T., Forman, J. L., Shapiro, R., & Pienkowski, D. (2010). Pediatric anthropometrics are inconsistent with current guidelines for assessing rider fit on all-terrain vehicles. *Accident; Analysis and Prevention, 42* (4), 1220–1225. https://doi.org/10.1016/j.aap.2010.01.015.

Brison, R. J., Pickett, W., Berg, R. L., Linneman, J., Zentner, J., & Marlenga, B. (2006). Fatal agricultural injuries in preschool children: Risks, injury patterns and strategies for prevention. *Canadian Medical Association Journal, 174*(12), 1723–1726.

Carman, A. B., Gillespie, S., Jones, K., Mackay, J., Wallis, G., & Milosavljevic, S. (2010). All terrain vehicle loss of control events in agriculture: Contribution of pitch, roll and velocity. *Ergonomics, 53*(1), 18–29. https://doi.org/10.1080/00140130903380919.

Cavallo, E., Gorucu, S., & Murphy, D. (2015). Jun 5). Perception of side rollover hazards in a Pennsylvania rural population while operating an all-terrain vehicle (ATV). *Work, 51*(2), 281–288. https://doi.org/10.3233/WOR-141864.

Chang, J. H., Fathallah, F. A., Pickett, W., Miller, B. J., & Marlenga, B. (2010). Limitations in fields of vision for simulated young farm tractor operators. *Ergonomics, 53*(6), 758–766. https://doi.org/10.1080/00140131003671983.

Chou, H.-Y., Khorsandi, F., Vougioukas, S. G., & Fathallah, F. A. (2022). Developing and evaluating an autonomous agricultural all-terrain vehicle for field experimental rollover simulations. *Computers and Electronics in Agriculture, 194*. https://doi.org/10.1016/j.compag.2022.106735.

Clay, L., Hay-Smith, E. J., Treharne, G. J., & Milosavljevic, S. (2015). Unrealistic optimism, fatalism, and risk-taking in New Zealand farmers' descriptions of quad-bike incidents: A directed qualitative content analysis. *Journal of Agromedicine, 20*(1), 11–20. https://doi.org/10.1080/1059924X.2014.976727.

Clay, L., Treharne, G. J., Hay-Smith, E. J. C., & Milosavljevic, S. (2014). Are agricultural quad bike loss-of-control events driven by unrealistic optimism? *Safety Science, 66*, 54–60. https://doi.org/10.1016/j.ssci.2014.02.002.

Concannon, E., Hogan, A., Lowery, A., Ryan, R. S., Khan, W., & Barry, K. (2012). Spectrum of all-terrain vehicle injuries in adults: A case series and review of the literature. *International Journal of Surgery Case Reports, 3*(6), 222–226. https://doi.org/10.1016/j.ijscr.2012.01.012.

CPSC (2006). Standards for All Terrain Vehicles and Ban of Three-Wheeled All Terrain Vehicles; Notice of Proposed Rulemaking. *US Consumer Product Safety Commission, Federal Register, 71*(154), 45904–45962.

David, J.-A. (1998). *Part III: Report on AU-Terrain Vehicle-Related Deaths, Jan. 1, 1985 - Dec. 31, 1996*. https://www.cpsc.gov/s3fs-public/pdfs/3548a5dc.pdf.

De Moura Araujo, G., & Khorsandi, F. (2020). *Ability of Youth Operators to Activate Agricultural All-Terrain Vehicles Controls* 2020 ASABE Annual International Virtual Meeting, July 13-15, 2020,

De Moura Araujo, G., Khorsandi, F., Kabakibo, S., & Kreylos, O. (2021). *Can Youth Reach Agricultural All-Terrain Vehicle Commands?* 2021 ASABE Annual International Virtual Meeting,

Denning, G. M., Harland, K. K., & Jennissen, C. A. (2014). Age-based risk factors for pediatric ATV-related fatalities. *Pediatrics, 134*(6), 1094–1102. https://doi.org/10.1542/peds.2014-1993.

DTI. (1995). *Childata: The Handbook of Child Measurements and Capabilities. Data for Design Safety*. Department of Trade and Industry (URN 95/681).

Fathallah, F. A., Chang, J. H., Berg, R. L., Pickett, W., & Marlenga, B. (2008). Forces required to operate controls on farm tractors: Implications for young operators. *Ergonomics, 51*(7), 1096–1108. https://doi.org/10.1080/00140130801961901.

Fathallah, F. A., Chang, J. H., Pickett, W., & Marlenga, B. (2009). Ability of youth operators to reach farm tractor controls. *Ergonomics, 52*(6), 685–694. https://doi.org/10.1080/00140130802524641.

FReSH. (2012). *Youth ATV safety*. https://ag-safety.extension.org/youth-atv-safety.

Garvey, P. M., Murphy, D. J., Yoder, A. M., & Hilton, J. W. (2008). National safe tractor and machinery operation program: Development and content evaluation. *Journal of agricultural safety and health, 14*(3), 333–349.

General Accounting Office (GAO). (2010). All-terrain vehicles: how they are used, crashes, and sales of adult-sized vehicles for children's use. Report to Congressional Committees (GAO-10-418). http://www.gao.gov/new.items/d10418.pdf.

Grzebieta, R., Rechnitzer, G., & McIntosh, A. S. (2015). *Quad Bike Performance Project (QBPP) Part 3: Rollover Crashworthiness Test Results* (The Quad Bike Performance Project, Issue.

Hard, D. L., & Myers, J. R. (2006). Fatal work-related injuries in the agriculture production sector among youth in the United States, 1992–2002. *Journal of agromedicine, 11*(2), 57–65.

Helmkamp, J. C., Marsh, S. M., & Aitken, M. E. (2011). Occupational All-Terrain Vehicle Deaths among Workers 18 Years and Older in the United States, 1992–2007. *Journal of agricultural safety and health, 17*(2), 147–155. https://doi.org/10.13031/2013.36498.

Hendricks, K. J., & Hard, D. J. (2014). *Working youth on farms*.

Hendricks, K. J., Myers, J. R., Layne, L. A., & Goldcamp, E. M. (2005). Household youth on minority operated farms in the United States, 2000: Exposures to and injuries from work, horses. *ATVs and tractors. J Safety Res, 36*(2), 149–157. https://doi.org/10.1016/j.jsr.2005.01.002.

House, T., Schwebel, D. C., Mullins, S. H., Sutton, A. J., Swearingen, C. J., Bai, S., & Aitken, M. E. (2016, Oct). Video intervention changes parent perception of all-terrain vehicle (ATV) safety for children. *Inj Prev, 22*(5), 328-333. https://doi.org/10.1136/injuryprev-2015-041880.

IPCH. (2018). *Riding the Right ATV* https://doi.org/https://intermountainhealthcare.org/-/media/files/facilities/pch/atv-right-one-sheet.pdf?la=en.

Jennissen, C. A., Miller, N. S., Tang, K., & Denning, G. M. (2014). Optimising seat design for all-terrain vehicle injury prevention: Wide variability illustrates need for evidence-based standardisation. *Injury Prevention, 20*(2), 88–98. https://doi.org/10.1136/injuryprev-2013-040786.

Jennissen, C. A., Wetjen, K., Hoogerwerf, P., O'Donnell, L., & Denning, G. M. (2018). *The All-Terrain Vehicle Exposure and Crash Experiences of Iowa FFA Members National Occupational InjuryResearch Symposium*. WV: Morgantown.

Jinnah, H. A., & Stoneman, Z. (2016). Youth Personality Factors as Predictors of Risky All-Terrain Vehicle (ATV) Driving Behaviors. *Journal of agricultural safety and health, 22*(2), 134–147. https://doi.org/10.13031/jash.22.11330.

Khorsandi, F., Ayers, P., Denning, G., Jennissen, C., Jepsen, D., Myers, M., Oesch, S., Pate, M., & White, D. J. (2021). Agricultural All-Terrain Vehicle Safety: Hazard Control Methods Using the Haddon Matrix. *Journal of Agromedicine, 26*(4), 420–435. https://doi.org/10.1080/1059924X.2020.1837705.

Khorsandi, F., Ayers, P. D., & Fong, E. J. (2019). Evaluation of Crush Protection Devices for agricultural All-Terrain Vehicles. *Biosystems Engineering*. https://doi.org/10.1016/j.biosystemseng.2019.04.025.

Kreylos, O. (2016). Lighthouse tracking examined Retrieved 06/19 from http://doc-ok.org/?p=1478.

Landis, J. R., & Koch, G. G. (1977, Mar). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://www.ncbi.nlm.nih.gov/pubmed/843571.

Lower, T., & Herde, E. (2012). Non-intentional farm injury fatalities in Australia, 2003–2006. *NSW Public Health Bulletin, 23*(1–2), 21–26.

Lower, T., Monaghan, N., & Rolfe, M. (2016). Quads, Farmers 50+ Years of Age, and Safety in Australia. *Safety, 2*(2). https://doi.org/10.3390/safety2020012.

Marlenga, B., Pickett, W., & Berg, R. L. (2001). Assignment of work involving farm tractors to children on North American farms. *American Journal of Industrial Medicine, 40*, 15–22.

McBain-Rigg, K. E., Franklin, R. C., McDonald, G. C., & Knight, S. M. (2014). Why Quad Bike Safety is a Wicked Problem: An Exploratory Study of Attitudes, Perceptions, and Occupational Use of Quad Bikes in Northern Queensland, Australia. *Journal of agricultural safety and health, 20*(1), 33–50. https://doi.org/10.13031/jash.20.10380.

McHugh, M. (2012). Interrater reliability: The kappa statistic.pdf>. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031.
McIntosh, A. S., Patton, D. A., Rechnitzer, G., & Grzebieta, R. (2016). Injury mechanisms in fatal Australian quad bike incidents. *Traffic Injury Prevention, 17*(4), 386–390. https://doi.org/10.1080/15389588.2015.1091073.
Milosavljevic, S., McBride, D. I., Bagheri, N., Vasiljev, R. M., Carman, A. B., Rehn, B., & Moore, D. (2011). Factors associated with quad bike loss of control events in agriculture. *International Journal of Industrial Ergonomics, 41*(3), 317–321. https://doi.org/10.1016/j.ergon.2011.02.010.
Murphy, D. (2020). *National Safe Tractor and Machinery Operation Program*.
National 4-H Council. (2005). *4-H ATV Safety*. Retrieved 04-08 from http://www.atv-youth.org/.
NCCRAHS. (2018). *Operating an All-Terrain Vehicle, Can youth do this job safely?* National Children's Center for Rural and Agricultural Health and Safety (NCCRAHS).
Niehorster, D. C., Li, L., & Lappe, M. (2017, May-Jun). The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research. *Iperception, 8*(3), 2041669517708205. https://doi.org/10.1177/2041669517708205.
Pollack-Nelson, C., Vredenburgh, A. G., Zackowitz, I. B., Kalsher, M. J., & Miller, J. M. (2017). Adult Products That Kill and Injure Children. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*(1), 886–889. https://doi.org/10.1177/1541931213601513.
Reed, M. P., Ebert-Hamilton, S. M., Manary, M. A., Klinich, K. D., & Schneider, L. W. (2005). A New Database of Child Anthropometry and Seated Posture for Automotive Safety. *SAE Transaction, 114*, 2222–2235 https://www.jstor.org/stable/44725253.
Serre, T., Lalys, L., Bartoli, C., Christia-Lotter, A., Leonetti, G., & Brunet, C. (2010). Child pedestrian anthropometry: evaluation of potential impact points during a crash. *Accident; Analysis and Prevention, 42*(6), 1943–1948. https://doi.org/10.1016/j.aap.2010.05.016.
An inquiry into the death of Suzanne Claudia Ferguson, (2013). https://coronialservices.justice.govt.nz/assets/Documents/Decisions/csu-2010-whg-000160-ferguson-suzanne-claudia.pdf.
Snyder, R. G., Schneider, L. W., Owings, C. L., Reynolds, H. M., Golomb, D. H., & Schork, M. A. (1977). *Anthropometry of infants, children, and youths to age 18 for product safety design*. US Government Printing Office.
Thorbole, C., Aitken, M. E., Graham, J., Miller, B. J., & Mullins, S. H. (2012). *Assessment of the dynamic behavior of a single person ATV in presence of a passenger: Outcome on the rider and passenger crash impact kinematics using computational model*. Houston, TX: International Mechanical Engineering Congress and Exposition.
Towner, E., & Mytton, J. (2009). Prevention of unintentional injuries in children. *Paediatrics and Child Health, 19*(11), 517–521. https://doi.org/10.1016/j.paed.2009.10.007.
Voreacos, D. (1987). *Safety Panel Confirms 3-Wheel ATV Ban*. https://www.latimes.com/archives/la-xpm-1987-12-30-mn-21651-story.html.
Weichelt, B., & Gorucu, S. (2018). Supplemental surveillance: a review of 2015 and 2016 agricultural injury data from news reports on AgInjuryNews. org. *Injury prevention*, injuryprev-2017-042671.
Wiener, R. C., Waters, C., Harper, M., Shockey, A. K. T., & Bhandari, R. (2022). All-Terrain Vehicle-Related Emergency Department Visits: Interaction of Sex and Age, NEISS, 2019. *The Journal of Emergency Medicine*. https://doi.org/10.1016/j.jemermed.2022.02.005.

Guilherme De Moura Araujo received his PhD in 2022 from the University of California, Davis in Biological Systems Engineering. He is a postdoctoral researcher in the Department of Biological and Agricultural Engineering at the University of California, Davis in the Safety Ag. lab. His research at UC Davis focuses on applying technology to improve the safety of All-Terrain Vehicle operators.

Farzaneh Khorsandi Kouhanestani received her Ph.D. in 2017 from The University of Tennessee, Knoxville in Biological Systems Engineering. She is an Assistant Professor of Cooperative Extension in the Department of Biological and Agricultural Engineering at the University of California, Davis. Her research focuses on Engineering control methods to improve occupational health and safety among agricultural workers.

Fadi A. Fathallah received his Ph.D. in 1995 from The Ohio State University in Industrial and Systems Engineering. He is a Professor and Chair of the Department of Biological and Agricultural Engineering at the University of California, Davis. His research focuses on ergonomics, biomechanics, and occupational health and safety among agricultural workers.

Journal of Safety Research 84 (2023) 251–260

Contents lists available at ScienceDirect

Journal of Safety Research

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

A comparative study of collision types between automated and conventional vehicles using Bayesian probabilistic inferences

Norris Novat [a,*], Emmanuel Kidando [b], Boniphace Kutela [c], Angela E. Kitali [d]

[a] Graduate Research Assistant, Cleveland State University, 2121 Euclid Avenue, Cleveland, OH 44115, United States
[b] Department of Civil and Environmental Engineering, Cleveland State University, 2121 Euclid Avenue, Cleveland, OH 44115, United States
[c] Roadway Safety Program, Texas A&M Transportation Institute, 1111 RELLIS Parkway, Bryan, TX 77807, United States
[d] School of Engineering and Technology, University of Washington Tacoma, 1900 Commerce Street Tacoma, WA 98402-3100, United States

A R T I C L E I N F O

Article history:
Received 9 March 2022
Received in revised form 1 August 2022
Accepted 1 November 2022
Available online 10 November 2022

Keywords:
Automated vehicles
Traffic crashes severity
Bayesian networks

A B S T R A C T

Introduction: Automated vehicle (AV) technology is a promising technology for improving the efficiency of traffic operations and reducing emissions. This technology has the potential to eliminate human error and significantly improve highway safety. However, little is known about AV safety issues due to limited crash data and relatively fewer AVs on the roadways. This study provides a comparative analysis between AVs and conventional vehicles on the factors leading to different types of collisions. *Method:* A Bayesian Network (BN) fitted using the Markov Chain Monte Carlo (MCMC) was used to achieve the study objective. Four years (2017–2020) of AV and conventional vehicle crash data on California roads were used. The AV crash dataset was acquired from the California Department of Motor Vehicles, while conventional vehicle crashes were obtained from the Transportation Injury Mapping System database. A buffer of 50 feet was used to associate each AV crash and conventional vehicle crash; a total of 127 AV crashes and 865 conventional vehicle crashes were used for analysis. *Results:* Our comparative analysis of the associated features suggests that AVs are 43% more likely to be involved in rear-end crashes. Further, AVs are 16% and 27% less likely to be involved in sideswipe/broadside and other types of collisions (head-on, hitting an object, etc.), respectively, when compared to conventional vehicles. The variables associated with the increased likelihood of rear-end collisions for AVs include signalized intersections and lanes with less than 45 mph speed limit. *Conclusions:* Although AVs are found to improve safety on the road in most types of collisions by limiting human error leading to vehicle crashes, the current state of the technology shows that safety aspects still need improvement.

1. Background

Vehicular crashes occur due to various causes, but most of these causes are human errors (Singh, 2015). Examples of human errors contributing to road crashes include distracted driving, driving under the influence of alcohol and/or drugs, or mere fatigued driving (Goetz & Haleem, 2021; Khan & Habib, 2021; Monyo et al., 2021). Automated vehicle (AV) technology is anticipated to address these human errors by reducing human involvement in controlling the vehicle and therefore improving safety (Favarò et al., 2017). Currently, little is known about AV safety due to limited crash data and relatively fewer AVs on the road. However, strategic efforts are in place to ensure AV safety aspects are evaluated and tested before fully manifesting this technology.

One proposed way is to assess safety by testing AVs in a mixed traffic environment and observing their performance (Kalra & Paddock, 2016). The authors show that AVs would have to be driven hundreds of millions of miles to demonstrate their safety and reliability. This is due to limited miles-driven data and the capacity to test enough AVs on the road. That being the case, manufacturers, transportation officials, and researchers ought to principally come up with more ways of studying the safety extent of AVs, such as the use of statistical comparisons of AVs to human driver performance. Most current AV safety studies use simulations or survey questionnaires (Bansal & Kockelman, 2017; Combs et al., 2019; Arvin et al., 2020; Rahman, Abdel-Aty, & Wu, 2021). Tibljaš et al. (2018) used safety simulation analysis with a surrogate safety simulation model (SSAM) to assess AVs' safety levels at roundabouts. The study found that rear-end crashes are typical in AVs. Similarly,

* Corresponding author.
E-mail addresses: n.novatbaraba@vikes.csuohio.edu (N. Novat), e.kidando@csuohio.edu (E. Kidando), b-kutela@tti.tamu.edu (B. Kutela), akitali@uw.edu (A.E. Kitali).

https://doi.org/10.1016/j.jsr.2022.11.001
0022-4375/

Papadoulis et al. (2019) developed a decision-making algorithm in VISSIM software and used the SSAM for safety analysis. It was found that traffic conflicts were minimized with increased market penetration rates, although rear-end crashes were typical. Furthermore, Arvin et al. (2020) evaluated AVs' safety in a mixed traffic setting at intersections in a simulation framework. The study assessed the volatility of conflicts and crashes at different market penetration rates. It was concluded that the lower the market penetration rates, the minimal the conflicts, and the ideal road safety could be achieved with an optimal market penetration rate of 40%. Although in this study, Arvin et al. (2020) signify the safety importance of AVs over conventional vehicles in a simulation study, validation with actual field data was not incorporated in the analysis.

Few studies have been conducted on AV crashes using field data. Recent application of the publicly available California AV dataset has included researchers conducting mainly exploratory data analysis of the limited crash samples (Favaro et al., 2017; Favaro, Eurich, & Nader, 2018; Boggs, Arvin, & Khattak, 2020; Petrovic, Mijailović, & Pešić, 2020). This is partly due to limited market penetration; therefore, few crash data involving AVs are available. Favaro, Eurich, and Rizvi (2019) utilized the data published by the California Bureau of Motor Vehicles (CA BMV) from 2014 to 2018 to study AV Advanced Driver Assistance Systems potential risks in case of technology failures. This study found that AVs are susceptible to rear-end crashes and observed that most crashes occur at 10 miles per hour (mph). Furthermore, due to the integration of the two-vehicle types, if the penetration rate of AVs rises, the overall number of crashes may also rise.

Despite the difference not being statistically significant, Teoh and Kidd (2017) found that the crash rates of AVs were lower than conventional vehicles. Even though there is no confirmation in the police reports, the authors suspect that because Google cars are regularly impacted from behind, the technology may force the vehicles to brake abruptly. Additionally, Chen et al. (2021) explored the occurrence of AV crashes using machine learning algorithms on the CA DMV crash data. The study found that AVs are more prone to rear-end crashes. It was also established that factors such as vehicle damage, weather conditions, accident location, and driving mode are some of the most critical factors associated with AV crashes. A study by Boggs, Wali, and Khattak (2020) employed text analytics on the AV crash reports and a hierarchical Bayesian regression to examine AV crashes. The authors also found that rear-ended crashes are the most frequent type of collision (61.1%), while injury-resulting crashes are about 13.3%. This study also suggests that the likelihood of rear-end crashes is higher when the AV is engaged than when the AV is disengaged (human-driven). The study also indicated that the AVs have a higher crash propensity in areas of mixed land use settings compared to other land uses, with lower chances observed in public areas and school zones. Kutela, Das, and Dadashova (2022) evaluated non-motorists direct and indirect involvement in AV crashes in California. The study found that bicyclists and scooterists are likely to be directly involved, while pedestrians are likely to be indirectly involved in AV crashes. Even more, Kutela, Avelar, and Bansal (2022) jointly analyzed the associated factors of three interrelated outcome variables vehicle at fault, collision type, and injury outcome in AV-involved crashes. They showed that irrespective of the collision type, when the AV is at fault, the chance of the physical injury in a crash increases significantly. Nonetheless, the above-mentioned studies did not assess injuries and other contributing factors such as intersection signalization and, above all, did not draw out a comparative analysis of conventional vehicles type of collision.

Although studies have attempted to evaluate the safety aspects of AVs using crash data, the pattern of collision type has not been a topic of interest for most researchers. In fact, no study has compared characteristics of AV and conventional vehicle crashes that occurred in the same vicinity and explained the safety benefits of AVs. Furthermore, some road users appear concerned about AV system failure, system breaching/hacking, and safety issues (Bansal, Kockelman, & Singh, 2016). One study found that even though AVs can potentially minimize chances of crash occurrence, be environmentally friendly, timesaving, and increase mobility (Payre, Cestac, & Delhomme, 2014), users "do not feel safe if the car is driving itself," and most parents would not let their kids ride alone in a car with full driving automation (König & Neumayr, 2017). It will be several years before all vehicles on the roadways transition to AV; AVs and conventional vehicles may have different responses in different conditions (e.g., hard braking, deceleration). Nonetheless, limited published research tried to address the concern: What is the safety advantage of AVs when compared to conventional vehicles? What type of collisions are AVs most likely to be involved in? What strategies are needed to improve AVs safety? The current study seeks to answer the question of how well AVs that are currently on the road perform in comparison to conventional vehicles in terms of safety (manner of collision). For comparability purposes, only crashes involving conventional vehicles in a 50 ft buffer zone to crashes involving AVs were considered in the current study, assuming they share common characteristics such as road geometry and surface conditions.

Bayesian Networks (BNs) approach is used to investigate the prominent factors and compare various crash types associated with AVs and conventional vehicles. Furthermore, the study explored whether the predicted probabilities for AVs and conventional vehicles on the manner of collisions are statistically different. This comparison was made using the predicted posterior distributions. The graphical representation of BNs and how it accounts for interdependency between variables makes BNs superior to the classical regression models (Kidando et al., 2019). The study utilizes the AV crash data from California as reported by CA DMV (California DMV, 2021) and conventional vehicle crash data (TIMS, 2021).

The contribution of this study is twofold. First, this study provides a comparative assessment of limitations on road safety issues due to human driving (conventional vehicles) as opposed to AVs. Second, the inferences from the model will help transportation officials, manufacturers, and researchers, especially in the mixed traffic era, quantify the safety expectations of AVs as the technology is being developed and improved. The potential benefit of using BN over other traditional analytical approaches is its capacity to condition several explanatory variables. Moreover, the BN model provides an outstanding interpretation, explicitly presenting the probability relationships among variables in the model. The rest of the manuscript is structured as follows. The next section describes the study data followed by the approach used to geographically merge data from AV and conventional vehicles. Then the descriptive statistics of the merged data are presented, followed by the methodology used to develop BNs, results, and discussion, and finally, the conclusion and recommendation section.

2. Data

This study obtained crash data from CA DMV and the Transportation Injury Mapping System (TIMS). The CA DMV is an open repository, and it was initiated in 2014 to collect all the AVs crash data from manufacturers in California (California DMV, 2021). The AV manufacturers that have reported their vehicle crashes in this database include Google, General Motors, Delphi, Nissan, and Cruise Automation, to mention a few (Boggs et al., 2020; California DMV, 2021). The published data in the CA DMV repository contains a detailed report of each AV crash, including the geolocation of the crashes, type of collision, weather, road surface condition, and date and time of the collision. After auditing the published AVs' crash reports and screening erroneous and inconsistent data reporting, AV crashes that occurred between 2017 and 2020 were found to have consistent reporting and complete data to be used for accurate analysis. The data auditing process involved screening AV crash records that were not in fully autonomous mode; that is, the AV was disengaged, and the driver was in control of the vehicle. This data screening step was vital for the study in order to establish an effective comparison between conventional vehicles and AVs (when not under the influence of a human driver). A total of 127 reported AV crashes were then used for analysis in the selected study period. On the other hand, the TIMS database provides conventional vehicle crash data. This database is hosted and maintained by the Center for Safe Transportation Research and Education at the University of California Berkeley (TIMS, 2021). A total of 865 crash observations were used in this study. This count was obtained by the procedure explained in the merging section discussed in the following section.

2.1. Data merging

The AVs and conventional vehicle crashes were mapped in the geographic information system (GIS) application using geolocation (latitudes and longitudes) of the crashes. Only crashes in close vicinity were compared to perform a fair comparison between AVs and conventional vehicles. It is assumed that the closer the AV and conventional vehicle crashes, the higher the likelihood that the two crashes share similar physical attributes. Thus, a buffer of 50 feet (ft) radius around each AV crash was used to extract all associated conventional vehicle crashes. A similar geo-filtering approach to vehicle crashes has been used by the previous studies to identify traffic volume (i.e., annual average daily traffic (AADT), data within a proximity of 50 ft, as well as using to narrow down to crashes within 50 ft of an intersection; Mammadrahimli, 2015; Estep, Kim, & Nixon, 2016). This procedure revealed a total of 865 conventional vehicle (2017–2020) crashes to be within the vicinity of the 50 feet buffer zone of AV crashes. Fig. 1 presents a GIS map that shows the geo-filtered locations of the conventional vehicle crashes (green) and AV crashes (red) that occurred within the same area.

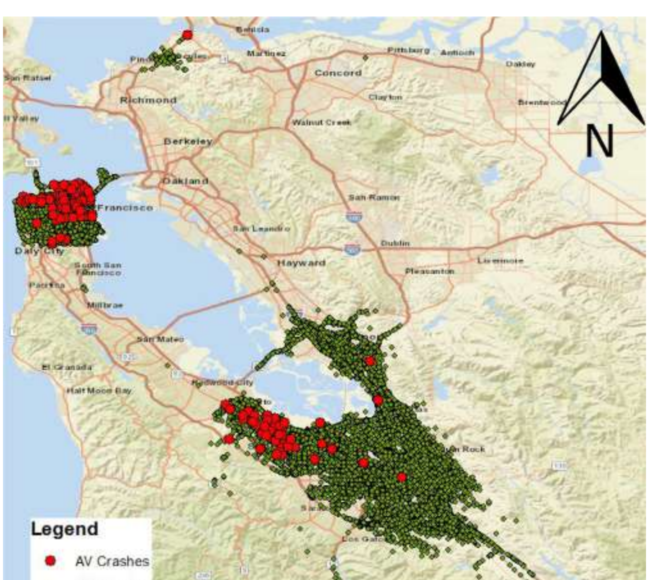


Fig. 1. Map of AVs and conventional vehicle crashes in California.

Based on the literature review, the study explored the following explanatory variables: type of collision, road classification, segment type, crash whether occurred near signalized, unsignalized intersection or segment, number of lanes, land use characteristics, and posted speed limit. These variables were obtained from the Google map.

2.2. Descriptive statistics of the data

Table 1 presents the descriptive statistics of the gathered data. The type of collision is divided into three categories: rear-end crashes, sideswipe/broadside crashes, and other types of crashes, composed of crashes that were infrequent to AV collisions (i.e., head-on, overturned, hit an object, pedestrian, etc.). The distribution for rear-end, sideswipe/broadside, and other crashes are 67%, 28%, and 6%, respectively, for AVs and 20%, 43%, and 37%, respectively, for conventional vehicles.

Table 1 also shows a distribution between AV and conventional vehicle crashes under different highway functional classifications. Many crashes occurred on local and collector roads, followed by arterial and fewer crashes on freeways/expressways for both conventional vehicles and AVs. Given the number of lanes on the corridor, crash occurrence appears to be increasing with an increase in the number of lanes for AVs (i.e., one lane, 29%; and two or more lanes, 71%), and conventional vehicles (i.e., one lane, 19%; and two or more lanes, 81%).

3. Methodology

This study developed a BN to establish a comparative analysis between AVs and conventional vehicles on the manner of collisions. The BN is one of the hybrid artificial intelligence (data + human knowledge) models, which expresses a probabilistic relationship between variables (Mittal & Kassim, 2007; Sheehan et al., 2019). The nodes are connected by links known as edges or sometimes referred to as arcs. The edge direction originates from the parent node to a child node, and the relationship between the two nodes is defined by conditional probability (Kidando et al., 2019; Kutela & Teng, 2019). The best network structure of the BNs in this study was established from learning the dataset using a scoring search algorithm and expert's knowledge to elucidate meaningful node connections of the BN. The search algorithm adopted is based on greedy hill climbing with Akaike Information Criterion (AIC) as a scoring function. This search algorithm involves an iteration process of adding and reversing the nodes and edges until the best network is obtained. The AIC scoring function used to establish the optimal network structure is presented in Equation (1).

$$AIC = 2 * L + 2 * n \tag{1}$$

where, L is the maximized log-likelihood; n is the number of parameters in a Bayesian Network.

3.1. Probabilistic reasoning and comparison

Once the optimal network structure was established, the predictive inference was conducted to evaluate the cause-effect relationship between the target (child) and parent variables. The predictive inference can also be referred to as the sensitivity analysis and is conducted by assigning evidence in the network structure and evaluating the impact on the target variable (e.g., collision type). Since the focus of this study is to compare the AV and CV on the manner of collisions, the probabilistic inference was computed following the two approaches. The first approach was to investigate individual evidence of AV and conventional vehicles on the

Table 1
Descriptive analysis of variables.

Variable	Category	Conventional Vehicles		AVs	
		Count	%	Count	%
Collision type	Rear end	171	20	85	67
	Sideswipe	376	43	35	28
	Others	318	37	7	6
Road classification	Collector and local	603	70	82	65
	Arterial	189	22	29	23
	Freeway/Expressway	73	8	16	13
Segment type	Non-intersection	22	3	13	7
	Unsignalized intersection	74	9	38	20
	Signalized intersection	769	89	141	73
Number of lanes	One lane	166	19	56	29
	Two or more lanes	699	81	136	71
Land use	Residential	154	18	36	28
	Commercial	448	52	50	39
	Mixed land use	263	30	41	32
Speed limit	Less than 45 mph	708	82	99	78
	45 mph or higher	157	18	28	22

outcome. This was conducted by setting certainty on a particular vehicle category (conventional vehicle or AV) in the BNs and observing the impact on the manner of collision. Equation 5 describes how the inference on the first approach was made in the analysis. In this equation, i is the probability of a certain type of crash and x represents the evidence of a hypothesis variable x.

$$P(type\ \ of\ \ collision = i|Evidence_x = 1) \tag{2}$$

On the second approach, the inference was made by evaluating the combined impact of conventional vehicles or AV with other hypothesis variables in predicting the manner of collisions. The combined evidence inference involves setting multiple pieces of evidence and predicting the probability of a target variable. This type of analysis can be used to reveal how, for example, the AV and segment type is associated with the likelihood of rear-end crashes. Herein, the hypothesis variable is referred to as the parent node, which has a direct connection with the target variable in the optimal network. The analysis was done using Equation (3) whereby the e_n is the evidence of a hypothesis variable n, and h_{vn} is the observed evidence category, v of hypothesis variable n. As with the individual evidence analysis, for the combined evidence, each observed evidence was assigned a certainty value of 1.

$$P(type\ \ of\ \ collison = i|e_1 = hv_1, e = hv_2 \cdots .e_n = h_{vn}) \tag{3}$$

Generally, the BN parameters, which are also referred to as conditional probabilities, are estimated using the maximum likelihood approach. While the BN gives us the estimated parameters, it does not account for the uncertainty, and it is difficult to estimate the confidence interval of the estimates. This study used a Bayesian framework based on the Markov Chain Monte Carlo (MCMC) simulations to address these limitations. The simulations were fitted using an open-source Python package called NumPyro version 0.7.2, developed by Uber AI Labs (Bingham et al., 2019; Phan, Pradhan, & Jankowiak, 2019). As with Bayesian inference, the distributions of the inferences were assumed to follow a non-informative prior. For variables with two groups, a Beta distribution was used to calibrate the probabilities, while with more than two groups, the Dirichlet distribution was adopted in the analysis. During sampling, the No-U-Turn Sampler (NUTS) with three chains was applied to calibrate the parameters of all variables in the network. The initial burn-in phases were set to 2,000 iterations, and subsequently, 2,000 iterations were used for inference. The model convergence was evaluated using the Gelman-Rubin diagnostic statistic. For a model to achieve convergence, the difference between chain variances, which is the Gelman-Rubin diagnostic statistic, must be equal to 1 (Makowski et al., 2019). Moreover, visual diagnostics based on the autocorrelation plot, density, and trace plot of each inference were also produced for visual interpretation.

4. Results and discussion

Fig. 2 shows the optimized Bayesian network structure. From this network, four variables had a direct dependence on the collision type: the *number of lanes*, *segment type*, *speed limit*, and *vehicle type*. These variables are also referred to as hypothesis variables. Other variables, including road classification and land use variables, were observed to influence collision type through the hypothesis variables. Sensitivity analysis and comparison of AVs and conventional vehicles were conducted using the optimal BN shown in Fig. 2.

4.1. Individual evidence prediction inference

This analysis was conducted to determine the probability of a specific crash type given a particular vehicle type (i.e., crashes involving AVs and those involving conventional vehicles). The probabilities were estimated by querying the optimal BN and setting the evidence probability to 1 on the vehicle type. Fig. 3 shows the optimal BN with the predicted average probabilities for different types of collisions involving AVs and conventional vehicles. The figure also shows the 95% highest density interval (HDI) and standard deviation (SD) of the predicted probabilities. The HDI and SD summarize the uncertainty of an estimate by indicating the range of most probable values. It is worth understanding that this study only compared different crash types involving AVs and conventional vehicles. It is also possible to conduct a sensitivity analysis and comparison of any node in the network and evaluate the probability change on the target variable.

The results in Fig. 3 show that AV-involved rear-end crashes are more prevalent than rear-end crashes involving conventional vehicles and other road users. Specifically, given that a collision involves an AV, the probability of that crash being a rear end is 0.64 (95% HDI [0.56, 0.71]). On the other hand, the probability of a collision involving a conventional vehicle being a rear end was 0.20 (95% HDI [0.18, 0.23]). Previous studies also reported a similar finding (Boggs, Wali, & Khattak, 2020; Wang et al., 2020; Kutela, Das, & Dadashova, 2022). Sideswipes and other crashes involving

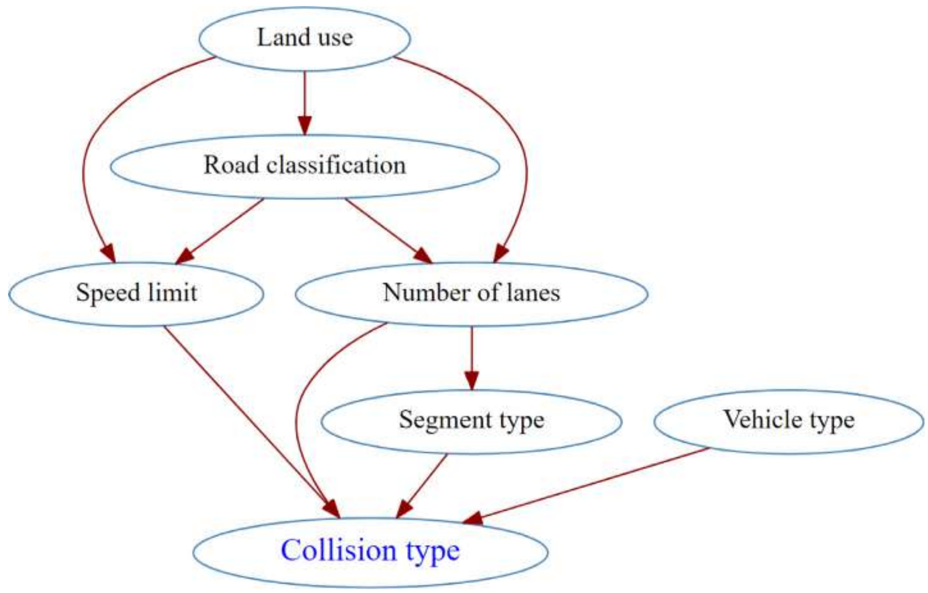


Fig. 2. Optimized BN structure.

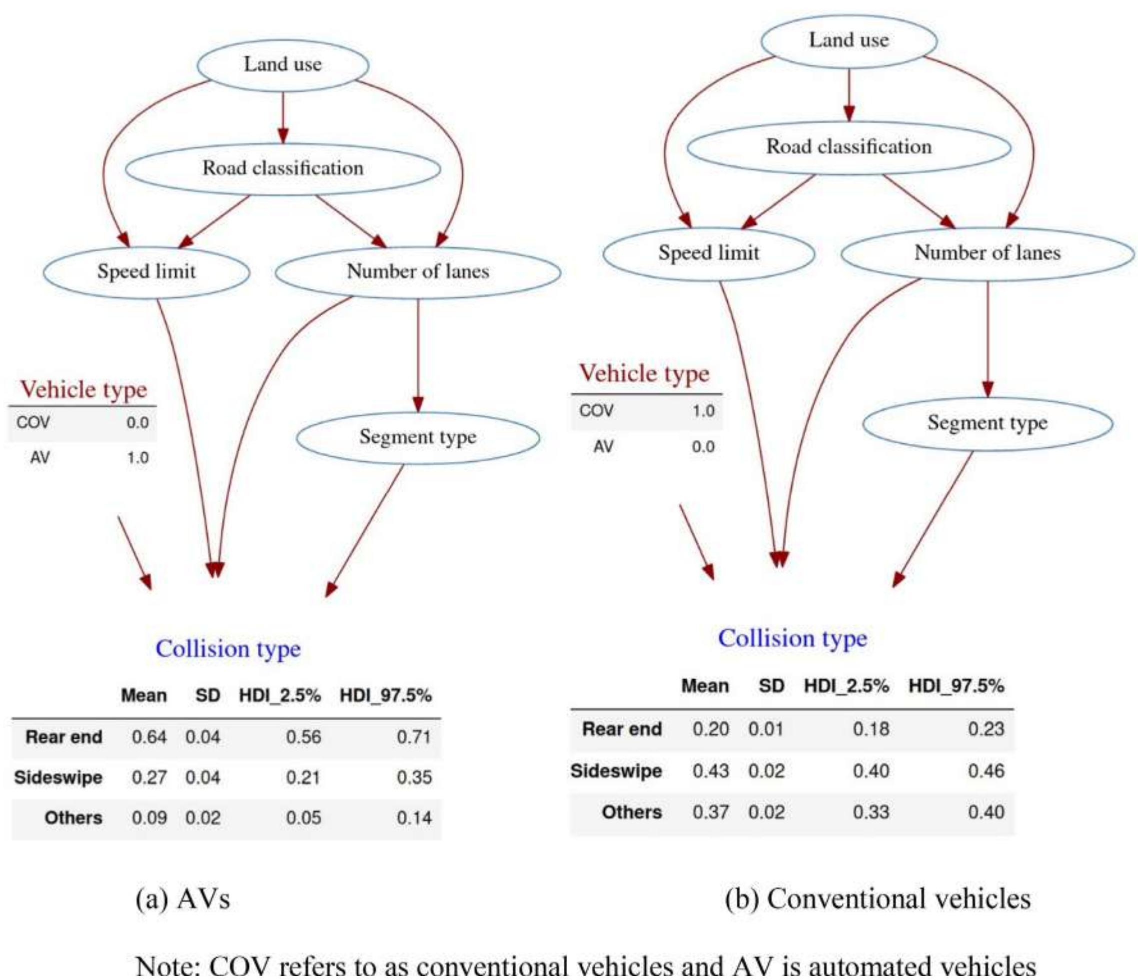


Fig. 3. Sensitivity analysis results for vehicle type, Note: COV refers to as conventional vehicles and AV is automated vehicles.

conventional vehicles were more prevalent than AV-involved sideswipes and other crashes.

Taking advantage of rich information produced by the Bayesian framework, the study further investigated whether the predicted probabilities for AVs and conventional vehicles on the manner of collisions are statistically different. This comparison was made using the predicted posterior distributions. In particular, the posterior indices were established by computing the posterior difference between the conventional vehicles' and AVs' probabilities of the same collision type. The null hypothesis for comparison was framed as the difference between the two fitted probabilities is the same. In contrast, the alternative hypothesis was that the probabilities between AVs and conventional vehicles are credibly different. Several previous studies have adopted this type of analysis to conduct the Bayesian hypothesis test or group comparison (Kruschke, 2011, 2013; Kidando, Moses, & Sando, 2019). Unlike the conventional maximum likelihood hypothesis testing, which relies on the p-value, Bayesian hypothesis testing describes the estimates in probability terms, accounts for uncertainty, and is not significantly affected by the data variations (Kruschke, 2011, 2013; Kidando, Moses, & Sando, 2019).

Fig. 4 shows the indices' posterior distribution (differences) for each collision type along with the mean difference and the 95% HDI represented by the horizontal line near the x-axis. The mean difference and 95% HDI are summary statistics, which facilitate concluding the null and alternative hypothesis. The mean credible values represent the best guess of the actual difference, and the 95% HDI represents the range where the actual difference has 95% credibility. When the difference in the posterior distributions is strictly positive or negative in the 95% HDI, the difference is considered credible at 95% HDI (Kruschke, 2013; Kidando, Moses, & Sando, 2019). This indicates that the limits of the 95% HDI, which are 2.5% and 97.5%, do not include zero as one of the credible values. Therefore, we reject the null hypothesis that the predicted probabilities are the same in favor of the alternative hypothesis. On the other hand, if the estimated HDI line in Fig. 4 crosses zero, the difference of zero is one of the credible values, and we fail to reject the null hypothesis that the conventional vehicles and AVs probabilities are statistically the same.

As shown in Fig. 4, the average probability difference between conventional vehicles and AVs is −0.43, with 95% HDI limits of −0.52 and −0.36 for rear-end crashes. The mean difference in probabilities for sideswipe is 0.16 with 95% HDI of 0.08 and 0.23. For other group types of collision, the estimated mean difference is estimated to be 0.27 with HDI limits of 0.22 and 0.33. The fitted 95% HDI limits for the differences are all positive (sideswipe and others) or negative (rear-end collisions) for all types of collision. Therefore, zero is not one of the credible values in the calculated posterior distributions. These results indicate that the estimated differences in probabilities between conventional vehicles and AVs among collision types are significant at 95% HDI.

4.2. Combined evidence prediction inference

In addition to assessing the influence of vehicle type alone on the predicted probabilities of collision type, the impact of concur-

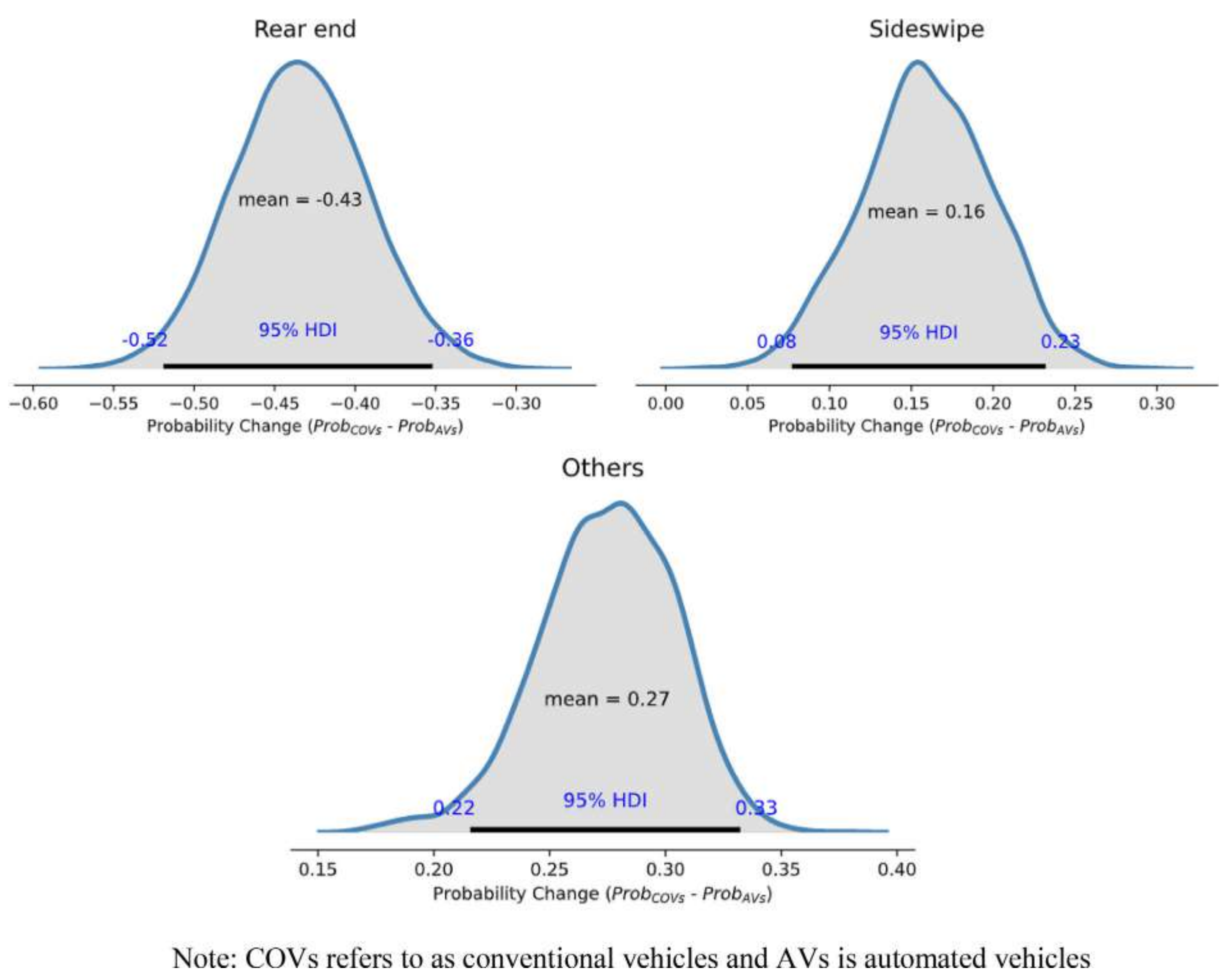


Fig. 4. Comparison of predicted probabilities for AVs and conventional vehicles on the manner of collision, Note: COVs refers to as conventional vehicles and AVs is automated vehicles.

rent evidence was assessed. The combined evidence analysis was conducted by setting evidence of concurrent variables, which are vehicle type and each hypothesis variable. The results of this analysis are presented in Table 2.

The segment type was grouped into three categories: non-intersection (segment), unsignalized, and signalized intersections. As indicated in Table 2, AV-involved-rear-end crashes were more prevalent in a segment than crashes involving conventional vehicles. A similar pattern was observed on the predicted probabilities for signalized and unsignalized intersections. The probabilities change between conventional vehicles and AVs for a rear-end collision are −0.42 (95% HDI [−0.61, −0.24]) and −0.44 (95% HDI [−0.54, −0.36]) for unsignalized and signalized intersection, respectively. These findings suggest that the large probability difference is at intersections compared to a segment for rear-end crashes. In fact, the probability difference for the segment was found insignificant at a 95% credible interval because the HDI crosses zero, which implies that zero is one of the credible values.

A comparison of predicted probabilities on each vehicle type reveals that rear-end collisions are more common at an intersection than a segment for AVs. This finding is similar to previous studies (Wang & Abdel-Aty, 2006; Pande, Abdel-Aty, & Das, 2010). Moreover, at signalized intersection level, AVs are expected to be more alert and respond to red-light stopping signal status, as opposed to human drivers who could delay stopping or sometimes deliberately run the red lights. Hence, this results in a higher likelihood for rear-end collision for AVs at signalized intersections as opposed to unsignalized intersections. On the other hand, the conventional vehicles estimated probabilities reveal a contradicting finding on the comparison of different segment types. The highest probability is recorded on the segment (mean = 0.27), followed by signalized intersection (mean = 0.21), and the last is unsignalized intersection (mean = 0.15).

The comparison of sideswipe type of a collision presented in Table 2 shows that the AV's probability is lower by 0.15 (95% HDI [−0.12, 0.40]) compared to conventional vehicles at a segment. For unsignalized intersections, the likelihood for a sideswipe collision is lower by 0.14 (95% HDI [−0.05, 0.33]) for AVs compared to conventional vehicles. The difference between AVs' and conventional vehicles' predicted probabilities are insignificant at 95% HDI, suggesting that the estimated probabilities are statistically the same. On the other hand, for signalized intersection the estimated difference is statistically different (mean = 0.16, 95% HDI [−0.07, 0.24]). Notably, the difference in likelihoods for AVs over conventional vehicles sideswipe type of collision is not that significant at the segment, unsignalized, or signalized intersection. The observed low likelihood for AVs in sideswipe crashes is due to advanced object detection features surrounding the vehicles, as well as advanced lane changing mechanisms (Liu et al., 2019; Rahman et al., 2019).

Moreover, the predicted probabilities on each vehicle type across different segment types indicate that sideswipe collisions are more common at an intersection than a segment for both conventional vehicles and AVs. Studies by Polders et al. (2015) and Alarifi, Abdel-Aty, and Lee (2018) also presented similar findings for the high likelihood for crashes on the side of the vehicle to be more apparent at intersections level than segment level due to frequent turning movements that lead to more conflicts.

The group of others (i.e., head-on, hitting an object) shows the predicted probability is lower by 0.02 (95% HDI [−0.24, 0.26]) for AVs in comparison to conventional vehicles at the segment level. For unsignalized intersections, the likelihood is lower by 0.29 (95% HDI [0.13, 0.44]) for AVs compared to conventional vehicles, while at signalized intersections, the likelihood is lower by 0.28 (95% HDI [0.22, 0.34]) for AVs compared to conventional vehicles. It is seen that, compared to conventional vehicles, AVs are less likely to be involved in a collision from the front of the vehicle, such as hitting an object at segment level than at intersection level (both unsignalized and signalized). One reason is due to unimpeded object detection at the segment level compared to the required multiple object detection at the intersection level (Aycard et al., 2011).

The number of lanes variable was grouped into highways with one lane and more than one lane. For collisions in one lane corridor, it was observed that AVs were 0.48 (95% HDI [−0.58, −0.36]) more likely to be involved in rear-end collisions than con-

Table 2
Probability of type of collision for AVs and conventional vehicle crashes BN MCMC inferences.

		AVs				**Conventional Vehicles**				**Difference in Probabilities**		
Hypothesis Variable	Target Variable (Collision type)	Mean	SD	HDI		Mean	SD	HDI		Mean	HDI	
				2.5%	97.5%			2.5%	97.5%		2.5%	97.5%
Segment type	Rear end	0.43	0.11	0.21	0.64	0.27	0.07	0.14	0.42	−0.16	−0.43	0.09
Non intersection	Sideswipe	0.26	0.11	0.08	0.47	0.41	0.08	0.25	0.57	0.15	−0.12	0.40
	Others	0.31	0.11	0.12	0.52	0.32	0.08	0.18	0.48	0.02	−0.24	0.26
Unsignalized intersection	Rear end	0.58	0.09	0.41	0.75	0.15	0.04	0.08	0.23	**−0.42**	**−0.61**	**−0.24**
	Sideswipe	0.30	0.08	0.16	0.47	0.44	0.05	0.34	0.55	0.14	−0.05	0.33
	Others	0.12	0.06	0.02	0.22	0.40	0.05	0.31	0.51	**0.29**	**0.13**	**0.44**
Signalized intersection	Rear end	0.65	0.04	0.56	0.73	0.21	0.02	0.17	0.24	**−0.44**	**−0.54**	**−0.36**
	Sideswipe	0.27	0.04	0.19	0.35	0.43	0.02	0.40	0.47	**0.16**	**0.07**	**0.24**
	Others	0.08	0.03	0.03	0.13	0.36	0.02	0.33	0.40	**0.28**	**0.22**	**0.34**
Number of Lanes	Rear end	0.62	0.06	0.51	0.72	0.14	0.02	0.11	0.18	**−0.48**	**−0.58**	**−0.36**
One lane	Sideswipe	0.27	0.05	0.16	0.36	0.42	0.03	0.37	0.46	**0.14**	**0.03**	**0.25**
	Others	0.10	0.04	0.04	0.17	0.44	0.03	0.39	0.49	**0.33**	**0.25**	**0.42**
Two or more lanes	Rear end	0.65	0.05	0.54	0.74	0.25	0.02	0.21	0.29	**−0.40**	**−0.50**	**−0.27**
	Sideswipe	0.27	0.05	0.18	0.37	0.44	0.02	0.39	0.48	**0.17**	**0.06**	**0.27**
	Others	0.08	0.03	0.02	0.14	0.31	0.02	0.27	0.35	**0.23**	**0.15**	**0.30**
Speed limit	Rear end	0.67	0.05	0.59	0.76	0.21	0.02	0.28	0.23	**−0.47**	**−0.56**	**−0.37**
Less than 45 mph	Sideswipe	0.23	0.04	0.16	0.32	0.42	0.02	−0.29	0.46	**0.19**	**0.10**	**0.27**
	Others	0.09	0.03	0.04	0.15	0.37	0.02	0.04	0.41	**0.28**	**0.21**	**0.34**
	Rear end	0.48	0.08	0.34	0.64	0.19	0.03	0.25	0.25	−0.29	−0.46	−0.14
45 mph or higher	Sideswipe	0.43	0.08	0.27	0.58	0.47	0.04	0.39	0.54	0.04	−0.12	0.21
	Others	0.09	0.04	0.02	0.16	0.34	0.04	0.27	0.41	**0.25**	**0.14**	**0.35**

Note: The differences between conventional vehicles and AVs were computed as $P_{x_i = \text{Conventional vehicles}}$ - $P_{x_i = AV}$; HDI = 95% high-density interval; the numbers in bold represent the differences that are statistically significant/credible at 95% HDI.

ventional vehicles. While, for sideswipe collisions, AVs are 0.14 (95% HDI [0.03, 0.25]) less likely in comparison to conventional vehicles. For other types of collisions, AVs are 0.33 (95% HDI [0.25, 0.42]) less likely compared to conventional vehicles. Sideswipe collisions are less likely to occur due to limited vehicle passing and lane-changing movement (Bakhit, Osman, & Ishak, 2017; Rista et al., 2017; Wang et al., 2020).

Additionally, for collisions occurring on two or more lanes corridors, AVs were 0.40 (95% HDI [−0.50, −0.27]) more likely to be involved in rear-end collisions than conventional vehicles. While, for sideswipe collisions, AVs are 0.17 (95% HDI [0.06, 0.27]) less likely when compared to conventional vehicles. For other types of collisions, AVs are 0.23 (95% HDI [0.15, 0.30]) less likely compared to conventional vehicles. The propensity for rear-end crashes on a single-lane road (0.48) is higher than for two-lane roads (0.40). Similarly, Rista et al. (2017) also found that rear-end crash frequencies tend to decrease with increased lane width and the number of lanes. This is due to improved room for lane changing and maneuverability to avoid a rear-end crash.

The speed limit variable had two categories: corridors with a speed limit of 45 mph and higher and corridors with a speed limit lower than 45 mph. The likelihood for a rear-end collision is higher by 0.47 (95% HDI [−0.56, −0.37]) for AVs compared to conventional vehicles on corridors with a speed limit lower than 45 mph. For corridors with a speed limit of 45 mph and higher, the likelihood for a rear-end collision is higher by 0.29 (95% HDI [−0.46, −0.14]) for AVs compared to conventional vehicles. This can be explained by the fact that when AVs travel at a slower speed than the speed limit, they are more prone to suffer rear-end collisions by conventional vehicles. This is contrary to the findings by Wang and Abdel-Aty (2006) whereby they observed a high number of rear-end collisions at areas with high-speed limits, although the study was based on conventional drivers only. Also, Dowling et al. (2016) argue that traveling under the speed limit creates queuing conditions and presents significant safety concerns, particularly with the increased potential for rear-end collisions.

On the contrary, the likelihood of sideswipe collision is lower by 0.19 (95% HDI [0.10, 0.27]) for AVs compared to conventional vehicles when the speed limit is lower than 45 mph. In contrast, the likelihood of sideswipe collision is lower by 0.04 (95% HDI [−0.12, 0.21]) for AVs compared to conventional vehicles when the speed limit is 45 mph or higher. The likelihood of a sideswipe collision when traveling on corridors with a speed limit of 45 mph and higher is higher compared to corridors with a speed limit of less than 45 mph due to loss of traction, reduced lateral control, and higher propensity to slip sideways (Duncan, Khattak, & Council, 1998). In addition to that, the likelihood for another type of collision is lower by 0.28 (95% HDI [0.21, 0.34]) for AVs compared to conventional vehicles when the speed limit is less than 45 mph. Moreover, the other type of collision is lower by 0.25 (95% HDI [0.14, 0.35]) when the speed limit is 45 mph or higher for AVs compared to conventional vehicles.

5. Conclusions and recommendations

Automated vehicle (AVs) technology has the potential to mitigate crashes associated with human behavior. Such behavior includes careless driving, over speeding, or driving under the influence of alcohol and/or drugs and distractions. Several studies have been conducted to understand the safety implications of AVs. Although there has been observed safety improvement associated with the introduction of AVs, less is known on the pattern changes in the manner of collision. This is especially important at a time with mixed conventional vehicles and AVs on the road. Understanding the changes in the pattern and manner of collisions would further improve the safety aspects of AVs by improving operations and mitigating the factors associated with the identified collision type.

This study presented a comparative analysis of the manner of vehicle collisions between AVs and conventional vehicles using CA BMV data. The study used AV and conventional vehicle crashes that occurred within the same vicinity of the 50ft buffer zone. The Bayesian Networks (BN) model was adopted to investigate prominent factors associated with different manners of collision for both AVs and conventional vehicles. The inferences were drawn from the trained optimal network structure. The results indicate that the patterns of the manner of collision for AVs and conventional vehicles differ significantly. That is, from the predicted probabilities, compared to conventional vehicles, AVs are 43% more likely to be involved in rear-end crashes while being 16% and 27% less likely to be involved in sideswipe/broadside and other types of collisions. Furthermore, the study explored whether the predicted probabilities for AVs and conventional vehicles on the manner of collisions are statistically different. This comparison was made using the predicted posterior distributions. The results also indicated that AV-involved rear-end crashes are more prevalent than rear-end crashes involving conventional vehicles and other road users.

Further, among other findings, the number of lanes, intersection signalization, speed limit, and vehicle type variables are influential factors for a certain type of collision. Rear-end collisions are more common at an intersection than a segment for AVs. The comparison of the sideswipe type of a collision shows that for AVs the mean credible value of probability is lower by 0.15 compared to conventional vehicles at a segment. For unsignalized intersections, the mean credible values likelihood of a sideswipe collision is lower by 0.14. The group of others (i.e., head-on, hitting an object) shows the predicted mean credible values probability is lower by 0.02 for AVs compared to conventional vehicles at the segment level. For unsignalized intersections, the mean credible values likelihood is lower by 0.29 for AVs compared to conventional vehicles, while for signalized intersections, the mean credible values likelihood is lower by 0.28 for AVs compared to conventional vehicles. Compared to conventional vehicles, AVs are less likely to be involved in a collision from the front of the vehicle, such as hitting an object at the segment level, than at the intersection level.

The study findings suggest that although AVs are expected to improve safety on the road by limiting human error leading to vehicle crashes, rear-end collisions are expected to increase with the current state of the technology. This is because AVs and conventional vehicles differ significantly in terms of operations. AVs are designed to react to any unfamiliar/unconventional outside environment (e.g., pedestrians waiting to cross a roadway at the crosswalk). On the other hand, conventional vehicle operations depend on the driver's judgments. Thus, the AV acceleration and deceleration behaviors do not cope with those of conventional vehicles. As human drivers are naturally not accustomed to a mixed traffic operation with AVs, this study calls for awareness creation programs by manufacturers, transportation officers, and planners. Such programs will hasten the adaptation of a harmonious mixed traffic environment and minimize the observed high risk for rear-end crash susceptibility of AVs.

Even so, this research suggests that the current state of AV technology in improving road safety can be maximized by modifying AVs' heavy braking intervention before a crash occurs. The manner of collisions that proved difficult for the AV technology can be handled by having controller area network (CAN) information readily available to the AV. Such CAN information includes steering wheel angle, real-time object detection, maneuverability, and so forth, so that object driver intent can be detected and interpreted sooner before a crash occurs. Moreover, improving availability of accurate real-world geo-positioning information to the AV by incorporating accurate satellite positioning data will allow the technology to be more versatile for rear-end, and near-collinear head-on crashes. This is because low approach angles that are common in such crashes are currently ignored to minimize false detection due to positioning uncertainty.

As a follow-up study, we aim to collect more field data for AV crashes and use the model to make inferences with a greater sample size and higher market penetration of AVs on the road. Furthermore, we plan to include other variables such as environmental roadway surface conditions and weather conditions and study how they differ in contributing to vehicle crashes. This model can also be employed in future studies to compare inferences of crashes per mile of conventional vehicles against AVs using mileage data reported to the CA DMV. The results of this study are expected to be used for deriving valuable policies and coming up with innovative and safer AV in-vehicle warning (e.g., rear-end collision warning) technologies to fully exploit the benefits of AVs. Nevertheless, this experiment has limitations including, a prior assumption to this study is that there was no spatial–temporal variation of factors such as land use, geometrical features, etc., during the study period (2017–2020) that could affect the findings of this study. Additionally, the data used were for selected roadbeds in California. Although the methodology can produce inference to other areas of study, the findings of this study may not necessarily be replicated in another area. Nevertheless, the full market penetration of AVs will not be accomplished in the near future. Thus, at the moment it is difficult to incorporate an effective full-scale field safety experiment with V2V communication.

Acknowledgments

The authors are grateful to the California Department of motor vehicles CA DMV for publishing the Automated Vehicle crashes and the Safe transportation research and education center at the University of California Berkeley (TIMS) for hosting the Conventional Vehicle crashes data. Any opinions, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the CAV manufacturers.

Disclosure statement

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

Authors' contribution statement

The authors' contributions to the preparation of this manuscript are as follows: study conception and design: Norris Novat; Boniphace Kutela; Emmanuel Kidando data collection: Boniphace Kutela; Norris Novat analysis and interpretation of results: Emmanuel Kidando, Norris Novat, Boniphace Kutela draft manuscript preparation: Norris Novat, Emmanuel Kidando, Boniphace Kutela, and Angela Kitali. All authors reviewed the results and approved the final version of the manuscript.

Data availability statement

Some or all data used during the study are available in a repository at the University of California Berkeley (TIMS, 2021) and the California Department of Motor Vehicles CA DMV (California DMV, 2021). The code and the model used in this study are available and can be requested from the authors.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Alarifi, S. A., Abdel-Aty, M., & Lee, J. (2018). A Bayesian multivariate hierarchical spatial joint model for predicting crash counts by crash type at intersections and segments along corridors. *Accident Analysis & Prevention, 119*, 263–273. https://doi.org/10.1016/J.AAP.2018.07.026.

Arvin, R. et al. (2020). Safety evaluation of connected and automated vehicles in mixed traffic with conventional vehicles at intersections. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 25*(2), 170–187. https://doi.org/10.1080/15472450.2020.1834392.

Aycard, O. et al. (2011). Intersection safety using lidar and stereo vision sensors. *IEEE Intelligent Vehicles Symposium, Proceedings*, 863–869. https://doi.org/10.1109/IVS.2011.5940518.

Bakhit, P. R., Osman, O. A., & Ishak, S. (2017). Detecting imminent lane change maneuvers in connected vehicle environments. *Transportation Research Record, 2645*(1), 168–175. https://doi.org/10.3141/2645-18.

Bansal, P., & Kockelman, K. M. (2017). Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies. *Transportation Research Part A: Policy and Practice, 95*, 49–63. https://doi.org/10.1016/j.tra.2016.10.013.

Bansal, P., Kockelman, K. M., & Singh, A. (2016). Assessing public opinions of and interest in new vehicle technologies: An Austin perspective. *Transportation Research Part C: Emerging Technologies, 67*, 1–14. https://doi.org/10.1016/J.TRC.2016.01.019.

Bingham, E. et al. (2019). Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research, 20*.

Boggs, A. et al. (2020a). *Exploratory analysis of automated vehicle crashes in California: A text analytics & hierarchical Bayesian heterogeneity-based approach*. Elsevier [Preprint].

Boggs, A. M., Arvin, R., & Khattak, A. J. (2020b). Exploring the who, what, when, where, and why of automated vehicle disengagements. *Accident Analysis and Prevention, 136*(12). https://doi.org/10.1016/j.aap.2019.105406.

Boggs, A. M., Wali, B., & Khattak, A. J. (2020c). Exploratory analysis of automated vehicle crashes in California: A text analytics & hierarchical Bayesian heterogeneity-based approach. *Accident Analysis and Prevention, 135*(October 2019). https://doi.org/10.1016/j.aap.2019.105354.

California DMV (2021) *Autonomous Vehicles - California DMV*. Available at: https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/ (Accessed: June 26, 2021).

Chen, H. et al. (2021). Exploring the mechanism of crashes with autonomous vehicles using machine learning. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2021/5524356.

Combs, T. S. et al. (2019). Automated Vehicles and Pedestrian Safety: Exploring the Promise and Limits of Pedestrian Detection. *American Journal of Preventive Medicine, 56*(1), 1–7. https://doi.org/10.1016/j.amepre.2018.06.024.

Deluka Tibljaš, A. *et al.* (2018). Introduction of Autonomous Vehicles: Roundabouts Design and Safety Performance Evaluation, *mdpi.com* [Preprint]. Available at: https://doi.org/10.3390/su10041060.

Dowling, R. et al. (2016). Performance benefits of connected vehicles for implementing speed harmonization. *Transportation Research Procedia, 15*, 459–470. https://doi.org/10.1016/J.TRPRO.2016.06.039.

Duncan, C. S., Khattak, A. J., & Council, F. M. (1998). Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions. *Transportation Research Record, 1635*, 63–71. https://doi.org/10.3141/1635-09.

Estep, E., Kim, K. and Nixon, D. (2016). Laboratory Exercise — Safety on Roads, *Engineering Applications of GIS - Laboratory Exercises* [Preprint].

Favarò, F., Eurich, S., & Nader, N. (2018). Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations. *Accident Analysis and Prevention, 110*(October 2017), 136–148. https://doi.org/10.1016/j.aap.2017.11.001.

Favarò, F. M. et al. (2017). Examining accident reports involving autonomous vehicles in California. *PLoS ONE, 12*(9). https://doi.org/10.1371/journal.pone.0184952.

Favaro, F. M., Eurich, S. O., & Nader, N. (2018). Analysis of disengagements in autonomous vehicle technology. In *Proceedings - Annual Reliability and Maintainability Symposium, 2018-Janua*. https://doi.org/10.1109/RAM.2018.8463123.

Favarò, F. M., Eurich, S. O., & Rizvi, S. S. (2019). 'Human' problems in semi-autonomous vehicles: Understanding drivers' reactions to off-nominal scenarios. *International Journal of Human-Computer Interaction, 35*(11), 956–971. https://doi.org/10.1080/10447318.2018.1561784.

Goetz, J. and Haleem, K. (2021). Investigating Distracted Driving at Roundabouts and Different Road Configurations Using a Driving Simulator.

Kalra, N., & Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice, 94*, 182–193. https://doi.org/10.1016/j.tra.2016.09.010.

Khan, N. A. and Habib, M. A. (2021). How built environment characteristics of pedestrian-vehicle collision locations affect pedestrian injury severity involving distracted driving?

Kidando, E. et al. (2019). Assessment of factors associated with travel time reliability and prediction: an empirical analysis using probabilistic reasoning approach. *Transportation Planning and Technology, 42*(4), 309–323. https://doi.org/10.1080/03081060.2019.1600239.
Kidando, E., Moses, R., & Sando, T. (2019). Bayesian regression approach to estimate speed threshold under uncertainty for traffic breakdown event identification [Preprint]. *Journal of Transportation Engineering, Part A: Systems*. https://doi.org/10.1061/jtepbs.0000217.
König, M., & Neumayr, L. (2017). Users' resistance towards radical innovations: The case of the self-driving car. *Transportation Research Part F: Traffic Psychology and Behaviour, 44*, 42–52. https://doi.org/10.1016/J.TRF.2016.10.013.
Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science, 6*(3), 299–312. https://doi.org/10.1177/1745691611406925.
Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology.*. https://doi.org/10.1037/a0029146. General [Preprint].
Kutela, B., Avelar, R. E., & Bansal, P. (2022). Modeling automated vehicle crashes with a focus on vehicle at-fault, collision type, and injury outcome. *Journal of Transportation Engineering, Part A: Systems, 148*(6), 04022024. https://doi.org/10.1061/JTEPBS.0000680.
Kutela, B., Das, S., & Dadashova, B. (2022). Mining patterns of autonomous vehicle crashes involving vulnerable road users to understand the associated factors. *Accident Analysis & Prevention, 165*, 106473. https://doi.org/10.1016/J.AAP.2021.106473.
Kutela, B., & Teng, H. (2019). Prediction of drivers and pedestrians' behaviors at signalized mid-block Danish offset crosswalks using Bayesian networks. *Journal of Safety Research, 69*, 75–83. https://doi.org/10.1016/j.jsr.2019.02.008.
Liu, B. et al. (2019). Trajectory planning for autonomous intersection management of connected vehicles. *Simulation Modelling Practice and Theory, 90*, 16–30. https://doi.org/10.1016/J.SIMPAT.2018.10.002.
Makowski, D. et al. (2019a). Indices of effect existence and significance in the bayesian framework. *Frontiers in Psychology, 10*. https://doi.org/10.3389/FPSYG.2019.02767.
Makowski, D. et al. (2019b). Indices of effect existence and significance in the bayesian framework. *Frontiers in Psychology, 10*(December), 1–14. https://doi.org/10.3389/fpsyg.2019.02767.
Mammadrahimli, A. (2015). Assessment of crash location improvements in map-based geocoding systems and subsequent benefits to geospatial crash analysis, *All Theses* [Preprint].
Mittal, A., & Kassim, A. (2007). Bayesian network technologies: Applications and graphical models. *IGI Global*. https://doi.org/10.4018/978-1-59904-141-4.
Monyo, D.E. *et al.* (2021). Are Older Drivers Safe on Interchanges? Analyzing Driving Errors Causing Crashes.
Pande, A., Abdel-Aty, M. and Das, A. (2010). A classification tree based modeling approach for segment related crashes on multilane highways. https://doi.org/10.1016/j.jsr.2010.06.004.
Papadoulis, A. et al. (2019). *Evaluating the safety impact of connected and autonomous vehicles on motorways*. Elsevier [Preprint].
Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*(PB), 252–263. https://doi.org/10.1016/J.TRF.2014.04.009.
Petrovic, D., Mijailović, R., & Pešić, D. (2020). Traffic accidents with autonomous vehicles: Type of collisions, manoeuvres and errors of conventional vehicles' drivers. In *Transportation Research Procedia* (pp. 161–168). Elsevier B.V. https://doi.org/10.1016/j.trpro.2020.03.003.
Phan, D., Pradhan, N. and Jankowiak, M. (2019). Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. Available at: http://arxiv.org/abs/1912.11554 (Accessed: January 21, 2022).
Polders, E. et al. (2015). Crash patterns at signalized intersections. *Transportation Research Record, 2514*, 105–116. https://doi.org/10.3141/2514-12.
Rahman, M. H., Abdel-Aty, M., & Wu, Y. (2021). A multi-vehicle communication system to assess the safety and mobility of connected and automated vehicles. *Transportation Research Part C: Emerging Technologies, 124*. https://doi.org/10.1016/j.trc.2020.102887.
Rahman, M. S. et al. (2019). Safety benefits of arterials' crash risk under connected and automated vehicles. *Transportation Research Part C: Emerging Technologies, 100*, 354–371. https://doi.org/10.1016/J.TRC.2019.01.029.
Rista, E. et al. (2017). Examining the safety impacts of narrow lane widths on urban/suburban arterials: Estimation of a panel data random parameters negative binomial model. *Journal of Transportation Safety & Security, 10*(3), 213–228. https://doi.org/10.1080/19439962.2016.1273291.
Sheehan, B. et al. (2019). Connected and autonomous vehicles: A cyber-risk classification framework. *Transportation Research Part A: Policy and Practice, 124*, 523–536. https://doi.org/10.1016/j.tra.2018.06.033.
Singh, S. (2015). Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey, *Traffic Safety Facts - Crash Stats* [Preprint].
Teoh, E. R., & Kidd, D. G. (2017). Rage against the machine? Google's self-driving cars versus human drivers. *Journal of Safety Research, 63*, 57–60. https://doi.org/10.1016/J.JSR.2017.08.008.
TIMS (2021). *TIMS - Transportation Injury Mapping System*. Available at: https://tims.berkeley.edu/about.php (Accessed: June 27, 2021).
Wang, L. et al. (2020). How many crashes can connected vehicle and automated vehicle technologies prevent: A meta-analysis. *Accident Analysis & Prevention, 136*, 105299. https://doi.org/10.1016/J.AAP.2019.105299.
Wang, X., & Abdel-Aty, M. (2006). Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis & Prevention, 38*(6), 1137–1150. https://doi.org/10.1016/J.AAP.2006.04.022.

Norris Novat is a Graduate Research and Teaching Assistant at Cleveland State University, Ohio. His area of research is on intelligent transportation systems, connected and automated vehicles, and traffic operations and safety. Currently in pursuit of a master's in civil engineering – Transportation from Cleveland State University, Ohio. He has demonstrated knowledge in transportation planning and design, traffic signal operations and optimization. Skilled in statistical data analysis, data pre-processing, data mining and data visualization.

Emmanuel Kidando, Ph.D., P.E., is an assistant professor in the Department of Civil & Environmental Engineering at Cleveland State University. His areas of expertise include intelligent transportation systems (ITS), connected vehicle systems, freeway operations, highway safety, and data science.

Boniphace Kutela, Ph.D., P.E., is an Assistant Research Scientist at Texas A&M Transportation Institute. He received his Ph.D. in Civil Engineering from University of Nevada Las Vegas in 2019. He has more than 10 years of experience in transportation safety and operations related research projects. His major areas of expertise include roadway data collection and analysis methodologies, machine learning, and predictive modeling with a focus on traffic safety. He has authored/co-authored over thirty publications in reputable peer-reviewed journals and presented over twenty presentations at national and local conferences.

Angela E. Kitali, Ph.D., is an assistant professor at the University of Washington Tacoma. She received her doctoral degree in civil engineering from Florida International University in 2020. Her research focuses on a wide range of topics in transportation safety, traffic incident management, traffic operations, and transportation equity.

Journal of Safety Research 84 (2023) 138–154

Contents lists available at ScienceDirect

Journal of Safety Research

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

Analysis of prevention through design studies in construction: A subject review

Ziyu Jin [a,*], John Gambatese [b], Ali Karakhan [c], Chukwuma Nnaji [d]

[a] Department of Civil, Construction and Environmental Engineering, University of New Mexico, Albuquerque, NM, USA
[b] School of Civil and Construction Engineering, Oregon State University, Corvallis, OR, USA
[c] Department of Reconstruction & Projects, University of Baghdad, Al-Jadriya, Baghdad, Iraq
[d] Department of Construction Science, Texas A&M University, College Station, TX, USA

ARTICLE INFO

Article history:
Received 26 October 2021
Received in revised form 12 June 2022
Accepted 20 October 2022
Available online 1 November 2022

Keywords:
Prevention through Design (PtD)
Construction safety
Construction technologies
Design for safety
Occupational health and safety

ABSTRACT

Introduction: The concept of addressing and minimizing construction site safety risks in the early phase of a project has generated research interest, especially since the National Institute for Occupational Safety and Health (NIOSH) launched its national *Prevention through Design* (PtD) initiative in July 2007. In the last decade, several studies on PtD with differeing goals and methods have been published in construction journals. To date, few systematic examinations of the development and trends associated with PtD research have been conducted in the discipline. *Method:* This paper presents a study of the latest PtD research trends in construction safety management through analysis of publications in prominent construction journals from 2008 to 2020. Both descriptive and content analyses were conducted based on the number of papers published annually and clusters of topics covered in the papers. *Results:* The study shows an increasing interest in PtD research in recent years. Research topics covered mainly focus on the perspectives of PtD stakeholders, PtD resources/tools/procedures, and technology applications to facilitate PtD implementation in practice. This review study provides an improved understanding of the state-of-the-art of PtD research in terms of accomplishments and research gaps. The study also compares the findings from journal articles with industry best practices related to PtD to guide future research in this domain. *Practical Application:* This review study is of significant value to researchers to overcome the limitations of the current PtD studies, and to extend the scope of PtD research, and can be used by industry professionals when considering and selecting appropriate PtD resources/tools in practice.

1. Introduction

Szymberski (1997) proposed a hypothetical time-safety influence curve, which suggests that the ability to influence safety decreases as a project progresses. The curve implies that the most effective phases during which to consider and address site safety issues are the conceptual and preliminary design phases before hazards are present on site. This hypothetical time-safety curve was empirically tested and validated with workplace safety and injury data in multiple previous studies including Lingard et al. (2015) and Karakhan et al. (2018). Moreover, several studies have attempted to quantify the relationship between construction safety and design. For example, Behm (2005) found that 42% of the reviewed fatality incidents could have been prevented or the severity of injuries could have been reduced if site safety was considered during design. Failing to address site hazards in the design phase can also result in higher economic and social costs (Gambatese et al., 2017).

Recognizing the importance and benefits of addressing site safety during design and planning phases of a project, many countries/regions around the world have put in place *Prevention through Design* (PtD) regulations. For instance, in Europe, Directive 92/57/EEC "*On the implementation of minimum safety and health requirements at temporary or mobile construction sites*" (Directive, 1992) is in place to address safety and health hazards throughout the project lifecycle. But, currently, the United States lacks PtD regulations that prescribe addressing construction safety in the design (Poghosyan et al., 2018). In July 2007, NIOSH held the first PtD workshop and launched a national initiative to advocate for PtD practices (NIOSH, 2013). The ultimate goal of the PtD initiative was to prevent or reduce occupational injuries, illnesses, and fatalities through the inclusion of precautions into all designs (e.g., per-

* Corresponding author.
E-mail addresses: zjin@unm.edu (Z. Jin), john.gambatese@oregonstate.edu (J. Gambatese), karakhan@uobaghdad.edu.iq (A. Karakhan), cnnaji@eng.ua.edu (C. Nnaji).

https://doi.org/10.1016/j.jsr.2022.10.015
0022-4375/

manent structures, temporary structures, and work environment) that impact workers (e.g., construction and maintenance workers). The PtD approach is aligned with the hierarchy of hazard controls that focuses on eliminating, reducing, and minimizing hazards during the design and re-design phases (NIOSH, 2015). PtD is also commonly referred to as design for safety (DfS), design for construction safety (DfCS), construction hazard prevention through design (CHPtD), safety by design (SbD), safety in design, and safe design (Goh & Chua, 2016; Toole et al., 2017).

Literature reviews about construction safety management have been conducted that also covered the topic of PtD, but most of them placed particular focus on specific aspects, such as digital design tools for construction safety (Zhou et al., 2012) and PtD implementation factors (Poghosyan et al., 2018). A 2019 study by Hardison and Hallowell reviewed PtD studies to investigate the efficacy of the concept at reducing actual risk (Hardison & Hallowell, 2019). While the review provides useful information on PtD, the focus of the review is narrow (based on research categories and the type of evidence). Additionally, a recent PtD work (Ibrahim et al., 2022) provides a review from a scientometric perspective in terms of research trends, co-occurrence networks, article citations, and co-occurring keywords. The study also provides qualitative analysis on five main research themes: concept and management, technological advancement, capability and competency, education, and sustainability, which are different than those summarized in the present work. In contrast with the previous review studies, the present study emphasizes PtD research from 2008 onward, aiming to provide a thorough and systematic review to understand PtD research trends, achievements, and shortcomings since the first PtD workshop held in 2007. The present study also provides a broader overview of PtD research topics - reviews twice the number of articles when compared to Hardison and Hallowell (2019) - and the achievements and shortcomings for each topic – critical components for formulating practical guidance and directing future research agenda.

Based on the aforementioned discussion, the research objectives of the study can be summarized in three main points: (1) summarize the nature of PtD research based on information in available publication; (2) provide an overview of main PtD research topics related to the construction industry; and (3) identify research gaps and recommend future research directions related to PtD in construction. Achieving these objectives contributes to the body of knowledge and practice on PtD in construction in two aspects. First, it highlights the research trends and themes of PtD studies and reinforces the crucial role of PtD in the construction safety management domain; and second, it provides researchers and practitioners an in-depth and thorough investigation of the extant PtD literature and offers potential opportunities to fill the identified gaps and advance the PtD frontier.

2. Methodology

The review of PtD research was performed using a systematic review method recommended by Denyer and Tranfield (2009). According to Denyer and Tranfield, a systematic review is different from a traditional literature review, as it explores a clearly specified set of questions. Therefore, the first step in conducting a systematic review is to formulate specific research questions and establish a set of relevance and quality criteria to identify literature that should be included in the review, and what information should be extracted from each document. The fundamental questions posed for the present study are: (1) What is known from existing literature about the concept of PtD in the construction industry? For example, who are the parties of interest? What are the common interventions of interest?; and (2) What are the research trends of the existing studies, in terms of the number of publications by year, by journal, by country/region, and by research topic?

After the research questions are formulated, the subsequent steps are to locate and select relevant studies to ensure the review incorporates high-quality research contributions. The researchers conducted a search of academic databases, similar to that conducted by Poghosyan et al. (2018). To select literature of interest and filter out irrelevant literature, the scope of the research is limited to the following criteria: (1) Databases: Science Direct, Taylor and Francis, Emerald Insight, American Society of Civil Engineers (ASCE), Engineering Village, and Google Scholar; (2) Publication type: English peer-reviewed journals; (3) Article type: technical papers, scholarly papers, and case studies; (4) Key descriptors: "prevention through design," "construction hazard prevention through design," "design for safety," "design risk management," and "safety in design:" (5) Period: 2008 to 2020; and (6) Discipline: architecture, engineering, and construction. The initial search was performed to select papers having those keywords within the title, abstract, or entire article. In addition to database searches, reference lists of review studies and reports related to the topic [e.g., CPWR (2018) and Poghosyan et al. (2018)] were included in the initial search, and screened for titles that include key terms. The initial search yielded 260 publications. The articles are further examined to eliminate those that are irrelevant to the reviewed subject (e.g., those that briefly mention the concept of PtD or simply cite a PtD-related research study). As a result, 140 articles of the 260 publications were selected and included in the analysis.

The last step was to extract the pertinent information from the collected documents, to make associations, and to summarize the findings using both descriptive and content analyses. This approach is helpful in providing an overview of the research studies, identifying research trends and gaps, and guiding future studies (Zhou et al., 2015). Five data fields were recorded during the coding process: (1) paper title, (2) publication year, (3) journal title, (4) the country/region of the first author's institute/company, and (5) research topic.

Lastly, the present paper summarizes the findings from the descriptive analysis by presenting the results in terms of research trends and research topics, as well as those from the content analysis. After reviewing the selected journal articles, the authors attempted to associate the findings described in the journal articles with current PtD industry practices and standards to discover gaps between research and practice, and to provide further evidence to support the findings and solidify recommendations for future research. The authors selected the SmartMarket reports from Dodge Data & Analytics (construction.com/toolkit/reports/), ANSI/ASSP Z590.3 Prevention through Design Standard, as well as other NIOSH and Centers for Disease Control and Prevention (CDC) related documents that pertain to the concept of PtD as sources to gain an understanding of the industry trends in the United States. The authors believe that the findings described in the identified reports and documents are potential indicators of current PtD practices given the large number of contractors and/or architects who participated in the studies. Based on the findings, conclusions are drawn and recommendations to future research are provided.

3. Overview of the literature

3.1. Publications distributed by year

The year profile of PtD publications from 2008 to 2020 is shown in Fig. 1. The dotted trend line in the figure was obtained using a

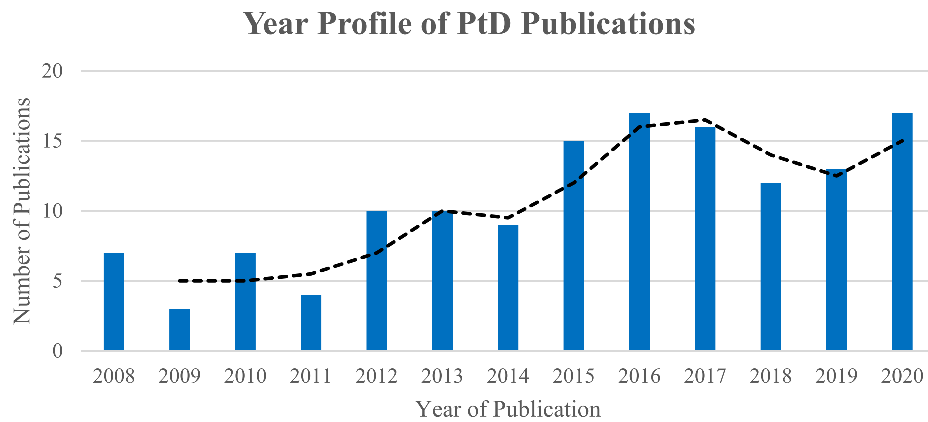


Fig. 1. Year Profile of PtD Publications (from 2008 to 2020).

simple moving average method over a two-year period. The figure reveals that the relevant papers published annually prior to 2011 were fewer than 10. There has been an increase in yearly journal articles published since 2015, with the number of publications more than doubling between 2008 and 2011. Fig. 1 highlights the general increasing trend of PtD publications and shows that more attention has been paid to the concept of PtD and its applications in the construction industry. The research finding is consistent with that of Sinyai and Choi (2020) – the increase in the number of publications pertaining to the topic of PtD is pronounced, and has become a new area of research in the construction occupational safety and health domain.

3.2. Publications distributed by journals

The 140 reviewed articles are distributed across 42 journals. The main sources are the *Journal of Construction Engineering and Management* (21), *Safety Science* (19), *Automation in Construction* (12), and *Construction Management and Economics* (10). This result suggests that researchers prefer these sources as mediums for disseminating PtD related findings.

3.3. Publications distributed by country/region

Based on the first/corresponding author's affiliation/company, 25 countries/regions contributed at least one publication to PtD research. Table 1 lists the countries/regions that published at least five articles. Even though PtD practices remain voluntary in the United States, the United States contributed the highest number of PtD publications (39.3%). Apparently, the concept of PtD has drawn researchers' attention worldwide, especially in the United States where PtD legislation is absent and efforts have been undertaken to actively investigate and promote PtD. However, only peer-reviewed journal articles published in English were included in the analysis, which may limit the scope of the analysis in terms of the countries/regions with PtD publications, and could limit generalization of the research findings to a larger group.

Table 1
Contributions by country/region.

Country/Region	Count	Percentage
USA	55	39.3%
UK	15	10.7%
China	13	9.3%
Australia	12	8.6%
Spain	8	5.7%
Singapore	6	4.3%
Others	34	22.1%

3.4. Publications distributed by research topic

The reviewed articles were categorized by content analysis based on the research topics investigated and discussed. Six main topic groups were identified based on the categorization, namely: (1) design-related factors and Occupational Safety and Health (OSH); (2) perspectives of PtD stakeholders; (3) barriers, enablers, and motivators to PtD implementations; (4) PtD interventions, including regulation and legislation as well as education and training; (5) PtD resources/tools/procedures and practical cases; and (6) construction technologies in PtD research. These six groups add up to more than 90% of all of the identified studies. Table 2 shows the count and percentage distribution of the research topics. Some articles discussed more than one identified theme and, therefore, the total count amounts to more than 140 articles and the percentage adds up to more than 100%. The most frequent study topic is PtD resources/tools/procedures and practical case descriptions, accounting for 49.3% of all studies. Since the number of studies on other PtD related topics such as PtD practice directions (Toole & Gambatese, 2008; Gambatese et al., 2017) is relatively small (n = 5), the content analysis only focuses on the six main groups.

4. Content analysis and discussion

4.1. Design and OSH

Prior to the first NIOSH PtD workshop in the United States, many studies (e.g., Behm, 2005) explored design's potential to reduce site risks through accident analysis. Several studies (16 were identified in the present study), using different methods,

Table 2
Topic categorization of the reviewed PtD articles.

Topics		Count	Percentage
Design and OSH		16	11.4%
Perspectives of PtD stakeholders		27	19.3%
PtD Barriers, Enablers and Motivators		14	10.0%
PtD Interventions	Regulations and Legislations	6	4.3%
	Education and Training	10	7.1%
PtD Resources/Tools/Procedures and Practical Cases		69	49.3%
Technologies and PtD		45	32.1%
Other PtD Topics (e.g., PtD practice directions, risk and financial impacts, PtD with lean construction etc.)		5	3.6%

have been conducted worldwide since then to validate the relationship between designs and OSH. Researchers mostly relied on historical injury/fatality databases and/or experts' perspectives (e.g., Burlet-Vienney et al., 2015; Gambatese et al., 2008; Wong et al., 2009) to examine and quantify if an injury/fatality related to design factors. Multiple other approaches were adopted to confirm the influence of designs on OSH, such as surveys (Wong et al., 2009), safety performance data along with interviews (Atkinson & Westall, 2010), empirical analysis (Lingard et al., 2012), and a systems approach (Alomari & Gambatese, 2016; Pirzadeh & Lingard, 2017).

Notably, only a few studies have attempted to measure if PtD design decisions actually reduced OSH risks during construction, similar to that conducted in longitudinal studies carried out by Weinstein et al. (2005) and Lingard (2013). Additionally, there is no well-established system that identifies whether design-related factors are relevant to OSH and helps in assessing the extent of influence on the frequency, severity, and/or exposure of a risk factor. Continued research efforts are required to establish systematic and consistent empirical methods to assess the causal relationship between the concept of PtD and OSH for workers involved in the entire life cycle of a project.

4.2. Perspectives of PtD Stakeholders

Recognizing design as a contributing factor in OSH incidents, the roles of involved parties in construction projects, their attitudes, awareness, and knowledge of PtD and related practices have been investigated in studies. Approximately-one-fifth of the reviewed articles (19.3%) discussed this topic.

- *Designers*

Most the investigated parties are designers. An essential question posed in the analysis is "who is the designer in PtD?" Besides architects, and civil and structural engineers, various occupations, such as designers of temporary structures, mechanical engineers, electrical engineers, interior designers, surveyors, and specialist suppliers, contractors or subcontractors that provide design input, have been included as "designers" in PtD studies (Breslin, 2007; Kikwasi & Smallwood, 2016; Öney-Yazıcı & Dulaimi, 2015; Toh et al., 2017). Indeed, a "designer" term has a broad meaning. For example, in the UK (Construction Design and Management [CDM] 2015), designers are defined as "organizations or individuals who as part of a business, prepare or modify designs for a building, product or system relating to construction work." When preparing or modifying designs, designers should eliminate, reduce, or control foreseeable risks that may arise during construction, maintenance, and use of a building. For example, for rooftop vegetation or skylights, designers could consider placing permanent guardrails around roof openings to prevent construction workers, maintenance workers, and end users from falling through the openings to lower levels (Behm, 2012). The responsibility of "designers" also includes providing information to other stakeholders to help them fulfill their health and safety duties.

With a predominate role in PtD practices, designer awareness of the concept, attitude towards the concept, and knowledge level of PtD are extremely important factors in PtD implementation. Investigating designers' PtD perspectives is the subject of interest. There is general agreement amongst designers from different countries (e.g., Goh & Chua, 2016; Karakhan & Gambatese, 2017a; Abueisheh et al. (2020) that designers have a high level of awareness of, and confidence in, the PtD concept, while they have limited knowledge about the hierarchy of controls used for hazard and risk mitigation in construction. As for PtD implementation practices, it was found that the level of engagement is low, and the PtD practices are immature, informal, and mostly a work-in-progress (Toh et al., 2017; Che Ibrahim & Belayutham, 2020). Furthermore, compared to engineers, owners, and contractors, designers were generally more resistant to PtD implementation (Tymvios & Gambatese, 2016b; Karakhan & Gambatese, 2017a); this finding, however, could be country-specific. Designer nationality, age, experience, professional background, and the safety culture of the organization they work in were found to be associated with their attitude towards PtD (Öney-Yazıcı & Dulaimi, 2015).

- *Owners/Clients*

Owners/clients have also been identified as having a vital role in construction safety. Their involvement can influence designer performance and reduce potential onsite risks (Liu et al., 2017; Votano & Sunindijo, 2014). Owner/client motivations for undertaking PtD have the greatest influence on integrating safety in designs (Goh & Chua, 2016; Toh et al., 2017). Owner influence, especially via an owner's effective leadership behaviors, was found to lead to high safety performance (Tymvios & Gambatese, 2016a; Wu et al., 2015). In practice, clients rarely contributed to improving the safety of construction workers in Tanzania (Kikwasi & Smallwood, 2016). In the United States, Tymvios and Gambatese (2016b), Toole et al. (2017) and Gambatese et al. (2017) found the vast majority of employees in the owner organizations investigated were unaware of the PtD concept. Nevertheless, the abovementioned U.S. studies revealed positive attitudes from owners/-clients towards PtD as an intervention for improving OSH.

- *Constructors*

Constructors traditionally have the sole responsibility for maintaining jobsite safety from legal and contractual perspectives, especially in the United States. PtD implementation requires close collaboration between the designer and constructor, such as having the constructor review the design as an additional safety check or having the designer observe construction progress periodically for potential identification of safety concerns (Toole, 2005). A survey conducted by Tymvios and Gambatese (2016b) revealed that only approximately 17% of the participating contractors stated that they knew of the PtD concept before the survey. More than 80% of the contractors felt that decisions made before the design phase, made during the design, and made during construction could help eliminate hazards. Larsen and Whyte (2013) reinforced the importance of PtD from the contractor's perspective in their study, and found that insufficient design documents increase unplanned rework, and late design changes increase the difficulty of planning safe procedures.

The involvement of designers and owners in enhancing construction site safety through the design is strongly supported by contractors (Tymvios & Gambatese, 2016b). However, contractors feel that designers might lack knowledge about how construction site operations and procedures take place to effectively adopt PtD in their designs (Karakhan & Gambatese, 2017b), which is consistent with designers' perspectives of their knowledge of the PtD concept.

To ensure successful PtD implementation, owners/clients, designers, and constructors all have to play their roles in practice. And the safety attitudes across PtD stakeholders can be compared using the inter-organizational safety climate instrument proposed by Saunders et al. (2017). In summary, PtD stakeholders generally hold positive attitudes regarding the concept of PtD and agree on the importance of addressing and dealing with safety and health hazards in designs. However, implementation of PtD in practice in the United States is quite limited due to identified barriers, which are discussed in the next section of this manuscript. Continued efforts are expected to explore ways to strengthen stakeholder confidence in the concept itself and implementation of the concept. For example, learning from experienced stakeholders who are from countries/regions with PtD regulations could lead to further understanding of how to successfully implement PtD in practice and quantify its impacts.

4.3. PtD barriers, enablers and motivators

Of the 140 reviewed articles, 14 articles (10.0%) discuss barriers, enablers, and motivators for PtD diffusion. With respect to barriers, most articles focus on the barriers experienced by designers. As summarized in Table 3, investigations in countries/regions with or without PtD regulations have revealed that design professionals face some of the same obstacles to the diffusion of PtD. These obstacles include: economic barriers (additional costs associated with PtD implementation), contractual barriers (changes in contract clauses), and knowledge/information barriers (designer lack of knowledge about safety or construction means and methods). Potential legal liability is identified as one of the most prominent impediments for designers to implement PtD in countries/regions without PtD regulations (Karakhan & Gambatese 2017a). There might be an enhanced risk of lawsuits for design firms that adopt PtD in their project designs (Karakhan & Gambatese, 2017c; Toole & Erger, 2019).

Researchers have also identified several factors and practices that enable PtD implementation. Examples of enablers include designers having the requisite knowledge and skills, adequate time available in the design to consider safety, construction safety given a high priority similar to other project objectives, and construction means and methods are well identified during design (Gambatese et al., 2017). Regarding motivators of PtD implementation, Bong et al. (2015) found that legal and regulatory factors were considered to be major motivators for designers in South Australia (where PtD regulations are present). For studies conducted in the United States (where no PtD regulations exist), ethical behavior was viewed by design professionals as the most prominent enabler (Karakhan & Gambatese, 2017a). Lastly, advancing the concept of social equity that requires design professionals to consider the safety and well-being of all construction stakeholders, including construction and maintenance workers, was found to be another PtD motivator (Toole & Carpenter, 2013).

As stated in Gambatese et al. (2017), the findings regarding PtD barriers, enablers, and motivators are mostly based on observational or anecdotal research. Quantitative studies supported by empirical data are needed to confirm the existence or absence of the identified barriers, enablers, and motivators to understand and facilitate PtD diffusion across the construction industry.

4.4. PtD Interventions

Aiming to address the identified barriers to PtD implementation, 16 of the reviewed articles (11.4%) conducted PtD intervention research. The 16 articles discuss two types of PtD interventions and their effectiveness: regulations/legislation, and PtD education and training.

- *Regulation and Legislation*

Many countries/regions (e.g., the UK, Australia, and Singapore) have implemented regulations and legislation that impose site safety duties on design professionals with the goal of encouraging the adoption of PtD in practice. Six reviewed articles examined the effectiveness of PtD-related regulations on OSH. In the EU, the studies performed by Martínez-Aires et al. (2010, 2016) revealed that from the year when the regulations were issued, 10 countries in the EU have experienced a greater than 10% drop in the workplace accident rate. Bong et al. (2015) found that the guidelines in Australia promote PtD requirements without significantly increasing designer and contractor workloads. Designers were aware of site hazards and were willing to address them during the design phase (Bong et al., 2015).

Saunders (2016) attempted to examine the reasons why the United States performed poorly in construction worker safety and health performance compared to Australia. Saunders found that safety decisions in Australia were generally made further upstream than in the United States due in part to U.S. designers' fear of liability. One additional possible contributor to poorer safety performance in the United States is the lack of PtD regula-

Table 3
Barriers to designer implementation of PtD.

Items		Previous Studies						
Study		Gambatese et al. (2017)	Bong et al. (2015)	Goh and Chua (2016)	Öney-Yazıcı and Dulaimi (2015)	Tymvios and Gambatese (2016b)	Karakhan and Gambatese (2017a)	Manu et al. (2019)
Country/Region Investigated		UK	Australia	Singapore	United Arab Emirates	US	US	Nigeria
Has PtD Regulation?		Yes	Yes	Yes	No	No	No	No
Investigated Stakeholders		D, C, O	D, C	D	D	D, C, O	D, C	D
Barriers	Regulatory		X				X	
	Economic		X	X		X	X	
	Legal		X		X	X	X	X
	Contractual			X		X	X	
	Ethical						X	
	Cultural						X	
	Knowledge / Information	X					X	X
	Other	Safety is not given higher priority		Client's attitude toward PtD	Lack of PtD education/training and resources		Lack of motivation; Safety is not given higher priority; Schedule and budget concerns	Lack of PtD education/training and resources

Notes Investigated Stakeholders: D = designers, C = constructors, O = owners.

tions - implementing PtD remains a voluntary effort in the United States. As a result, designers tend to avoid making OSH-related decisions and leave health and safety concerns to the contractors during the construction phase. Conversely, for countries with PtD legislation, such as Australia, there is legislation that addresses the liability of designers on construction projects; thereby OSH decisions are made earlier in the project development process, during the design and planning phases. The absence of PtD regulations was recognized as one major barrier to PtD implementation by many studies (see Table 3). Even though there is no obligation to implement PtD in the United States, some U.S. professional organizations have published new programs and standards to encourage PtD practice, such as the PtD-related pilot credit in the U.S. Green Building Council (USGBC) (2015) LEED rating system, and the ANSI/ASSP Z590.3-2011 standard on prevention through design. However, no study was found that investigated the prevalence and effectiveness of using these guidelines and standards.

To date, only limited studies (e.g., Mendeloff & Staetsky, 2014) have been carried out to examine the effectiveness of PtD regulations on reducing injuries and fatalities. The authors of the present study did not find a comparative study that investigates the differences in PtD regulations that are put in place in different countries/regions. For researchers from a country/region without PtD regulations, such as the United States, studies could be carried out to learn from countries that have successfully put PtD regulations in place. For example, studies could be conducted on what should be included in a PtD review process, how to clearly formulate contractual obligations, and which PtD approach is more effective: having principal designers and contractors (the approach used in the UK) or having design for safety registers and/or professionals (the approach used in Singapore).

- *Education and Training*

PtD has been a mandatory practice in Europe for over two decades, but studies have found that PtD is not embedded successfully in civil engineering curricula throughout all EU countries. In Spain, Cortés et al. (2012) proposed that a separate course on occupational risk prevention (a major PtD focus area) should be included in all engineering degree programs. However, the inclusion of PtD in courses is insufficient to help future professionals to be qualified in practicing PtD. Significant improvements in course designs are also needed (Lopez-Arquillos et al., 2015).

In the United States, Mann Iii (2008) suggested that the most effective way to introduce PtD in educational curricula is through modules, rather than in complete courses. Behm et al. (2014) developed and examined a PtD educational intervention, a 70-min lecture as a part of an Engineering Project Management class, through a before-and-after study of engineering students. The study showed that such an intervention helped in the development of students' safe design thinking over time.

To identify the best approaches to deliver the PtD concept and practices to students, researchers have discovered several innovative safety training methods besides traditional methods (i.e., lectures), such as a Building Information Modeling (BIM) enabled training module (Clevenger et al., 2015), and an energy-based training module (Tixier et al., 2018). These studies (Clevenger et al., 2015; Tixier et al., 2018) have shown such innovative training methods improved trainee learning experiences and provide a better platform to obtain PtD knowledge.

For designers no longer enrolled in university courses, there are several PtD training materials, such as those from NIOSH (www.cdc.gov/niosh/topics/ptd/pubs.html). However, after investigating design professionals' preferred type of PtD training in Nigeria, Manu et al. (2019) found that professionals have a higher preference for attending a seminar/workshop, and a moderate preference for an online course/study. Öney-Yazıcı and Dulaimi (2015) recommended influencing designers by introducing hands on experience with the concept because of the prevalence of the "learn by doing" method in the industry.

Apparently, the current level of PtD education and training is inadequate and there is no consensus on the best ways to introduce the PtD concept, procedures, and best practices for students (future professionals) and for existing professionals. Studies could be conducted to examine ways to integrate PtD content into existing university curricula and convey the effectiveness in preventing accidents when applying academic theories in practices. As for training methods for professionals, except for exploring effective ways for PtD knowledge dissemination, researchers are also encouraged to investigate professionals' preferred training methods and to develop an appropriate PtD training program.

4.5. PtD resources/tools/procedures and practical cases

Designers typically lack sufficient knowledge about construction safety to adopt PtD in designs, as well as to identify and assess hazards during the design/planning phase. Indeed, the hazard identification levels in construction projects are far from ideal (Carter & Smith, 2006). Hallowell and Hansen (2016) found that designers are only capable of identifying 38% of construction hazards from design documents. However, formal identification of hazards is one of the fundamental steps to ensure the success of safety management (Carter & Smith, 2006). Meanwhile, conducting a risk assessment is essential for designs as designers could benefit from having information that enables analyzing potential hazards. To improve designer hazard recognition skills, and their abilities to assess potential hazards and select safe designs, 61 of the reviewed articles proposed and developed PtD resources and tools. Examples of the PtD resources and tools are shown in Table 4 and Table 5. In addition, seven studies provide practical cases on how PtD was implemented for different applications, such as for greenery systems (e.g., Behm 2012), for heavy construction projects (Ezisi & Issa, 2019), and for solar installations (Ho et al., 2020).

Most of the proposed PtD tools (shown in Tables 4 and 5) fall under the domain of risk management, especially for hazard identification and risk assessment. One set of research studies focuses on identifying a single type of hazard only, such as falls (Cooke et al., 2008; Zhang, Sulankivi, et al., 2015), and sight obstructions (Cheng & Teizer, 2014; Marks & Teizer, 2013). Another set of studies emphasizes hazard identification and assessment for specific types of construction operations, including underground construction (Seo & Choi, 2008), highway construction (Esmaeili & Hallowell, 2013), and multistory buildings (Dharmapalan et al., 2015). Few tools were developed to provide risk mitigation advice, recommend design alternatives, or facilitate site-planning processes (Dharmapalan et al., 2015), while such information could be highly useful for designers and constructors when practicing PtD. Even though it might be impossible to generalize successful PtD work given that most findings are the results of experience-based approaches (Lopez-Arquillos et al., 2015), it is important to investigate ways to apply PtD consistently during design and to ensure determined safety measures are implemented on sites.

Traditional PtD resources/tools (Table 4) consist of risk-assessment matrices, design guides, and suggestions from lessons learned databases, which are consistent with the findings from Gambatese et al. (2017). With the help of construction technologies, safety management could be performed more effectively during design through visualization, simulation, data mining, and integration of a safe design database (details about construction technology usage for PtD can be found in the next section).

The definition of PtD in the construction sector often applies to "designing out" potential OSH hazards in the process of designing

Table 4
Traditional PtD resources/tools.

Author(s) (year)	Target Stakeholders		Function				Description
	Designers	Constructors	HI	RA	DS	SP	
Seo and Choi (2008)	X		X	X			A risk assessment model linking risk events with design items for underground construction projects (with a case study focused on an open-cut type subway construction project)
Frijters and Swuste (2008)	X		X	X	X		A risk assessment method to evaluate construction risks at an activity level (with a case study focused on the construction of different types of floor systems)
Zou et al. (2008)	X	X	X	X			A program titled ROAD (Risk and Opportunity at Design) that helps to perform risk and opportunity analyses at the design stage of building projects (with two case studies focused on a five-star green rating office building and a modern university educational building project)
Nussbaum et al. (2009)	X		X	X			A tool that aims to assist panel designers minimize ergonomic risk for residential carpenters
Gangolells et al. (2010)	X		X	X	X		A risk assessment method that helps designers to assess safety-related performance of residential construction designs
Fung et al. (2010)	X	X		X			A risk assessment model (RAM) that helps with assessing risk levels by analyzing risk factors of major types of trades (with a case study focused on a construction project in Hong Kong)
Kim et al. (2011)	X	X		X			A risk assessment study that evaluates levels of risk related to the design of prefabricated (panelized) walls
Dewlaney and Hallowell (2012)	X	X					Risk mitigation strategies for high performance sustainable building construction
Fortunato et al. (2012)	X		X				A risk identification approach for high-performance sustainable construction projects
Dewlaney et al. (2012)	X			X	X		A risk quantification approach for high-performance sustainable construction projects
Esmaeili and Hallowell (2013)	X	X	X	X			A risk assessment model to quantify highway construction risks by combining a safety risk database and a project schedule
Gangolells et al. (2013)	X	X	X	X	X		An integrated model for assessing and controlling environmental, health, and safety risks at the project level (with a case study focused on construction projects in a small construction company)
Dharmapalan et al. (2015)	X		X	X			An online tool (SliDeRulE) that links specific design features with construction risks to assess safety risks associated with the design of multistory buildings
Esmaeili et al. (2015)	X		X	X			An attribute-based risk identification and analysis method to help with assessing risks associated with a set of construction activities or building components
Karakhan and Gambatese (2017b)	X		X	X			A risk identification and assessment method that assesses OSH risk associated with the Leadership in Energy and Environmental Design (LEED) rating system
Penaloza et al. (2017)	X	X	X	X			A risk identification and assessment method for temporary edge protection systems in buildings
Ning et al. (2018)	X	X				X	An ant colony site layout planning optimization model with considerations of the facility safety relationship, geographic safety relationship, and cost reduction (with a case study focused on 13 temporary facilities on a construction site)
Fargnoli et al (2018)	X	X		X			A tool that integrates quality function deployment (QFD) and the analytic network process method to help hazard identification and risk assessment at a working task level (with a case study focused on a construction company)

Notes Function: HI = Hazard identification; RA = Risk assessment; DS = Design suggestions; SP = Site planning.

the end product to be constructed (Weidman et al., 2016). However, the definition of PtD goes beyond that; it could be applied to the design of all tools, equipment, materials, temporary structures, and work processes that are used during the construction process, and to the design of the constructed environment itself (Young-Corbett, 2014). Only a small number of studies have explored the applications of the PtD concept in relation to temporary structures (e.g., Zhang et al., 2013), construction equipment designs (e.g., Marks et al., 2013), and the construction environment (e.g., Gangolells et al., 2013). Moreover, only a few PtD research studies have focused on developing PtD solutions for the OSH of maintenance workers (Behm, 2012; Chew et al., 2019).

Furthermore, the majority of the PtD tools/resources developed aim at addressing worker safety; only six articles place an emphasis on worker health issues, especially on work-related musculoskeletal disorders (WMSDs) [e.g., Nussbaum et al. (2009)]. Thus, more studies are anticipated to explore PtD solutions for temporary works, construction equipment, and the construction environment, as well as other frequently experienced health issues by construction and maintenance workers.

4.6. Technologies and PtD

Thirty-nine of the reviewed articles (27.9%) discussed or applied construction technologies to address and control OSH risks in early stages of a project, a significant percentage that supports PtD diffusion in the construction industry. Examples of technology-based PtD tools/resources are shown in Table 5. Fig. 2 displays the number of different construction technologies discussed in the identified studies. It is evident that BIM and visualization technologies, such as virtual prototyping (VP) and virtual reality (VR), are the most popular technologies used in PtD research to improve health and safety management in construction.

BIM is as a powerful information platform for stakeholders to collaborate. With respect to safety, BIM mainly helps with the identification, assessment, and control of construction safety hazards in designs (Malekitabar et al., 2016), assists with site safety planning, and shows potential to integrate with PtD regulations (e.g., CDM regulations) Mzyece et al. (2019). A variety of BIM-based PtD tools have been invented to assist designers with conducting automatic safety checks (e.g., Zhang et al., 2013), performing 4D site safety planning (e.g., Jin et al., 2019), and offering safety

Table 5
Technology-based PtD resources/tools.

Construction Technology	Author(s) (year)	Target Stakeholders			Function				Description
		Designers	Constructors	Others	HI	RA	DS	SP	
Infographics	Edirisinghe et al. (2016)	X			X		X		An infographic approach that provides suggestions for wicked design problems (with a case study focused on façade design)
	Lingard (2018)	X			X				An infographics approach that helps designers to identify potential hazards (with a case study focused on façade design)
3D/4D CAD	Benjaoran and Bhokha (2010)	X	X		X		X	X	A CAD-based system to identify work-at-height hazards, provide advice regarding proper safety measures, and help with site planning (with a case study focused on a three-floor hotel)
	Rwamamara et al. (2010)	X	X		X			X	A 3D/4D CAD visualization approach to identify safety risks in the design process (e.g., clash detection, work tasks sequence, and workspace congestion), and to assist in site planning (with a case study focused on a construction project that consists of four building blocks)
BIM	Kim et al. (2011)		X					X	A BIM-based method to generate scaffold plans by identifying locations of scaffolds, creating building models, and simulating the work with schedules (with a case study focused on a five-story office building)
	Zhang et al. (2013), Melzner et al. (2013) and Zhang et al. (2015)	X	X		X		X	X	BIM-based rule-checking systems to identify potential fall hazards, provide safety suggestions, and perform site planning with schedules (with case studies focused on building projects)
	Qi et al. (2014)	X	X		X		X		Two BIM-based PtD checking tools to check for fall hazards in BIM models and provide design alternative suggestions (with a case study focused on a three-story building)
	Zhang et al. (2015)		X		X		X	X	A BIM-based tool that supports automated ontology-based job hazard analysis to assist site safety planning (with a case study focused on a masonry construction project)
	Teo et al. (2016)	X			X	X			A BIM-based Construction Safety Audit Scoring System (ConSASS) that supports hazard checks and provides control measure suggestions
	Ding et al. (2016)		X		X		X		A BIM-based tool that integrates risk knowledge to facilitate a construction risk analysis process including risk factor identification, risk path reasoning and prevention plan recommendations (with a case study focused on deep foundation pit excavation activities)
	Kim et al. (2016)	X	X		X		X		A BIM-based tool that assists with scaffold planning with considerations of work sequences and movements of work crews (with a case study focused on a single story commercial building project)
	Hossain et al. (2018)	X	X		X		X		A BIM-integrated rule-based risk review system for building projects (with a case study of a 5-story building project)
	Yuan et al. (2019)	X			X		X		A Revit plug-in that helps designers check safety risks in a building model (with a case study focused on a six-story building project)
	Jin et al. (2019)	X	X			X		X	A BIM-based risk assessment approach that integrates with a work breakdown structure and a construction schedule for site planning (with a case study focused on a three-story concrete building)
	Cortes-Perez et al. (2020)	X			X	X		X	A BIM-based method that integrates with the Spanish health and safety regulations to assess risks and generate health and safety plans for building projects
	Kim et al. (2020) and Lee et al. (2020)	X			X	X			A BIM-based risk evaluation system that incorporates safety guidelines to assess risks associated with common construction hazards (e.g., falls and collisions)
	Rodrigues et al. (2021)	X			X				BIM-based plugin that allows automation in fall hazard detection and safety object generation, and integrates with a safety database
Remote Sensing	Marks et al. (2013)	X		Equipment Manufactures	X		X		An approach that measures construction equipment blind spots by using laser scanning data, and provides design suggestions based on the data

(*continued on next page*)

Table 5 (*continued*)

Construction Technology	Author(s) (year)	Target Stakeholders			Function				Description
		Designers	Constructors	Others	HI	RA	DS	SP	
									collected
	Cheng and Teizer (2014)	X	X		X			X	An approach that identifies blind spaces that obstruct the field-of-view of a tower crane operator based on data sets collected by a laser scanner, and assists in hazard identification and site-planning (with a case study focused on the construction of a four-story campus building)
Other Visualization Technology	Patrucco et al. (2010)	X			X				A Computer Image Generation for Job Simulation (CIGJS) system that assists job safety analysis (with a case study focused on extractive activities in a crushing plant)
	Chun et al. (2012)	X			X				A hazard identification approach in a virtual environment (with a case study focused on a mega structure consisting of an exhibition hall extension)
	Sacks et al. (2015)	X	X		X				A Cave Automated Virtual Environment (CAVE) that allows performing safety reviews through dialogues between designers and builders
	Golabchi et al. (2015)	X				X			A risk assessment approach to evaluate ergonomic risk factors of jobs (with a case study focused on a production line of a construction modular prefabrication company)
	Golabchi et al. (2018)	X			X	X			An integrated simulation and visualization-based safety analysis framework that enables early identification of ergonomic risks (with a case study focused on a masonry operation)
Wearable Devices	Nath et al. (2017)	X			X	X			A method that utilizes smartphone sensory data to assess the risk levels of awkward postures for ergonomic analysis (with an experiment on manual screw driving tasks)
	Umer et al. (2018)	X	X		X	X			A static balance monitoring tool of construction workers that enables early identification of construction task hazards and personal risk factors using a wearable inertial measurement unit (IMU) and a smartphone (with an experiment on rebar tying postures)
	Antwi-Afari et al. (2018)	X	X		X	X			A wearable insole pressure system which could be inserted into workers' safety boot to alert workers to mitigate the risks of work-related musculoskeletal disorders
Other Information Technology	Cooke et al. (2008)	X			X	X			A web-based tool (ToolSHeD) to assist with assessing the risks of falling from heights using argument trees
	Tixier et al. (2017)	X	X		X				A safety clash detection approach using data mining techniques
	Goh and Guo (2018)	X	X				X		A web-based tool (FPSWizard) that assists active fall protection system (AFPS) selection (with a case study focused on a work-at-height problem)
	Hare et al. (2020)	X			X		X		A web-based multi-media tool that assists designers in identifying hazards and finding suitable controls

Notes Function: HI = Hazard identification; RA = Risk assessment; DS = Design suggestions; SP = Site planning
Technologies mentioned:
3D/4D CAD: accurate visual representations of construction projects created using 3D CAD tools. 4D CAD models combine 3D CAD models with construction schedules to create visual representations of construction sequences, and can be used to facilitate team collaboration (Benjaoran and Bhokha, 2010).
BIM: a virtual software and process that involves integration of design and construction elements into accurate virtual models digitally (Eastman et al., 2011).
Infographics: a chart, diagram, or illustration that uses graphic elements to present information in a visually striking way (Merriam-Webster.com, 2022).
Remote sensing: the process of detecting and monitoring an area's physical properties by measuring its reflected and emitted radiation at a distance (USGS, 2022). Commonly used remote sensing technologies in the design and construction phases include global positioning systems (GPS), digital imaging acquired by cameras, and point clouds captured by laser scanners (Moselhi et al., 2020).
Wearable devices: a set of electronic devices that can be attached to humans as accessories or embedded in clothing to monitor worker physiological metrics, locations, or environmental conditions (Awolusi et al., 2018).
Web-based tools: digital multimedia tools consisting of photographs, videos, databases, and web search capabilities (Hare et al., 2020).

control measures (e.g., Yuan et al., 2019). Additionally, BIM-based training was found to be an effective method to introduce safety knowledge (Clevenger et al., 2015). Other visualization technologies such as VP and VR provide a virtual and visual construction environment, and are mainly used to assist safety training and education (e.g., Sacks et al., 2013). Trainee learning interests and performance related to attaining safety knowledge is improved when using these technologies. Studies have also shown that VP is an effective tool in assisting hazard identification (Chun et al., 2012). Similarly, VR helps users experience a strong sense of presence (Wang, 2002) when conducting a job safety analysis (Patrucco et al., 2010).

Furthermore, 3D/4D computer-aided design (CAD) facilitates the visualization of the design and planning of a construction project, enables intuitive comprehension of the construction processes for identifying space–time conflicts and potential safety hazards including working-at-height hazards (Benjaoran & Bhokha, 2010), and offers a collaboration tool for all parties of interest

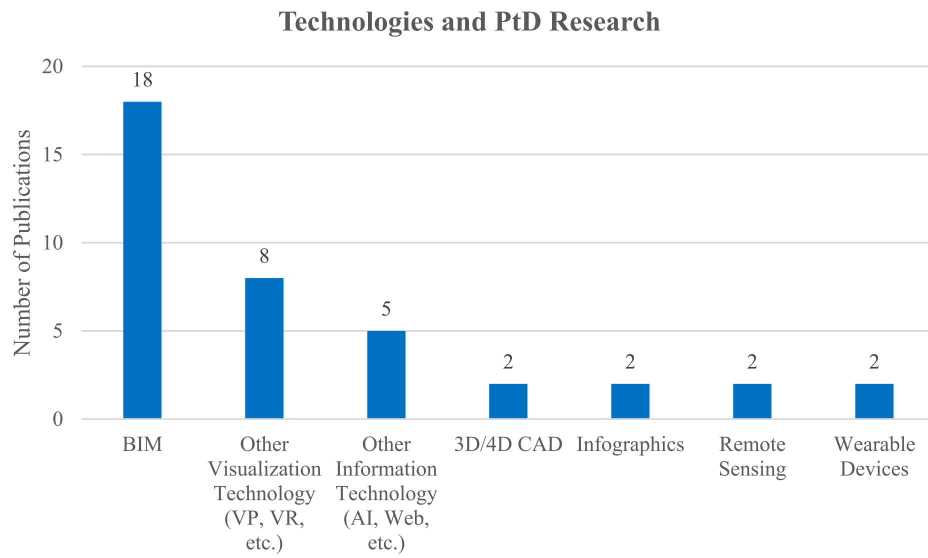


Fig. 2. Number of PtD publications by construction technology.

(Rwamamara et al., 2010). Different from 3D/4D CAD and BIM, infographics provide graphic visual representations of information in 2D. Prior studies (e.g., Edirisinghe et al., 2016) have shown that the use of infographics increased the number of identified potential OSH issues in designs and enhanced communication.

Focusing on site safety planning, the use of laser scanning data helps in the identification of hazard proximity including blind spot measurement for operators of loaders (Marks et al., 2013) and tower cranes (Cheng & Teizer, 2014). In addition, wearable devices were used to understand and monitor the postures of construction workers, and the devices were found to have great potential to prevent work-related musculoskeletal disorders (WMSDs; Antwi-Afari et al., 2018; Nath et al., 2017) and fall accidents (Umer et al., 2018). The studies performed by Nnaji et al. (2020; 2021) further show the potential of wearable smart devices for incident prevention, and point out the important roles of the physiological and environmental functions offered by wearable smart devices in improving worker safety and health. Other information technologies such as artificial intelligence (AI) provide new means for researchers to retrieve safety knowledge. Examples include the work conducted by Goh and Guo (2018) for active fall protection system designs, and that by Tixier et al. (2017) for clash identification. Moreover, as demonstrated by Cooke et al. (2008) and Goh and Guo (2018), web-interface offers an easy access platform for sharing PtD knowledge among the construction industry.

Technology advancements have benefited the diffusion of PtD by providing information platforms for stakeholders when identifying and assessing OSH hazards, assisting pre-construction site safety planning, sharing and communicating design and construction information, and providing PtD training and education effectively. Though various types of PtD technology applications have been developed, research-to-practice gaps exist since most of the applications are limited to academic research, limited implementation has been carried out (Zhou et al., 2013), and the technologies have not experienced wide-spread use in practice. Bridging the gap between research and practice requires continuous efforts to inform stakeholders regarding the availability, workability, and effectiveness of the developed PtD tools.

4.7. Industry PtD practices

Two SmartMarket reports (Dodge Data & Analytics, 2017, 2020) provide industry insights on the concept of PtD and the related industry practices. The 2017 study surveyed both contractors and architects, and the 2020 study only surveyed contractors. The survey results are consistent with the findings from the reviewed journal articles in terms of PtD awareness, barriers, and drivers, and provide additional information on the specific PtD practices implemented by architects and contractors. For example, some PtD practices frequently implemented by architects are the identification of prefabrication opportunities in collaboration with constructors and performing safety reviews. As for constructors, the installation of permanent safety features (e.g., permanent roof anchors for fall protection), prefabrication/modularization, and the use of BIM are cited as the most frequently implemented PtD practices.

Both designers and contractors view prefabrication/modularization as a PtD approach given that working with prefabricated assemblies and modules may reduce worker exposure in confined spaces, at heights, and in hazardous environmental conditions compared to onsite construction. Prefabrication/modularization was also mentioned as one of the PtD trajectories by Toole and Gambatese (2008). But, because prefabricated assemblies are often large in size, and may require additional heavy equipment (e.g., cranes) to assemble and/or install, the use of prefabricated modules may not always have a positive impact on site safety (McGraw-Hill Construction, 2011). The topic of construction prefabrication/modularization was rarely covered in the reviewed journal articles. Only the study conducted by Rubio-Romero et al. (2014) provided quantitative results showing that industrialized building systems could not claim to be safer than on-site traditional construction systems in Spain. Future studies should be conducted to investigate whether prefabrication/modularization is an effective means to practice PtD and determine factors that influence its effectiveness on improving site safety.

Another topic that was not discussed in the reviewed journal articles but was investigated in the 2017 report relates to the LEED pilot credit for PtD. The pilot credit, launched in 2015, requires conducting design reviews before the completion of schematic design that consider worker safety and health in the construction and operation phases. Most of the surveyed architects were not familiar with the PtD pilot credit even though they are widely familiar with LEED, and only a few expressed interest in using the pilot credit at the time of this investigation. Research could be conducted to compare the safety performance of projects that have registered for the pilot credit to those that have not, and then identify challenges designers have faced when implementing such a credit.

5. Proposed future research

Based on the increased number of publications on PtD topics since the PtD initiative was first started by NIOSH in 2007, it can be said that interest in PtD research is evident. The analysis of 140 PtD-related journal articles that were published between 2008 and 2020 revealed six primary research topics. Table 6 shows the list of topics and provides a summary of the current achievements and shortcomings within each topic.

PtD diffusion is still at its early stage, especially in countries like the United States that have no mandatory PtD-related requirements. Barriers to PtD implementation exist in both people and processes, which include, among others, a low-level awareness of and engagement in PtD, lack of support from clients/owners and constructors due to schedule and budget concerns, lack of PtD training and resources, and lack of regulatory and legislative support. Thus, there is a need for continuous effort in both research and practice that influences and enhances the awareness and implementation of PtD in the long run. The following section provides further discussion on the review findings and points out possible directions of future research. Conducting research in these directions will eventually improve PtD research design, solidify current research findings, and strengthen the understanding of the need for, and adoption of, PtD.

5.1. Use of mixed research methods and empirical evidence

Regardless of the investigated PtD topics, a common issue is related to the method employed to conduct the research. Through a careful examination of the adopted research methods, only about one-fifth of the studies described in the reviewed articles adopted a mixed-methods approach. Case study and surveys are the two most frequently adopted research methods, used by 36% and 24% of the studies, respectively. For studies that utilized a case study as the primary research method, the findings often could not be generalized to all construction projects due to the limited number of case study projects investigated. Data collected with surveys were often based on participant perceptions or their ability to recall information, which were often retrospective. Within the reviewed articles, only the study conducted by Lingard et al. (2015) provided empirical data to reveal the significance of PtD from a prospective view. As stated by Abowitz and Toole (2009), no single method is best. It is suggested to use a mixed-methods approach for PtD research that aligns with the social sciences to improve the validity and reliability of the results by combining quantitative and qualitative approaches in research design and data collection.

A lack of empirical studies is another notable issue with PtD studies, which is in line with the findings from Tymvios et al. (2020). In addition, there is minimal longitudinal research that evaluates the impacts of OSH design decisions over the life cycles of projects, not only on the effectiveness of PtD in reducing OSH risks, but also on the impacts of other important performance indicators such as cost, schedule, productivity, quality, and their potential influence on the choice of project delivery methods. Comprehensive evaluations of PtD design decisions on multiple performance indicators using a weighted system or other evaluation methods could be conducted. To support diffusion, it is essential that information on the cost impacts of PtD be assessed. Even though prefabrication/modularization was cited as one of the most frequently used PtD industry practices in the United States, the effectiveness of construction prefabrication/modulation on improving safety performance still needs further confirmation. Additional understanding of the actual influence of implementation of PtD would help stakeholders make informed decisions regarding whether to adopt PtD or not, when to adopt it, and to what degree.

5.2. PtD regulations, legislations, education, and training

Compared to other PtD topics, fewer studies explored the topics of PtD regulations and legislations, and PtD education and training, which are two influential PtD implementation factors. To date, only limited research studies have been carried out to examine the effectiveness of existing PtD regulations, and no study was found to compare the differences in the policies and practices from different countries. For sustainable/green construction projects in the United States, little research (Behm & Pearce, 2017; Karakhan & Gambatese, 2017a) has investigated the influence of implementing the LEED PtD pilot credit on project safety performance. Further studies could be conducted to compare different sets of PtD regulations in different countries and to investigate the status of applying the PtD pilot credit (or other PtD related guidelines and regulations), including the level of awareness and interest, challenges faced, and effectiveness in OSH improvement. The incorporation of PtD in sustainable design and construction could yield significant safety and non-safety gains (Karakhan & Gambatese, 2017a; Kamas et al., 2019). Moreover, utilizing visualization methods such as VR and augmented reality (AR) to develop training courses/modules related to PtD is a topic that requires further research. An increasing number of research studies have recently attempted to develop training courses/modules related to construction site safety (Jeelani et al., 2020; Le et al., 2015; Sacks et al., 2013). However, there have been no studies that focused on developing PtD training courses/modules for designers (both architects and engineers) using VR and AR. Developing such training would help overcome one of the most notable barriers (designer lack of knowledge about safety or construction means and methods) to greater diffusion of PtD across the construction industry. Emerging technology has been shown to be crucial for improved workplace conditions and workplace safety management (Nnaji et al., 2020).

5.3. PtD related to worker health and environment, construction equipment, and temporary structures

Though a variety of PtD tools/resources, with or without the integration of construction technologies, have been developed to help designers and constructors implement the concept of PtD, the primary focus is on worker safety when constructing permanent structures. As a highly fragmented and complex industry, additional attention should be given to address worker health issues, improve construction equipment and the work environment, and enhance the safety of temporary structures. With these research studies, the industry could better understand the feasibility of implementing PtD in broader aspects to improve worksite conditions and ensure worker protection.

5.4. PtD solutions at high levels of hazard control

As shown in Tables 4 and 5, the majority of the currently available PtD tools focus on providing information on hazard identification and risk assessment instead of providing viable and workable PtD solutions for designers to use in design alternative selection or site planning. Consistent with the findings of Hardison and Hallowell (2019), further research studies are needed to examine whether the risk-based tools are practical to use on real life projects, and whether they are effective in improving lifecycle safety

Table 6
Achievements and shortcomings of previous PtD research and possible future research directions.

PtD Topic	Achievement	Shortcomings	Future Directions
Design and OSH	There is a relationship between designs and construction accidents	1) Studies relied on historical injury/fatality databases or experts' opinions; 2) Studies focused on construction safety. There is a lack of research on PtD related to health issues, and for other phases of the entire life cycle of a project; and 3) No well-established method to identify the causal relationship and to quantify the magnitude of the relationship.	1) Using a mixed-method approach to collect both quantitative and qualitative data to improve research validity and reliability; 2) Conducting longitudinal studies to evaluate the impacts of PtD design decisions; and 3) Developing a systematic and consistent method to assess the relationship between design factors and safety and health for workers over the whole life cycle of a project.
Perspectives of PtD Stakeholders	Generally positive attitudes towards PtD, a low level of PtD knowledge, and low engagement in PtD	1) Limited samples were investigated; 2) Findings were based on the perspectives of stakeholders; and 3) No empirical data were provided.	1) Collecting more representative and randomly selected samples from a wide range of PtD stakeholders; and 2) Soliciting empirical data to confirm the findings from previous PtD research in terms of stakeholder knowledge levels of PtD and their PtD practices.
PtD Implementation Barriers, Enablers and Motivators	Identified barriers (e.g., legal, economic, and contractual), enablers (e.g., requisite PtD knowledge and adequate time), and motivators (e.g., legal and regulatory factors, and ethics) related to designer and owner/client implementation of PtD	Findings were mostly based on observational or anecdotal research	Conducting quantitative studies to confirm the existence or absence of barriers, enablers, and motivators
PtD Interventions	Regulations and legislation PtD regulations worldwide	1) For research studies conducted in countries/regions that have PtD regulations, no comprehensive investigation was conducted to verify the effectiveness of regulations on OSH; and 2) No studies investigated the differences in PtD regulations that have been put in place in different countries.	1) Performing research to examine the effectiveness of PtD regulations on accident prevention; 2) Conducting comparative studies to examine the differences in PtD regulations; and 3) For researchers from a country/region that has no PtD regulations, performing studies to learn from countries that have successfully put PtD regulations in place.
	Education and training PtD education and training materials and opportunities	1) No agreement on how to incorporate the PtD concept in existing university curricula; and 2) No thorough investigation on the preference and effectiveness of PtD training methods for professionals.	Establish a systematic PtD education and training program for both students and professionals
PtD resources/ tools/ procedures and practical cases	Developed PtD resources/tools/ procedures to help designers and constructors identify hazards, assess risk levels, propose design alternatives and assist in site planning. Provided practical cases to demonstrate how to apply the PtD concept in real life scenarios.	Studies were not comprehensive: 1) The majority of the studies emphasized risk identification and assessment for very specific hazards such as falls; 2) The majority of the studies focused on permanent structures; 3) Health issues and OSH of maintenance workers were barely discussed; and 4) The impacts of PtD solutions were not well assessed.	1) More research required on health issues of workers, and on maintenance workers; 2) More research on the interactions between workers and temporary structures, construction equipment, and the work environment; 3) Developing more PtD tools that can be applied to a wide range of applications, and can assist in design optimization; and 4) Conducting more research on the impacts of PtD solutions/ resources/ tools in terms of safety, cost, productivity, and quality from practical cases.
Technologies and PtD	Identified frequently used PtD construction technologies (e.g., BIM and other visualization technology such as VP and VR) and their usage (e.g., safety reviews/checks and safety training)	Limited practical implementations	

and health for workers. Through an exploratory study, Karakhan et al. (2019) showed that construction technologies have great potential to lessen or eliminate hazards, which are the two higher levels of the hierarchy of controls and which are generally considered as the most effective means of risk control. However, only a limited number of PtD tools have the capability of addressing the most effective levels of the hierarchy of controls. It is expected that design deficiencies related to jobsite hazards and time–space-activity conflicts could be detected, lessened, or eliminated with the help of visualization tools such as BIM, VR, and geographic information systems (GIS) (Bansal, 2016), in order to achieve improved effectiveness of risk control before hazards are present on sites.

5.5. PtD knowledge integration

Currently, there is no consensus on the most effective formats and contents for delivering PtD knowledge and skill training, and how to inform designers of the available PtD resources and tools. This deficiency is a major barrier that limits widespread PtD implementation. Future research could be conducted to determine what key PtD concepts should be delivered (e.g., situational awareness, and hierarchy of controls), to identify the existing PtD education and training opportunities for both existing and future design professionals, and to propose a systematic and practical PtD educational framework in construction to improve designer awareness and their ability to ensure worker protection.

In addition, knowledge regarding PtD regulations, procedures, tools, and best practices is scattered and fragmented. There is a lack of clarity as to how to communicate the benefits of PtD implementation to legislators, professional societies, and designers and constructors globally, and how the gap in research-to-practice can be narrowed to better serve the industry. Researchers, practitioners, and legislators could benefit from future studies on how to foster effective communication of critical information. For instance, providing information on the cost effectiveness or return on investment associated with PtD could play a central role in integrating PtD knowledge, given the important role that finance plays in the adoption of new safety tools and processes. Providing a streamlined process for informing practitioners and legislators on the advancements in PtD research and practices would also improve PtD awareness and knowledge integration.

Limited studies have attempted to integrate PtD knowledge with design authoring tools to reduce or eliminate design flaws identified during the design and planning phases. With the integration and development of emerging technologies including sensors, drones, and data mining, researchers could collect more reliable safety, health, and site data. Designers and constructors would certainly benefit by having additional information, if information is exchanged in an explicit and efficient manner, and via a collaborative platform. Future research could be conducted to explore the framework of PtD information exchange, which includes what types of data to collect, in what formats, when to exchange, on what platforms, which parties are responsible for the information exchange, and how to incorporate health and safety knowledge to facilitate PtD diffusion.

6. Limitations

Similar to other review papers, the present work has some limitations. Firstly, the analysis only covered journal articles in English, using a limited set of keywords for searching within specific databases. The literature analyzed in this study could be affected by the database used and choice of keywords. Moreover, additional studies on PtD could be available in other languages. In addition, as shown in Table 1, the review only covers PtD research studies conducted in 25 countries/regions. While reviewing studies from 25 countries provides valuable insight, it should be noted that the findings presented in this study might be limited to the countries covered and those with similar characteristics. Therefore, researchers and practitioners from countries not represented in this study should be cautious when applying the findings presented in this study. Secondly, the study placed a primary focus on PtD research studies instead of industry practices. This action was largely driven by the limited resources on industry practices that are currently available (only a few reports from the United States were available and included). The lack of discussion on PtD industry practices from the perspective of different countries is another limitation of this study. As construction means and methods vary by country, so do the implementation of PtD industry practices and tools.

7. Conclusions and recommendations

The concept of PtD has attracted attention from both the construction industry and academia since it has the potential to effectively eliminate hazards and prevent worker injuries/fatalities throughout the whole life cycle of a project. This study investigates the current state-of-the-art of PtD research by examining 140 relevant peer-reviewed journal papers that were published between 2008 and 2020. This paper provides a systematic review based on descriptive and content analysis, and then compares the findings with industry practices. The study reveals that the annual number of publications shows a growing interest in the PtD concept, and researchers in the United States contributed the most PtD research.

A wide variety of PtD topics have been investigated in peer-reviewed articles by researchers worldwide since 2007. The existing PtD research studies mainly fall into six topic categories: design and OSH; perspectives of PtD stakeholders; PtD implementation barriers, enablers, and motivators; PtD interventions; PtD resources/tools/procedures and practical cases; and construction technologies in PtD research studies. The achievements and shortcomings of each topic have been discussed in the paper as well.

7.1. Recommendations for future research

Recommendations for future PtD research, which could be considered by researchers to overcome the limitations of the current studies and extend the scope of PtD research, are summarized in Table 6. The most important recommendations are highlighted below:

- Utilize empirical data and adopting mixed research methods in PtD studies to improve research quality;
- Conduct longitudinal studies to evaluate the impact of PtD design decisions on OSH, as well as on other key indicators (e.g., cost, productivity, and quality) for the entire life cycle of a project;
- Perform studies on the effects of PtD when integrated with prefabrication/modularization and sustainability (e.g., the implementation of the LEED PtD Pilot Credit);
- Conduct comparative studies that assess PtD legislation and regulations;
- Conduct studies on the best ways to promote PtD awareness and facilitate communication and collaboration between academia, industry, and governing organizations;
- Develop a PtD education and training framework for existing and future professionals;
- Conduct studies on implementing PtD concepts to address worker health issues, improve construction equipment and the work environment, and enhance safety related to temporary structures;

- Utilize technology (e.g., BIM, VR, and GIS) use and integration with PtD to address the higher levels of the hierarchy of controls; and finally
- Develop a PtD information exchange framework.

7.2. Recommendations for practitioners (practical applications)

Results presented in this paper, especially those related to the available PtD tools/resources/procedures, could benefit industry practitioners in understating the current status of PtD research and in making decisions on when and how to adopt PtD, and identify the appropriate PtD tools to use in order to improve OSH effectively.

Conflicts of Interest

The authors declare that there is no conflict of interest related to the work presented in this paper. This research was not supported by any grants from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abowitz, D. A., & Toole, T. M. (2009). Mixed method research: Fundamental issues of design, validity, and reliability in construction research. *Journal of Construction Engineering and Management, 136*(1), 108–116.

Abueisheh, Q., Manu, P., Mahamadu, A.-M., & Cheung, C. (2020). Design for safety implementation among design professionals in construction: The context of Palestine. *Safety Science, 128*, 104742. https://doi.org/10.1016/j.ssci.2020.104742.

Alomari, K., & Gambatese, J. (2016). Degrees of Connectivity Impact on Construction Worker Safety. *Construction Research Congress, 2016*, 2783–2792. https://doi.org/10.1061/9780784479827.277.

Antwi-Afari, M. F., Li, H., Yu, Y., & Kong, L. (2018). Wearable insole pressure system for automated detection and classification of awkward working postures in construction workers. *Automation in Construction, 96*, 433–441. https://doi.org/10.1016/j.autcon.2018.10.004.

Atkinson, A. R., & Westall, R. (2010). The relationship between integrated design and construction and safety on construction projects. *Construction Management and Economics, 28*(9), 1007–1017. https://doi.org/10.1080/01446193.2010.504214.

Awolusi, I., Marks, E., & Hallowell, M. (2018). Wearable technology for personalized construction safety monitoring and trending: Review of applicable devices. *Automation in Construction, 85*, 96–106.

Bansal, V. K. (2016). Potential Application Areas of GIS in Preconstruction Planning. *Journal of Professional Issues in Engineering Education and Practice, 142*(1), 06015002. https://doi.org/10.1061/(ASCE)EI.1943-5541.0000257.

Behm, M. (2005). Linking construction fatalities to the design for construction safety concept. *Safety Science, 43*(8), 589–611.

Behm, M. (2012). Safe design suggestions for vegetated roofs. *Journal of Construction Engineering and Management, 138*(8), 999–1003. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000500.

Behm, M., & Pearce, A. (2017). Occupational Safety and Health in Green Buildings: LEED PtD Pilot Credit Analysis'. *Small Study Final Report, the Center for Construction Research and Training (CPWR). Available at Http://Bit. Ly/2yKvxge*.

Behm, M., Culvenor, J., & Dixon, G. (2014). Development of safe design thinking among engineering students. *Safety Science, 63*, 1–7. https://doi.org/10.1016/j.ssci.2013.10.018.

Benjaoran, V., & Bhokha, S. (2010). An integrated safety management with construction management using 4D CAD model. *Safety Science, 48*(3), 395–403.

Bong, S., Rameezdeen, R., Zuo, J., Li, R. Y. M., & Ye, G. (2015). The designer's role in workplace health and safety in the construction industry: Post-harmonized regulations in South Australia. *International Journal of Construction Management, 15*(4), 276–287. https://doi.org/10.1080/15623599.2015.1094850.

Breslin, P. (2007). Improving OHS standards in the building and construction industry through safe design. *Journal of Occupational Health and Safety, Australia and New Zealand, 23*(1), 89.

Burlet-Vienney, D., Chinniah, Y., Bahloul, A., & Roberge, B. (2015). Occupational safety during interventions in confined spaces. *Safety Science, 79*, 19–28. https://doi.org/10.1016/j.ssci.2015.05.003.

Carter, G., & Smith, S. D. (2006). Safety hazard identification on construction projects. *Journal of Construction Engineering and Management, 132*(2), 197–205.

CDM (The Construction Design and Management Regulations). (2015). Managing health and safety in construction. https://www.hse.gov.uk/pubns/priced/l153.pdf.

Che Ibrahim, C. K. I., & Belayutham, S. (2020). A knowledge, attitude and practices (KAP) study on prevention through design: A dynamic insight into civil and structural engineers in Malaysia. *Architectural Engineering and Design Management, 16*(2), 131–149. https://doi.org/10.1080/17452007.2019.1628001.

Cheng, T., & Teizer, J. (2014). Modeling Tower Crane Operator Visibility to Minimize the Risk of Limited Situational Awareness. *Journal of Computing in Civil Engineering, 28*(3), 04014004. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000282.

Chew, M. Y., Conejos, S., & Azril, F. H. B. (2019). Design for maintainability of high-rise vertical green facades. *Building Research & Information, 47*(4), 453–467.

Chun, C. K., Li, H., & Skitmore, M. (2012). The use of virtual prototyping for hazard identification in the early design stage. *Construction Innovation, 12*(1), 29–42.

Clevenger, C., Lopez del Puerto, C., & Glick, S. (2015). Interactive BIM-enabled Safety Training Piloted in Construction Education. *Advances in Engineering Education, 4*(3), n3.

Cooke, T., Lingard, H., Blismas, N., & Stranieri, A. (2008). ToolSHeDTM: The development and evaluation of a decision support tool for health and safety in construction design. *Engineering Construction and Architectural Management*.

Cortés, J. M., Pellicer, E., & Catalá, J. (2012). Integration of Occupational Risk Prevention Courses in Engineering Degrees: Delphi Study. *Journal of Professional Issues in Engineering Education and Practice, 138*(1), 31–36. https://doi.org/10.1061/(ASCE)EI.1943-5541.0000076.

Cortes-Perez, J. P., Cortes-Perez, A., & Prieto-Muriel, P. (2020). BIM-integrated management of occupational hazards in building construction and maintenance. *Automation in Construction, 113*, 12. https://doi.org/10.1016/j.autcon.2020.103115.

CPWR (The Center for Construction Research and Training). (2018). *Topics in Construction Safety and Health – Prevention through Design (PtD): An Interdisciplinary Annotated Bibliography*.

Denyer, D., & Tranfield, D. (2009). Producing a systematic review. In *The Sage handbook of organizational research methods* (pp. 671–689). Sage Publications Ltd..

Dewlaney, K. S., & Hallowell, M. (2012). Prevention through design and construction safety management strategies for high performance sustainable building construction. *Construction Management and Economics, 30*(2), 165–177. https://doi.org/10.1080/01446193.2011.654232.

Dewlaney, K. S., Hallowell, M. R., & Fortunato Iii, B. R. (2012). Safety risk quantification for high performance sustainable building construction. *Journal of Construction Engineering and Management, 138*(8), 964–971. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000504.

Dharmapalan, V., Gambatese, J. A., Fradella, J., & Moghaddam Vahed, A. (2015). Quantification and assessment of safety risk in the design of multistory buildings. *Journal of Construction Engineering and Management, 141*(4). https://doi.org/10.1061/(ASCE)CO.1943-7862.0000952.

Ding, L., Zhong, B., Wu, S., & Luo, H. (2016). Construction risk knowledge management in BIM using ontology and semantic web technology. *Safety Science, 87*, 202–213.

Directive, C. (1992). 92/57/EEC of 24 June 1992 on the implementation of minimum safety and health requirements at temporary or mobile construction sites (eighth individual Directive within the meaning of Article 16 (1) of Directive 89/391/EEC). *Official Journal L, 245*(26), 08.

Dodge Data & Analytics. (2017). *Safety Management in the Construction Industry 2017*. Dodge Data & Analytics. http://images.marketing.construction.com/Web/DDA/%7B22eaad80-f9af-475b-9717-927c4cac5f8e%7D_Safety_Management_2017_SmartMarket_Report.pdf.

Dodge Data & Analytics. (2020). *Safety Management in the Construction Industry 2020*. Dodge Data & Analytics,. https://proddrupalcontent.construction.com/s3fs-public/Safety-Management-2020-SmartMarket-Report-27Jan.pdf.

Eastman, C. M., Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors*. John Wiley & Sons.

Edirisinghe, R., Stranieri, A., & Blismas, N. (2016). Information visualisation for the wicked problem of safe construction design. *Architectural Engineering and Design Management, 12*(4), 296–310. https://doi.org/10.1080/17452007.2016.1182890.

Esmaeili, B., & Hallowell, M. (2013). Integration of safety risk data with highway construction schedules. *Construction Management and Economics, 31*(6), 528–541. https://doi.org/10.1080/01446193.2012.739288.

Esmaeili, B., Hallowell, M. R., & Rajagopalan, B. (2015). Attribute-Based Safety Risk Assessment. I: Analysis at the Fundamental Level. *Journal of Construction Engineering and Management, 141*(8), 04015021. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000980.

Ezisi, U., & Issa, M. H. (2019). Case study application of prevention through design to enhance workplace safety and health in manitoba heavy construction projects. *Canadian Journal of Civil Engineering, 46*(2), 124–133. https://doi.org/10.1139/cjce-2017-0454.

Fargnoli, M., Lombardi, M., Haber, N., & Guadagno, F. (2018). Hazard function deployment: A QFD-based tool for the assessment of working tasks–a practical study in the construction industry. *International Journal of Occupational Safety and Ergonomics*, 1–22.

Fortunato, B. R., Hallowell, M. R., Behm, M., & Dewlaney, K. (2012). Identification of Safety Risks for High-Performance Sustainable Construction Projects. *Journal of Construction Engineering and Management, 138*(4), 499–508. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000446.

Frijters, A. C., & Swuste, P. H. (2008). Safety assessment in design and preparation phase. *Safety Science, 46*(2), 272–281.

Fung, I. W., Tam, V. W., Lo, T. Y., & Lu, L. L. (2010). Developing a risk assessment model for construction safety. *International Journal of Project Management, 28*(6), 593–600.

Gambatese, J. A., Behm, M., & Rajendran, S. (2008). Design's role in construction accident causality and prevention: Perspectives from an expert panel. *Safety Science, 46*(4), 675–691.

Gambatese, J. A., Gibb, A. G., Brace, C., & Tymvios, N. (2017). Motivation for Prevention through Design: Experiential Perspectives and Practice. *Practice Periodical on Structural Design and Construction, 22*(4). https://doi.org/10.1061/(ASCE)SC.1943-5576.0000335.

Gambatese, J., Gibb, A., Bust, P., & Behm, M. (2017). Expanding Prevention through Design (PtD) in practice: Innovation, change, and a path forward. *Journal of Construction Project Management and Innovation, 7*(2), 1995–2006.

Gambatese, J. A., Michael Toole, T., & Abowitz, D. A. (2017). Owner perceptions of barriers to prevention through design diffusion. *Journal of Construction Engineering and Management, 143*(7). https://doi.org/10.1061/(ASCE)CO.1943-7862.0001296.

Gangolells, M., Casals, M., Forcada, N., Fuertes, A., & Roca, X. (2013). Model for enhancing integrated identification, assessment, and operational control of on-site environmental impacts and health and safety risks in construction firms. *Journal of Construction Engineering and Management, 139*(2), 138–147. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000579.

Gangolells, M., Casals, M., Forcada, N., Roca, X., & Fuertes, A. (2010). Mitigating construction safety risks using prevention through design. *Journal of Safety Research, 41*(2), 107–122. https://doi.org/10.1016/j.jsr.2009.10.007.

Goh, Y. M., & Chua, S. (2016). Knowledge, attitude and practices for design for safety: A study on civil structural engineers. *Accident Analysis and Prevention, 93*, 260–266. https://doi.org/10.1016/j.aap.2015.09.023.

Goh, Y. M., & Guo, B. H. (2018). FPSWizard: A web-based CBR-RBR system for supporting the design of active fall protection systems. *Automation in Construction, 85*, 40–50.

Golabchi, A., Han, S., & AbouRizk, S. (2018). A simulation and visualization-based framework of labor efficiency and safety analysis for prevention through design and planning. *Automation in Construction, 96*, 310–323. https://doi.org/10.1016/j.autcon.2018.10.001.

Golabchi, A., Han, S., Seo, J., Han, S., Lee, S., & Al-Hussein, M. (2015). An Automated Biomechanical Simulation Approach to Ergonomic Job Analysis for Workplace Design. *Journal of Construction Engineering and Management, 141*(8), 04015020. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000998.

Hallowell, M. R., & Hansen, D. (2016). Measuring and improving designer hazard recognition skill: Critical competency to enable prevention through design. *Safety Science, 82*, 254–263. https://doi.org/10.1016/j.ssci.2015.09.005.

Hardison, D., & Hallowell, M. (2019). Construction hazard prevention through design: Review of perspectives, evidence, and future objective research agenda. *Safety Science, 120*, 517–526.

Hare, B., Kumar, B., & Campbell, J. (2020). Impact of a multi-media digital tool on identifying construction hazards under the UK construction design and management regulations. *Journal of Information Technology in Construction, 25*, 482–499. https://doi.org/10.36680/j.itcon.2020.028.

Ho, C., Lee, H. W., & Gambatese, J. A. (2020). Application of Prevention through Design (PtD) to improve the safety of solar installations on small buildings. *Safety Science, 125*, 104633. https://doi.org/10.1016/j.ssci.2020.104633.

Hossain, M. A., Abbott, E. L. S., Chua, D. K. H., Nguyen, T. Q., & Goh, Y. M. (2018). Design-for-Safety knowledge library for BIM-integrated safety risk reviews. *Automation in Construction, 94*, 290–302. https://doi.org/10.1016/j.autcon.2018.07.010.

Ibrahim, C. K. I. C., Manu, P., Belayutham, S., Mahamadu, A. M., & Antwi-Afari, M. F. (2022). Design for safety (DfS) practice in construction engineering and management research: A review of current trends and future directions. *Journal of Building Engineering, 52*, 104352.

Jeelani, I., Han, K., & Albert, A. (2020). Development of virtual reality and stereo-panoramic environments for construction safety training. *Engineering, Construction and Architectural Management, 27*(8), 1853–1876. https://doi.org/10.1108/ECAM-07-2019-039.

Jin, Z., Gambatese, J., Liu, D., & Dharmapalan, V. (2019). Using 4D BIM to assess construction risks during the design phase. *Engineering, Construction and Architectural Management, 26*(11), 2637–2654. https://doi.org/10.1108/ECAM-09-2018-0379.

Kamas, W. M., Hasan, A. A., & Fadel, A. H. (2019). Economic Benefits for the Application of Standards of Sustainability in Construction Projects. *Journal of Engineering, 25*(3), 117–126.

Karakhan, A. A., & Gambatese, J. A. (2017a). Integrating Worker Health and Safety into Sustainable Design and Construction: Designer and Constructor Perspectives. *Journal of Construction Engineering and Management, 143*(9), 04017069. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001379.

Karakhan, A. A., & Gambatese, J. A. (2017b). Identification, quantification, and classification of potential safety risk for sustainable construction in the United States. *Journal of Construction Engineering and Management, 143*(7). https://doi.org/10.1061/(ASCE)CO.1943-7862.0001302.

Karakhan, A. A., & Gambatese, J. A. (2017c). Safety Innovation and Integration in High-Performance Designs: Benefits, Motivations, and Obstacles. *Practice Periodical on Structural Design and Construction, 22*(4), 04017018.

Karakhan, A., Rajendran, S., & Gambatese, J. (2018). Validation of time-safety influence curve using empirical safety and injury data-Poisson regression. *In Construction Research Congress, 2018*, 389–399.

Karakhan, A., Xu, Y., Nnaji, C., & Alsaffar, O. (2019). Technology alternatives for workplace safety risk mitigation in construction: Exploratory study. In *Advances in informatics and computing in civil and construction engineering* (pp. 823–829). Springer.

Kikwasi, G. J., & Smallwood, J. (2016). Construction health and safety (H&S) performance improvement : Are clients and consultants playing their part? *Journal of Construction Project Management and Innovation, 6*(1), 1339–1352.

Kim, H.-J., Ahn, H.-S., & Kim, W.-Y. (2011). 3-D temporary facility visualization using bim (building information modeling) technology. *Journal of Construction Engineering and Project Management, 1*(2), 37–42.

Kim, K., Cho, Y., & Zhang, S. (2016). Integrating work sequences and temporary structures into safety planning: Automated scaffolding-related safety hazard identification and prevention in BIM. *Automation in Construction, 70*, 128–142.

Kim, I., Lee, Y., & Choi, J. (2020). BIM-based Hazard Recognition and Evaluation Methodology for Automating Construction Site Risk Assessment. *Applied Sciences, 10*(7), 2335.

Kim, S., Nussbaum, M. A., & Jia, B. (2011). Low back injury risks during construction with prefabricated (panelised) walls: Effects of task and design factors. *Ergonomics, 54*(1), 60–71. https://doi.org/10.1080/00140139.2010.535024.

Larsen, G. D., & Whyte, J. (2013). Safe construction through design: Perspectives from the site team. *Construction Management and Economics, 31*(6), 675–690. https://doi.org/10.1080/01446193.2013.798424.

Le, Q. T., Pedro, A., & Park, C. S. (2015). A social virtual reality based construction safety education system for experiential learning. *Journal of Intelligent & Robotic Systems, 79*(3), 487–506.

Lee, Y., Kim, I., & Choi, J. (2020). Development of BIM-Based Risk Rating Estimation Automation and a Design-for-Safety Review System. *Applied Sciences, 10*(11), 21. https://doi.org/10.3390/app10113902.

Lingard, H. (2013). Prevention through design: Trades-offs in reducing occupational health and safety risk for the construction and operation of a facility. *Built Environment Project and Asset Management, 3*(1), 7–23. https://doi.org/10.1108/BEPAM-06-2012-0036.

Lingard, H. (2018). Making the invisible visible. *Engineering, Construction and Architectural Management, 25*(1), 39–61. https://doi.org/10.1108/ECAM-07-2016-0174.

Lingard, H. C., Cooke, T., & Blismas, N. (2012). Designing for construction workers' occupational health and safety: A case study of socio-material complexity. *Construction Management and Economics, 30*(5), 367–382. https://doi.org/10.1080/01446193.2012.667569.

Lingard, H., Saunders, L., Pirzadeh, P., Blismas, N., Kleiner, B., & Wakefield, R. (2015). The relationship between pre-construction decision-making and the effectiveness of risk control: Testing the time-safety influence curve. *Engineering, Construction and Architectural Management, 22*(1), 108–124. https://doi.org/10.1108/ECAM-08-2013-0074.

Liu, H., Jazayeri, E., & Dadi, G. B. (2017). Establishing the influence of owner practices on construction safety in an operational excellence model. *Journal of Construction Engineering and Management, 143*(6), 04017005.

Lopez-Arquillos, A., Rubio-Romero, J. C., & Martinez-Aires, M. D. (2015). Prevention through Design (PtD). The importance of the concept in Engineering and Architecture university courses. *Safety Science, 73*, 8–14. https://doi.org/10.1016/j.ssci.2014.11.006.

Malekitabar, H., Ardeshir, A., Sebt, M. H., & Stouffs, R. (2016). Construction safety risk drivers: A BIM approach. *Safety Science, 82*, 445–455.

Mann Iii, J. A. (2008). Education Issues in Prevention through Design. *Journal of Safety Research, 39*(2), 165–170. https://doi.org/10.1016/j.jsr.2008.02.009.

Manu, P., Poghosyan, A., Mshelia, I. M., Iwo, S. T., Mahamadu, A.-M., & Dziekonski, K. (2019). Design for occupational safety and health of workers in construction in developing countries: A study of architects in Nigeria. *International Journal of Occupational Safety and Ergonomics, 25*(1), 99–109. https://doi.org/10.1080/10803548.2018.1485992.

Marks, E. D., Cheng, T., & Teizer, J. (2013). Laser scanning for safe equipment design that increases operator visibility by measuring blind spots. *Journal of Construction Engineering and Management, 139*(8), 1006–1014.

Marks, E. D., & Teizer, J. (2013). Method for testing proximity detection and alert technology for safe construction equipment operation. *Construction Management and Economics, 31*(6), 636–646.

Martínez-Aires, M. D., Rubio Gámez, M. C., & Gibb, A. (2010). Prevention through design: The effect of European Directives on construction workplace accidents. *Safety Science, 48*(2), 248–258. https://doi.org/10.1016/j.ssci.2009.09.004.

Martínez-Aires, M. D., Rubio Gámez, M. C., & Gibb, A. (2016). The impact of occupational health and safety regulations on prevention through design in construction projects: Perspectives from Spain and the United Kingdom. *Work, 53*(1), 181–191.

McGraw-Hill Construction. (2011). *Prefabrication and Modularization: Increasing Productivity in the Construction Industry*. McGraw-Hill Construction. http://images.marketing.construction.com/Web/DDA/%7Bd6edb745-0e7d-433c-b3b4-a1a01d1d90cf%7D_Prefabrication___Modularization_SMR.pdf.

Melzner, J., Zhang, S., Teizer, J., & Bargstadt, H.-J. (2013). A case study on automated safety compliance checking to assist fall protection design and planning in building information models. *Construction Management and Economics, 31*(6), 661–674. https://doi.org/10.1080/01446193.2013.780662.

Mendeloff, J., & Staetsky, L. (2014). Occupational fatality risks in the United States and the United Kingdom. *American Journal of Industrial Medicine, 57*(1), 4–14.

Merriam-Webster.com (2022). Infographic. Merriam-Webster, https://www.merriam-webster.com/dictionary/infographic. Accessed May 31, 2022.

Michael Toole, T., & Carpenter, G. (2013). Prevention through Design as a Path toward Social Sustainability. *Journal of Architectural Engineering, 19*(3), 168–173. https://doi.org/10.1061/(ASCE)AE.1943-5568.0000107.

Michael Toole, T., & Erger, K. (2019). Prevention through Design: Promising or Perilous? *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 11*(1), 04518023. https://doi.org/10.1061/(ASCE)LA.1943-4170.0000284.

Moselhi, O., Bardareh, H., & Zhu, Z. (2020). Automated data acquisition in construction with remote sensing technologies. *Applied Sciences, 10*(8), 2846.
Mzyece, D., Ndekugri Issaka, E., & Ankrah Nii, A. (2019). Building information modelling (BIM) and the CDM regulations interoperability framework. *Engineering, Construction and Architectural Management, 26*(11), 2682–2704. https://doi.org/10.1108/ECAM-10-2018-0429.
Nath, N. D., Akhavian, R., & Behzadan, A. H. (2017). Ergonomic analysis of construction worker's body postures using wearable mobile sensors. *Applied Ergonomics, 62*, 107–117. https://doi.org/10.1016/j.apergo.2017.02.007.
National Institute for Occupational Safety and Health (NIOSH). (2013). *Prevention through Design*. https://www.cdc.gov/niosh/topics/ptd/default.html.
National Institute for Occupational Safety and Health (NIOSH). (2015). *Hierarchy of Controls*. https://www.cdc.gov/niosh/topics/hierarchy/default.html.
Ning, X., Qi, J., Wu, C., & Wang, W. (2018). A tri-objective ant colony optimization based model for planning safe construction site layout. *Automation in Construction, 89*, 1–12.
Nnaji, C., Awolusi, I., Park, J. W., & Albert, A. (2021). Wearable sensing devices: Towards the development of a personalized system for construction safety and health risk mitigation. *Sensors, 21*(3), 682.
Nnaji, C., Gambatese, J., Karakhan, A., & Osei-Kyei, R. (2020). Development and application of safety technology adoption decision-making tool. *Journal of Construction Engineering and Management, 146*(4), 04020028.
Nnaji, C., Okpala, I., & Awolusi, I. (2020). Wearable sensing devices: Potential impact & current use for incident prevention. *Professional Safety, 65*(04), 16–24.
Nussbaum, M. A., Shewchuk, J. P., Kim, S., Seol, H., & Guo, C. (2009). Development of a decision support system for residential construction using panellised walls: Approach and preliminary results. *Ergonomics, 52*(1), 87–103.
Öney-Yazıcı, E., & Dulaimi, M. F. (2015). Understanding designing for construction safety: The interaction between confidence and attitude of designers and safety culture. *Architectural Engineering and Design Management, 11*(5), 325–337. https://doi.org/10.1080/17452007.2014.895697.
Patrucco, M., Bersano, D., Cigna, C., & Fissore, F. (2010). Computer image generation for job simulation: An effective approach to occupational Risk Analysis. *Safety Science, 48*(4), 508–516. https://doi.org/10.1016/j.ssci.2009.12.025.
Penaloza, G. A., Saurin, T. A., & Formoso, C. T. (2017). Identification and assessment of requirements of temporary edge protection systems for buildings. *International Journal of Industrial Ergonomics, 58*, 90–108. https://doi.org/10.1016/j.ergon.2017.02.005.
Pirzadeh, P., & Lingard, H. (2017). Understanding the Dynamics of Construction Decision Making and the Impact on Work Health and Safety. *Journal of Management in Engineering, 33*(5), 05017003. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000532.
Poghosyan, A., Manu, P., Mahdjoubi, L., Gibb, A. G. F., Behm, M., & Mahamadu, A.-M. (2018). Design for safety implementation factors: A literature review. *Journal of Engineering, Design and Technology, 16*(5), 783–797. https://doi.org/10.1108/JEDT-09-2017-0088.
Qi, J., Issa, R. R. A., Olbina, S., & Hinze, J. (2014). Use of Building Information Modeling in Design to Prevent Construction Worker Falls. *Journal of Computing in Civil Engineering, 28*(5), A4014008. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000365.
Rodrigues, F., Antunes, F., & Matos, R. (2021). Safety plugins for risks prevention through design resourcing BIM. *Construction Innovation-England, 21*(2), 244–258. https://doi.org/10.1108/ci-12-2019-0147.
Rubio-Romero, J. C., Suárez-Cebador, M., & Abad, J. (2014). Modeling injury rates as a function of industrialized versus on-site construction techniques. *Accident Analysis & Prevention, 66*, 8–14. https://doi.org/10.1016/j.aap.2014.01.005.
Rwamamara, R., Norberg, H., Olofsson, T., & Lagerqvist, O. (2010). Using visualization technologies for design and planning of a healthy construction workplace. *Construction Innovation, 10*(3), 248–266.
Sacks, R., Perlman, A., & Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management and Economics, 31*(9), 1005–1017.
Sacks, R., Whyte, J., Swissa, D., Raviv, G., Zhou, W., & Shapira, A. (2015). Safety by design: Dialogues between designers and builders using virtual reality. *Construction Management and Economics, 33*(1), 55–72.
Saunders, L. W. (2016). International benchmarking for performance improvement in construction safety and health. *Benchmarking: An International Journal, 23*(4), 916–936. https://doi.org/10.1108/BIJ-11-2013-0105.
Saunders, L. W., Kleiner, B. M., McCoy, A. P., Ellis, K. P., Smith-Jackson, T., & Wernz, C. (2017). Developing an inter-organizational safety climate instrument for the construction industry. *Safety Science, 98*, 17–24. https://doi.org/10.1016/j.ssci.2017.04.003.
Seo, J. W., & Choi, H. H. (2008). Risk-based safety impact assessment methodology for underground construction projects in Korea. *Journal of Construction Engineering and Management, 134*(1), 72–81. https://doi.org/10.1061/(ASCE)0733-9364(2008)134:1(72).
Sinyai, C., & Choi, S. (2020). Fifteen years of American construction occupational safety and health research. *Safety Science, 131*, 104915.
Szymberski, R. T. (1997). Construction project safety planning. *Tappi Journal, 80*(11), 69–74.
Teo, A. L. E., Ofori, G., Tjandra, I. K., & Kim, H. (2016). Design for safety: theoretical framework of the safety aspect of BIM system to determine the safety index. *Construction Economics and Building, 16*(4). 1–1.
Tixier, A. J.-P., Albert, A., & Hallowell, M. R. (2018). Proposing and Validating a New Way of Construction Hazard Recognition Training in Academia: Mixed-Method Approach. *Practice Periodical on Structural Design and Construction, 23*(1), 04017027. https://doi.org/10.1061/(ASCE)SC.1943-5576.0000347.
Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2017). Construction Safety Clash Detection: Identifying Safety Incompatibilities among Fundamental Attributes using Data Mining. *Automation in Construction, 74*, 39–54. https://doi.org/10.1016/j.autcon.2016.11.001.
Toh, Y. Z., Goh, Y. M., & Guo, B. H. W. (2017). Knowledge, Attitude, and Practice of Design for Safety: Multiple Stakeholders in the Singapore Construction Industry. *Journal of Construction Engineering and Management, 143*(5). https://doi.org/10.1061/(ASCE)CO.1943-7862.0001279.
Toole, T. M. (2005). Increasing engineers' role in construction safety: Opportunities and barriers. *Journal of Professional Issues in Engineering Education and Practice, 131*(3), 199–207.
Toole, T. M., & Gambatese, J. (2008). The trajectories of prevention through design in construction. *Journal of Safety Research, 39*(2), 225–230.
Toole, T. M., Gambatese, J. A., & Abowitz, D. A. (2017). Owners' Role in Facilitating Prevention through Design. *Journal of Professional Issues in Engineering Education and Practice, 143*(1). https://doi.org/10.1061/(ASCE)EI.1943-5541.0000295.
Tymvios, N., & Gambatese, J. A. (2016a). Direction for Generating Interest for Design for Construction Worker Safety - A Delphi Study. *Journal of Construction Engineering and Management, 142*(8). https://doi.org/10.1061/(ASCE)CO.1943-7862.0001134.
Tymvios, N., & Gambatese, J. A. (2016b). Perceptions about design for construction worker safety: Viewpoints from contractors, designers, and university facility owners. *Journal of Construction Engineering and Management, 142*(2). https://doi.org/10.1061/(ASCE)CO.1943-7862.0001067.
Tymvios, N., Hardison, D., Behm, M., Hallowell, M., & Gambatese, J. (2020). Revisiting Lorent and Szymberski: Evaluating how research in prevention through design is interpreted and cited. *Safety Science, 131*, 104927. https://doi.org/10.1016/j.ssci.2020.104927.
U.S. Geological Survey (USGS). (2022). What is remote sensing and what is it used for? What is remote sensing and what is it used for? | U.S. Geological Survey. Retrieved May 31, 2022, from https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used#:~:text=Remote%20sensing%20is%20the%20process, sense%22%20things%20about%20the%20Earth.
U.S. Green Building Council (USGBC). (2015). *New LEED Pilot Credit: Prevention through design*. https://www.usgbc.org/articles/new-leed-pilot-credit-prevention-through-design.
Umer, W., Li, H., Lu, W., Szeto, G. P. Y., & Wong, A. Y. (2018). Development of a tool to monitor static balance of construction workers for proactive fall safety management. *Automation in Construction, 94*, 438–448.
Votano, S., & Sunindijo, R. Y. (2014). Client Safety Roles in Small and Medium Construction Projects in Australia. *Journal of Construction Engineering and Management, 140*(9), 04014045. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000899.
Wang, G. G. (2002). Definition and review of virtual prototyping. *Journal of Computing and Information Science in Engineering, 2*(3), 232–236.
Weidman, J. E., Dickerson, D. E., & Koebel, C. T. (2016). Technology Champions: A Theory-Based Intervention to Improve Adoption of Occupational Safety Innovations. *International Journal of Construction Education and Research, 12*(3), 193–207. https://doi.org/10.1080/15578771.2015.1118417.
Weinstein, M., Gambatese, J., & Hecker, S. (2005). Can design improve construction safety?: Assessing the impact of a collaborative safety-in-design process. *Journal of Construction Engineering and Management, 131*(10), 1125–1134.
Wong, F. K., Chan, A. P., Yam, M. C., Wong, E. Y., Kenny, T., Yip, K. K., & Cheung, E. (2009). Findings from a research study of construction safety in Hong Kong: Accidents related to fall of person from height. *Journal of Engineering, Design and Technology*..
Wu, C., Fang, D., & Li, N. (2015). Roles of owners' leadership in construction safety: The case of high-speed railway construction projects in China. *International Journal of Project Management, 33*(8), 1665–1679. https://doi.org/10.1016/j.ijproman.2015.07.005.
Young-Corbett, D. E. (2014). Prevention through design: Health hazards in asphalt roofing. *Journal of Construction Engineering and Management, 140*(9). https://doi.org/10.1061/(ASCE)CO.1943-7862.0000892.
Yuan, J., Li, X., Xiahou, X., Tymvios, N., Zhou, Z., & Li, Q. (2019). Accident prevention through design (PtD): Integration of building information modeling and PtD knowledge base. *Automation in Construction, 102*, 86–104. https://doi.org/10.1016/j.autcon.2019.02.015.
Zhang, S., Boukamp, F., & Teizer, J. (2015). Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA). *Automation in Construction, 52*, 29–41. https://doi.org/10.1016/j.autcon.2015.02.005.
Zhang, S., Sulankivi, K., Kiviniemi, M., Romo, I., Eastman, C. M., & Teizer, J. (2015). BIM-based fall hazard identification and prevention in construction safety planning. *Safety Science, 72*, 31–45.
Zhang, S., Teizer, J., Lee, J.-K., Eastman, C. M., & Venugopal, M. (2013). Building Information Modeling (BIM) and Safety: Automatic Safety Checking of Construction Models and Schedules. *Automation in Construction, 29*, 183–195. https://doi.org/10.1016/j.autcon.2012.05.006.
Zhou, Z., Goh, Y. M., & Li, Q. (2015). Overview and analysis of safety management studies in the construction industry. *Safety Science, 72*, 337–350.
Zhou, Z., Irizarry, J., & Li, Q. (2013). Applying advanced technology to improve safety management in the construction industry: A literature review. *Construction Management and Economics, 31*(6), 606–622.

Zhou, W., Whyte, J., & Sacks, R. (2012). Construction safety and digital design: A review. *Automation in Construction, 22*, 102–111.
Zou, P. X., Redman, S., & Windon, S. (2008). Case studies on risk and opportunity at design stage of building projects in Australia: Focus on safety. *Architectural Engineering and Design Management, 4*(3–4), 221–238.

Ziyu Jin, Ph.D. is a lecturer in the Department of Civil, Construction and Environmental Engineering at the University of New Mexico. Dr. Jin holds a B.S. degree in Construction Management from Suzhou University of Science and Technology, an M.S. degree in Engineering Management from Syracuse University, and a Ph.D. in Civil Engineering from Oregon State University. Her research is focused on construction safety and health, temporary structures, and technology applications in construction. Dr. Jin can be reached via email at zjin@unm.edu.

John Gambatese, Ph.D. is a Professor in the School of Civil and Construction Engineering at Oregon State University. Dr. Gambatese's educational background includes Bachelor and Master of Science degrees in Civil Engineering from the University of California at Berkeley, and a PhD in Civil Engineering from the University of Washington. He has worked in industry as a structural engineer in San Francisco, and as a project engineer for a construction management firm in Seattle. Dr. Gambatese's research interests are in the broad area of construction engineering and management, and more specifically related to safety, prevention through design, work zone safety, risk and reward, innovation, and sustainable and lean practices. He is a Fellow in the American Society of Civil Engineers (ASCE), and a member of the American Society of Safety Professionals (ASSP). He is a licensed Professional Civil Engineer in California. Dr. Gambatese can be reached via email at john.gambatese@oregonstate.edu.

Ali Karakhan, Ph.D. is a faculty member at University of Baghdad. Dr. Karakhan holds a B.S. degree in Building and Construction Engineering from University of Technology, an M.S. degree in Civil Engineering from University of Baghdad, and a Ph.D. in Civil Engineering from Oregon State University. Dr. Karakhan is a member of the CI Construction Safety Committee within the American Society of Civil Engineers (ASCE). His area of research interests includes construction site safety, workforce development in construction, sustainable design and construction, technology adoption in construction. Dr. Karakhan can be reached via email at karakhan@uobaghdad.edu.iq.

Chukwuma Nnaji, Ph.D. is an assistant professor in the Department of Construction Science at Texas A&M University. Dr. Nnaji holds a B.S. degree in Building Engineering from Imo State University, an M.S. degree in Civil Engineering from Oregon State University, and a Ph.D. in Civil Engineering from Oregon State University. His research is focused on construction safety, technology adoption and diffusion, and organizational behavior. Dr. Nnaji can be reached at cnnaji@tamu.edu.

Journal of Safety Research 84 (2023) 7–17

Contents lists available at ScienceDirect

Journal of Safety Research

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

An investigation of brake failure related crashes and injury severity on mountainous roadways in Wyoming

Muhammad Tahmidul Haq [a,*], Vincent-Michael Kwesi Ampadu [b], Khaled Ksaibati [c]

[a] Wyoming Technology Transfer Center, University of Wyoming, 1000 E. University Ave., EN 3029, Laramie, WY 82071, USA
[b] Department of Civil & Architectural Engineering, University of Wyoming, 1000 E University Avenue, Laramie, WY 82071, USA
[c] Wyoming Technology Transfer Center, 1000 E. University Ave., Dept. 3295, Laramie, WY 82071, USA

ARTICLE INFO

Article history:
Received 16 March 2022
Received in revised form 20 July 2022
Accepted 17 October 2022
Available online 31 October 2022

Keywords:
Traffic safety
Brake failure
Injury severity
Occupant protection
Bayesian binary logit model

ABSTRACT

Introduction: Although the braking system plays a key role in a safe and smooth vehicular operation, it has not been given proper attention and hence brake failures are still underrepresented in traffic safety. The current body of literature on brake failure-related crashes is very limited. Moreover, no previous study was found to extensively investigate the factors associated with brake failures and the corresponding injury severity. This study aims to fill this knowledge gap by examining brake failure-related crashes and assessing the factors associated with the corresponding occupant injury severity. *Method:* The study first performed a Chi-square analysis to examine the relationship among brake failure, vehicle age, vehicle type, and grade type. Three hypotheses were formulated to investigate the associations between the variables. Based on the hypotheses, vehicles aged more than 15 years, trucks, and downhill grade segments seemed to be highly associated with brake failure occurrences. The study also applied the Bayesian binary logit model to quantify the significant impacts of brake failures on occupant injury severity and identified various vehicle, occupants, crash, and roadway characteristics. *Conclusions and Practical Applications:* Based on the findings, several recommendations regarding enhancing statewide vehicle inspection regulation were outlined.

1. Introduction

Of all the features constituting a vehicle, the braking system is likely to be the most important. Its function is to enable the driver to control the vehicle speed when the need arises in order to protect the vehicle, driver, and other road users from crashes that might be fatal. Vehicle stability and operation can deteriorate significantly by defective brakes. Although the failures of modern vehicle braking systems appears to be rare, their performance can degrade constantly as vehicles age. In 2018, the National Motor Crash Causation Survey (NMVCCS) reported that of the approximately 44,000 vehicle defect-related crashes, nearly 22% (10,000) were due to defective brakes (USDOT, 2018). The distributions of other vehicle defect types reported were tires/wheels-related (accounting for about 35%), steering/suspension/transmission/engine (accounting for about 3%), and other/unknown vehicle-related problems (accounting for about 40%). Being unable to stop is almost impossible to avoid a crash when it occurs. Considering the fact, the National Traffic and Motor Vehicle Safety Act authorizes the National Highway Traffic Safety Administration (NHTSA) to investigate issues relating to motor-vehicle safety, and requires manufacturers to notify NHTSA of all safety-related defects (e.g., brakes, tires, lighting) involving unreasonable risk of a crash, death, or injury (USDOT, 2020). Therefore, there is a need to conduct research on investigating brake failure-related crashes and the corresponding injury severity.

Wyoming with its extensive network of mountainous roads has the second-highest traffic fatality rate (21.2 death per 100,000 population) in the nation (NHTSA, 2019). A significant portion of Wyoming roads goes through mountainous and rolling terrain, resulting in severe vertical grades and horizontal curves. For instance, Interstate 80 (I-80) provides a major corridor connection between the west coast and major cities in the east. About 9% of I-80 in Wyoming (in both directions) is within vertical grades of more than 3%, where certain sections reach grades of close to 7%. These conditions with a high elevation tend to accelerate the vehicle defects quicker than the normal condition. Crashes involving defective brakes are a growing concern in Wyoming because of the presence of steep downgrades, which accelerates heavy vehicles' brake temperatures exceeding the critical temperature, leading to a higher level of injury severity and significant economic

* Corresponding author.
E-mail address: mhaq@uwyo.edu (M.T. Haq).

https://doi.org/10.1016/j.jsr.2022.10.003
0022-4375/

impact (Moomen et al., 2019, 2020; Rezapour, Moomen, & Ksaibati, 2019). Among the vehicle defect-related crashes in Wyoming, brake failures contributed about 25% based on the Wyoming Department of Transportation (WYDOT) 10 years of crash reports (2010–2019). This study aims to provide some insights to fulfill the knowledge gap by analyzing brake defect-related crashes in Wyoming.

2. Literature review

Crashes with vehicle defects tend to have consequences in terms of property damage, loss of goods, and loss of life. Yet, a limited number of studies have been found investigating vehicle defects and associated safety implications. Previous studies reported brake and tire failures as the most significant factors in mechanical failures resulting in crashes (Automobil, 2015; Schoor, Niekerk, & Grobbelaar, 2001; Solah et al., 2017). In order to explore the causes of brake failure, a study administered a structured questionnaire (Owusu-Ansah, Alhassan, Frimpong, & Agyemang, 2014). The survey results indicated that overheating of the brake assembly due to prolonging the application of the brakes is the main reason behind brake failures. While investigating the effect of brake failure on road traffic, Oduru (2012) concluded that brake failure could result in a fatal crash and, hence, the vehicle should be inspected regularly to reduce brake failure. Das, Dutta, and Geedipally (2021) examined the association between crash severity and vehicle defect types by applying a Bayesian data mining approach and found that vehicle age is associated with severe injury crashes resulting from defective brakes and worn tires. In order to detect and diagnose car brake failure, Ibitayo, Mohammed, Rabiu, and Abdulrahman (2016) developed Expert System (ES), one of the leading Artificial Intelligence techniques that provides input and output equations in assisting mechanical technicians for car brake failure detection and diagnosis via mathematical Differential Equations in form of Dynamic Control System (DCS).

Trucks carrying huge loads are more prone to brake failures, especially while descending downgrades because of the generation of excessive heat on brakes. Yan and Xu (2018) showed that the brake temperature is positively correlated to the truck weight and the percentage and length of the downgrade. Dinh, Vu, Mcllroy, Plant, and Stanton (2020) developed Systems Theoretic Accident Model and Process and its corresponding Causal Analysis using Systems Theory (STAMP-CAST) model to analyze downhill truck crashes caused by a brake failure. The results indicated the driver's inexperience, together with the truck's low quality and severe road conditions, as potential factors directly leading to the corresponding crashes. In order to improve the safety performance of heavy commercial vehicles when a braking failure occurs under turning conditions, Lu, Wang, and Zhang (2020) carried out the Trucksim-Simulink joint simulation and hardware in the loop experiment to jointly verify the control effect of the brake. The results showed that the influence of vehicle brake failure on vehicle driving stability is reduced, the performance of vehicle brake under long and downhill conditions and turning conditions are improved, and the safety stability of vehicle brake failure in hardware loop is improved, which meets the requirements of national standards for vehicle braking. While examining the injury severity resulting from brake failure, Wang and Prato (2019) found a positive relation of truck brake failure with a significant rise in the fatality probability (14.6%).

Based on the previous studies discussed, vehicle condition (e.g., vehicle age), vehicle configuration (e.g., vehicle type), and roadway geometry (e.g., vertical grade) were reported as the most potential factors associated with brake failure related crashes. On the other hand, occupant, vehicle, crash, and roadway characteristics were frequently utilized in the literature to investigate injury severity analysis. Table 1 summarizes the type of data previous studies used with key findings and variables. Hence these factors were used in this study to examine the occurrence of brake failures and the corresponding injury severity.

The literature review indicates that the current research on brake failure crashes is limited, and no study was found that extensively investigated injury severity in brake failure-related crashes considering various injury-related characteristics. The contributions of this study include: (a) exploring the relationship among brake failure, vehicle age, vehicle type, and grade type; (b) assessing the impacts of brake failures on occupant injury severity while accounting for possible intra-crash correlation (effects of the common crash-specific unobserved factors in occupant injury severity within the same crash); and (c) demonstrating the necessity of statewide vehicle inspection regulations.

3. Research methodology

The first analysis performed in this study was Chi-square analysis to examine the relationship among brake failure, vehicle age, vehicle type, and grade type. Three hypotheses were formulated to investigate the associations between the variables.

i. Null hypothesis (H_o): The brake failure occurrence in newer vehicles is the same as those in older vehicles regardless of the vehicle type and the grade type.

Alternate hypothesis (H_a): The brake failure occurrence in newer vehicles is not the same as those in older vehicles regardless of the vehicle type and the grade type.

ii. Null hypothesis (H_o): The brake failure occurrence is the same across different vehicle types regardless of the age of the vehicle and the grade type.

Alternate hypothesis (H_a): The brake failure occurrence is not the same across different vehicle types regardless of the age of the vehicle and grade type.

iii. Null *hypothesis* (H_o): The brake failure occurrence is the same across different grade types regardless of the age and type of vehicle.

Alternate hypothesis (H_a): The brake failure occurrence is not the same across different grade types regardless of the age and type of vehicle.

The Chi-square test is frequently used in the literature to test the difference between what is actually observed and what would be expected if there were truly no relationship between the variables of interest (Agresti, 2009). The vehicle age groups used for the Chi-square analyses were categorized based on the previous literature (Liu & Subramanian, 2020; Nambisan, Boakye, & Yu, 2021), which include 1–6 years, 7–11 years, 12–15 years, and greater than 15 years. Note that, less than 1% of total crashes investigated in this study do not have vehicle year information and thus those crashes were discarded from the analysis.

To investigate the injury severity resulting from the brake failures, binary logit models with a Bayesian inference approach were applied to examine the effects of the vehicle, occupant, crash, and roadway factors contributing to fatal or any other injuries. Apart from the traditional logit or probit models, Bayesian statistics are gaining popularity in traffic safety analysis because of their better performance over the traditional maximum likelihood estimation (MLE) based approach (Haque, Chin, & Huang, 2010; Huang & Abdel-Aty, 2010; Huang, Chin, & Haque, 2008; Ma & Kockelman, 2006; Xie, Zhang, & Liang, 2009). The benefits of using the Bayesian inference model is listed below:

Table 1
Previous studies related to vehicle defects.

Authors	Model/Method	Variables Used	Key Findings
Schoor et al. (2001)	Roadside survey	Vehicle defects, maintenance history, overloading, high speed	Tires and brakes were found as the two most dominant components that contribute to the mechanical defects causing accidents, with overloading an additional factor to consider.
Oduru (2012)	A questionnaire survey	Brake fluid, brake overheating, and brake servicing period	The results indicated that brake failure is caused by low or shortage of brake fluid and brake overheating. It was also recommended that vehicle should be inspected regularly to reduce brake failure
Owusu-Ansah et al. (2014)	A structured questionnaire survey	Responses from bus terminals, automotive workshops, and government institutions	The survey results showed that brake failure in commercial minibuses is caused mainly by Overheating of the brake assembly due to prolong application of the brakes.
Solah et al. (2017)	Logistic regression	Roadworthiness inspection results (Pass and Fail)	It was found that the two most common private passenger vehicle defects were worn out tire (or lack of tread) and structural integrity.
Yan and Xu (2018)	Mathematical model	Vehicle, roadway, and environmental characteristics	The results indicated that brake temperature is positively correlated to the truck weight and the percentage and length of the downgrade.
Wang and Prato (2019)	Partial proportional odds model	Geometric, driver, crash, truck, and environmental characteristics.	The study found a positive relation of truck brake failure with a significant rise in the fatality probability.
Moomen et al. (2019)	Logistic regression	Driver, environmental, crash, traffic, and geometric features	Geometric factors, driver, weather, lighting, and road conditions, and day of week contributed to truck crashes on downgrades.
Rezapour et al. (2019)	Ordered logistic model	Driver, environmental, crash, traffic, and geometric features	Several variables including driver, vehicle, geometric and traffic factors were found to impact single- and multi-vehicle crashes on downgrades.
Dinh et al. (2020)	Systems Theoretic Accident Model	User, vehicle, and road environment characteristics	The results indicated the driver's inexperience, together with the truck's low quality and severe road conditions as potential factors directly leading to the brake failure-related crashes.
Lu et al. (2020)	TruckSim-Simulink joint simulation	Vehicle dynamics characteristics	The results showed that the influence of vehicle brake failure is increased under long and downhill conditions.
Moomen et al. (2020)	Negative binomial model	Geometric factors	Downgrade length, number of lanes, shoulder width, among others were identified as important geometric factors.
Das et al. (2021)	Bayesian data mining method	Crash, vehicle, and geometric characteristics	The findings showed that vehicle age is associated with severe injury in vehicle defects-related crashes.

- The variables are treated as random and the data are used to simulate the behavior of the variables in assessing their distributional properties.
- There is flexibility in selecting the parametric family for prior probability distributions.
- Bayesian inference performs better with small datasets with a multitude of factors/variables and can handle complex models much better than MLE-based methods with the power of Markov Chain Monte Carlo (MCMC) sampling techniques.

Having good knowledge of priors and selecting an appropriate prior distribution help to overcome the drawbacks of the datasets with a small sample size. On the other hand, the robust nature of the Bayesian inference via MCMC methods makes it suitable to handle complex models. Among the three priors, the selection of informative prior was limited since there has not been any previous study that applied Bayesian analysis into occupant injury severity resulting from brake failures. While selecting between the weak informative and non-informative prior, Lemoine (2019) in his study suggested the use of weak informative prior and indicated that the use of non-informative prior produces a similar result to the frequentist model. Moreover, it was found that models with non-informative prior take an unreasonably long time to converge compared to models with weakly informative prior. Therefore, the authors decided to move away from the non-informative prior and applied the weak informative prior to keep the estimates as unbiased as possible. Recently, the No-U-Turn Hamiltonian Monte Carlo (NUT HMC) has been introduced as an advancement of the MCMC sampling technique (Hoffman & Gelman, 2014). The application of the NUT HMC technique was previously rare in the literature, mainly due to the underlying sophisticated mathematics and statistics. However, recent improvements in the Stan programming language in R® statistical software (i.e., "brms;" Bayesian Regression Models using Stan) make it more user-friendly to easily implement the Bayesian models and investigate the posterior distributions using NUT HMC (Bürkner, 2017; Carpenter et al., 2017; Haq et al., 2020a, 2020b, 2020c, 2021a, 2021b). Therefore, this study developed Bayesian binary logit models utilizing the NUT HMC sampling technique in R® installed with the "brms" package.

The Bayesian inference model was developed by going through rigorous fine-tuning of model specification, where the adaptive rejection was set up as 3000 iterations with 1000 as burn-ins. Three parallel MCMC chains were run to simulate posterior distributions resulting in a total of 6000 (2000 runs $\times$ 3 chains) posterior samples. Gelman-Rubin statistic ($\hat{R}$) was used to ensure the convergence of the model (Gelman & Rubin, 1992). The value equal to 1 for each parameter in the model referred to perfect convergence.

For this study, both response and explanatory variables were converted into a binary format of either one or zero (1 or 0). If the binary responses have respective probabilities of p and $1-p$, then the general form of logistic regression can be expressed in Equation (1).

$$logit(\pi) = log\left(\frac{p}{1-p}\right) = (\beta_0 + u_j) + \sum \beta X \quad (1)$$

where β_0 is the intercept, X is the vector of the explanatory variables, β is the regression coefficients for the explanatory variables to be estimated, and u_j is the random effect parameter accounting for the random variation. In a Bayesian modeling framework, the posterior distribution of the parameters is given by the following equations.

$$Posterior = \frac{Likelihood \times Prior}{Average\ \ Likelihood} \tag{2}$$

$$p(\theta|y) = \frac{p(y|\theta)xp(\theta)}{p(y)} \tag{3}$$

where θ = parameters to be estimated, $p(y|\theta)$ = likelihood function, $p(\theta)$ prior information of the parameters, $p(y)$= marginal distribution of y as shown in Equation (4).

$$p(y) = \int_{\theta} p(y, \theta)d\theta \tag{4}$$

Since $p(y)$ is independent of θ, the posterior is only proportional to the product of likelihood and prior, which is clarified in Equation (5).

$$p(\theta|y) \propto p(y|\theta)xp(\theta) \tag{5}$$

An intra-class correlation (ICC) coefficient is defined as a quantitative measure that determines the similarity between individuals within groups. In this study, ICC coefficient was used to assess the random effects, which can examine the correlation among crashes within a specific group (Jones & Jørgensen, 2003). ICC can be defined in terms of random effects model (Gelman & Hill, 2007):

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \tag{6}$$

where, y_{ij} is i-th observation for j-th group, μ is overall mean, α_j is a random effect, and ε_{ij} is error term. α_j is assumed to be normal with a mean of zero and variance σ^2, denoted as $\alpha_j \sim N(0, \sigma^2)$. ICC can be calculated using the following equation (Huang et al., 2008).

$$ICC = \frac{{\sigma_c}^2}{{\sigma_c}^2 + {\sigma_o}^2} \tag{7}$$

where ${\sigma_c}^2$ is the between-crash variance of the random intercept model, and ${\sigma_o}^2$ is the occupant-level variance, which is equal to $\frac{\pi^2}{3} = 3.29$ for a hierarchical logistic distribution. An ICC value close to 1 indicates the significance of between-crash variance in describing total variance and justifies the use of the hierarchical model in the study (Huang et al., 2008; Kutner, Nachtsheim, Neter, & Li, 2005).

4. Data preparation and description

The study was conducted using 10 years (2010 – 2019) of historical crash data in Wyoming roadways. Two primary data sources: Critical Analysis Reporting Environment (CARE) and WYDOT's Roadway Database were utilized to integrate and manage crash data. CARE database has a separate crash, vehicle, and person file. The first step was to combine the three files based on the unique crash case number to obtain more details about the crash, vehicle, and occupant characteristics. The next step was to merge the person-vehicle crash file to the roadway data file based on the milepost. At this stage, the crashes with the involvement of brake failures were filtered out and used for further analysis. This resulted in 1,415 occupant-level brake failure-related crash data. The comprehensive final crash data provided detailed information associated with each crash. Table 2 shows the exact percentages of crash observations in each of the five injury severity levels. Due to the limited number of crash observations in each injury severity level, the five injury severity outcomes were combined into two categories (i.e., fatal or any injury type and no injury). Note that, this aggregation is not expected to substantially affect the inference and a similar approach was commonly found in past studies (Haq et al, 2020a, 2021a; Xu et al., 2016).

The variables used in the injury severity models were converted into categorical predictors and set up as binary (1 or 0) whether the corresponding factor was involved in a fatal/injury-related crash resulting from brake failure. Table 3 shows the descriptive statistics of the investigated variables used for the injury severity model, where the presence of fatal and any injuries (FI) or no injuries involved in brake failure-related crashes was selected as the response variable. The factors were broadly classified into vehicle characteristics, occupant characteristics, crash characteristics, and roadway geometrics to observe their effects on injury severity. The analyzed vehicle characteristics included vehicle age and type. Vehicles aged more than 15 years at the time of the brake failure-related crash were explored. In this study, trucks are defined as any light (weigh less than 10,000 pounds), medium (weight between 10,000 and 26,000 pounds), or heavy (weight more than 26,000 pounds). Note that, pickup trucks and SUVs are not considered light trucks in the WYDOT database. Also, SUVs and pickups were combined into one group because of their similar physical and operational characteristics (Mokhtar & Pervez, 2012). The occupant characteristics analyzed for this model included driver and passenger involvement, age and gender, airbag deployment, drug use, license type, and citation records. The driver's license type was categorized as a commercial driving license (CDL) and non-CDL type as the combination of heavy loads, steep inclines, and long downgrade lengths raise the probability of brake failure to the commercial vehicle resulting from brake heating. The investigated crash characteristics included season, day of the week, and the possible subsequent effects of brake failures (i.e., hitting a guardrail, rollover, and fixed object). The season was reduced to two levels, winter and summer. The day of the week was labeled as weekday and weekend. The horizontal alignment was categorized into curved and straight segments. Table 3 presents a summary of the investigated variables with the count and percentage of each category. Zero (0) was taken as the reference category for each variable and hence the modeling results were observed for the opposite (1) coded factors.

Table 2
Injury information of the brake failure-related crashes (occupant-level).

Injury Severity Type	Count	Percentage
Fatal	13	1%
Incapacitating Injury	38	3%
Non-Incapacitating Injury	141	10%
Possible Injury	96	7%
No Injury	1,127	80%
Total	1,415	100%

Crash is a very complex process in which it is very difficult to capture all the contributory factors during the modeling process. This leads to a very important phenomenon called unobserved heterogeneity. The possibility of such unobserved heterogeneity in the crash modeling process could be substantial, for which several previous studies analyzed and captured such heterogeneity using random effects model (Saeed, Hall, Baroud, & Volovski, 2019; Waseem, Ahmed, & Saeed, 2019). Random effect models have the ability to control for unobserved heterogeneity when the heterogeneity is constant over time and not correlated with independent variables. This constant can be eliminated from longitudinal data through differencing, since taking a first difference will eliminate any time invariant components of the model (Wooldridge, 2010). There are two common assumptions for the random effects and fixed effects model. The random effects assumption considers the individual unobserved heterogeneity as uncorrelated with the independent variables, while the fixed effect assumption considers the individual specific effect as correlated with the independent variables (Wooldridge, 2010). If the random effects assumption holds, the random effects estimator performs more efficient than the fixed effects model.

Table 3
Descriptive statistics.

Variables		Description	Coded Response	Count	Percent
Vehicle characteristics	Vehicle age	Vehicle age	1 = More than 15 years	911	64%
			0 = Otherwise	504	36%
	SUV/pickup	SUV/pickup involvement	1 = Yes	626	44%
			0 = No	789	56%
	Truck	Truck involvement	1 = Yes	451	32%
			0 = No	964	68%
Occupant characteristics	Driver	Driver presence	1 = If the injured person is a driver	922	65%
			0 = Otherwise	493	35%
	Passenger	Passenger presence	1 = If the injured person is a passenger	493	35%
			0 = Otherwise	922	65%
	Young	Young (Age < 25)	1 = Yes	451	32%
			0 = No	964	68%
	Middle	Middle ($25 \leq$ age ≤ 55)	1 = Yes	547	39%
			0 = No	868	61%
	Old	Old (Age > 55)	1 = Yes	304	21%
			0 = No	1111	79%
	Gender	Gender	1 = Female	378	27%
			0 = Other	1037	73%
	Airbag	Airbag deployment	1 = Deployed	101	7%
			0 = Not deployed	1314	93%
	Illegal drugs	Drug involvement	1 = Involved	18	1%
			0 = Not involved	1397	99%
	License	License type	1 = CDL	219	15%
			0 = Other	1196	85%
	Citation	Citation records	1 = At least one	293	21%
			0 = Otherwise	1122	79%
Crash characteristics	Season	Season	1 = Summer	769	54%
			0 = Otherwise	646	46%
	Day of week	Day of week	1 = Weekend	372	26%
			0 = Weekday	1043	74%
	Guardrail	Guardrail related	1 = Yes	53	4%
			0 = No	1362	96%
	Rollover	Rollover related	1 = Yes	191	13%
			0 = No	1224	87%
	Fixed object	Fixed object related	1 = Yes	309	22%
			0 = No	1106	78%
Roadway geometrics	Horizontal alignment	Horizontal alignment	1 = Curve	367	26%
			0 = Straight	1048	74%
	Downhill grade	Downhill grade	1 = Yes	500	35%
			0 = No	915	65%

For occupant-level injury severity crash data, it is most likely that the injury severity levels sustained by the occupants involved in the same crash are correlated (Eluru, Paleti, Pendyala, & Bhat, 2010; Haq et al., 2020a, 2020b, 2020c, 2021a, 2021b; Shaheed, Gkritza, Carriquiry, & Hallmark, 2016; Zhu & Srinivasan, 2011). The parameter estimates could be biased for neglecting such intra-crash correlation in crash data. Therefore, the variability in the injury severity across occupants in the same crash was examined. To explore this unobserved heterogeneity, the random-effects Bayesian approach is applied to the injury severity model.

5. Results and discussions

5.1. Brake failure by vehicle type and vehicle age

Table 4 shows the results for the first hypothesis tested. In Table 4, the occurrence of brake failures was analyzed based on the age of the vehicle for various vehicle types. In addition to this, the analyses were performed by each vertical grade type separately (i.e., level grade, downhill grade, and uphill grade), and the combination of all categories. In Table 4, the results were presented in a set of rows for each vehicle type, sets of columns with results pertaining to the three grade categories, and one column for all categories combined. Within each grade category, the table included three columns for the sample size, the percentage, and the p-value for the Chi-square test. Within individual rows were the four age categories and a row that was for all specific vehicles with no regard to age.

As shown in Table 4, the lowest brake failure occurrence in all vehicles (rows 16 to 19) was 6.4% (Cell N-16) among the vehicles 1–6 years old, with the highest rate (64.9%) among the vehicles over 15 years old. The p-value of less than 0.05 from the Chi-square test denotes a statistical significance of better than 95%. Based on this, the null hypothesis is rejected and concludes that the brake failure occurrence involved in crashes is NOT the same across the age categories of the vehicle. Thus, the analysis suggests that regardless of the type of vehicle (i.e., passenger cars, SUV/pickups, and trucks) and grade type, people in older vehicles were more likely to experience brake failures than those who traveled in newer vehicles. This pattern was consistent across all combinations of individual vehicle types and various grade categories, except for two insignificant scenarios (i.e., Cells L-3 and L-13). The reason behind the insignificance findings should be the presence of uphill segments, which made the drivers avoid brake application while ascending, resulting in no potential difference of brake failures across the age categories of vehicle.

5.2. Brake failure by vehicle age and vehicle type

Table 5 shows the results for the second hypothesis testing. Here the brake failure occurrence in crashes was summarized based on the type of vehicle for various vehicle age categories. In addition to this, the analyses were performed by each vertical grade type separately (i.e., level grade, downhill grade, and uphill grade), and the combination of all categories. Table 5 was prepared

Table 4
Brake failure occurrence by vehicle type and vehicle age.

Vehicle Type	Row #	Vehicle Age (years)	Brake Failure (Level Grade)			Brake Failure(Downhill Grade)			Brake Failure(Uphill Grade)			Brake Failure (TOTAL)		
			Total	Percent	P-Value	Total	Percent	P-Value	Total	Percent	P-Value	Total	Percent	P-Value
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
PAS-CAR	1	1–6	16	6.6%	<0.001	2	3.0%	<0.001	0	0.0%	0.528	18	5.3%	<0.001
	2	7–11	24	9.9%		3	4.5%		4	13.3%		31	9.2%	
	3	12–15	49	20.2%		13	19.7%		8	26.7%		70	20.7%	
	4	> 15	153	63.2%		48	72.7%		18	60.0%		219	64.8%	
	5	Total	242	100.0%		66	100.0%		30	100.0%		338	100.0%	
SUV/PU	6	1–6	4	1.0%	0.001	3	1.7%	<0.001	4	6.5%	0.007	11	1.8%	<0.001
	7	7–11	31	7.9%		22	12.8%		4	6.5%		57	9.1%	
	8	12–15	46	11.7%		46	26.7%		6	9.7%		98	15.7%	
	9	> 15	311	79.3%		101	58.7%		48	77.4%		460	73.5%	
	10	Total	392	100.0%		172	100.0%		62	100.0%		626	100.0%	
TRUCK	11	1–6	34	22.7%	<0.001	23	8.8%	<0.001	4	10.3%	0.534	61	13.5%	<0.001
	12	7–11	37	24.7%		26	9.9%		8	20.5%		71	15.7%	
	13	12–15	39	26.0%		30	11.5%		10	25.6%		79	17.5%	
	14	> 15	40	26.7%		183	69.8%		17	43.6%		240	53.2%	
	15	Total	150	100.0%		262	100.0%		39	100.0%		451	100.0%	
TOTAL	16	1–6	54	6.9%	< 0.001	28	5.6%	<0.001	8	6.1%	<0.001	90	6.4%	<0.001
	17	7–11	92	11.7%		51	10.2%		16	12.2%		159	11.2%	
	18	12–15	134	17.1%		89	17.8%		24	18.3%		247	17.5%	
	19	> 15	504	64.3%		332	66.4%		83	63.4%		919	64.9%	
	20	Total	784	100.0%		500	100.0%		131	100.0%		1415	100.0%	

Table 5
Brake failure occurrence by vehicle age and vehicle type.

Vehicle Age (years)	Row #	Vehicle Type	Brake Failure (Level Grade)			Brake Failure(Downhill Grade)			Brake Failure(Uphill Grade)			Brake Failure (TOTAL)		
			Total	Percent	P-Value	Total	Percent	P-Value	Total	Percent	P-Value	Total	Percent	P-Value
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1–6	1	PAS-CAR	16	29.6%	0.004	2	7.1%	0.081	0	0.0%	0.008	18	20.0%	<0.001
	2	SUV/PU	4	7.4%		3	10.7%		4	50.0%		11	12.2%	
	3	TRUCK	34	63.0%		23	82.1%		4	50.0%		61	67.8%	
	4	TOTAL	54	100.0%		28	100.0%		8	100.0%		90	100.0%	
7–11	5	PAS-CAR	24	26.1%	0.015	3	5.9%	0.008	4	25.0%	0.447	31	19.5%	<0.001
	6	SUV/PU	31	33.7%		22	43.1%		4	25.0%		57	35.8%	
	7	TRUCK	37	40.2%		26	51.0%		8	50.0%		71	44.7%	
	8	TOTAL	92	100.0%		51	100.0%		16	100.0%		159	100.0%	
12–15	9	PAS-CAR	49	36.6%	0.007	13	14.6%	<0.001	8	33.3%	0.183	70	28.3%	0.081
	10	SUV/PU	46	34.3%		46	51.7%		6	25.0%		98	39.7%	
	11	TRUCK	39	29.1%		30	33.7%		10	41.7%		79	32.0%	
	12	TOTAL	134	100.0%		89	100.0%		24	100.0%		247	100.0%	
> 15	13	PAS-CAR	153	30.4%	<0.001	48	14.5%	<0.001	18	21.7%	0.633	219	23.8%	<0.001
	14	SUV/PU	311	61.7%		101	30.4%		48	57.8%		460	50.1%	
	15	TRUCK	40	7.9%		183	55.1%		17	20.5%		240	26.1%	
	16	TOTAL	504	100.0%		332	100.0%		83	100.0%		919	100.0%	
TOTAL	17	PAS-CAR	242	30.9%	<0.001	66	13.2%	< 0.001	30	22.9%	0.002	338	23.9%	<0.001
	18	SUV/PU	392	50.0%		172	34.4%		62	47.3%		626	44.2%	
	19	TRUCK	150	19.1%		262	52.4%		39	29.8%		451	31.9%	
	20	TOTAL	784	100.0%		500	100.0%		131	100.0%		1415	100.0%	

similarly to Table 4, but to reflect the difference in the hypothesis being tested.

Table 5 shows that the lowest brake failure occurrence in all vehicles (rows 17 to 19) was 23.9% (Cell N-17) among occupants in passenger cars, with the highest rate (44.2%) among occupants in SUV/pickups. The p-value of less than 0.05 from the Chi-square test denotes the rejection of the null hypothesis and concludes that the brake failure occurrence is NOT the same across different vehicle types regardless of the age of the vehicle and the grade type. These results were similar across various vehicle age categories, with the exception of a few for uphill grade categories. As mentioned earlier, the possible explanation behind the insignificant results should be the fact that drivers typically do not need to apply brakes while ascending upgrades, resulting in a relatively less vulnerable situation for brake failures to occur.

The analysis results provided in Table 5 indicate that, overall, occupants of SUV/pickups were more likely to encounter brake failures compared to occupants in passenger cars and trucks. However, it is interesting to note that the finding was not consistent for the downhill grade category. Here trucks were found to experience a major portion (Cell H-19) of brake failure-related crashes. This is attributed to the fact that the combination of heavy loads, steep inclines, and long downgrade lengths increases the risk of brake failure resulting from brake heating. As trucks descend downgrades, large amounts of potential energy are generated and absorbed by the truck's service brakes. This potential energy is then converted to heat energy. This is then absorbed by the braking system, which increases the braking temperature. This ultimately results in brake failure and truck runaway.

5.3. Brake failure by vehicle age and grade type

The results for the third hypothesis testing are summarized in Table 6. The brake failure occurrence in crashes was analyzed based on three grade types (i.e., level, downhill, and uphill) for various vehicle age categories. In addition to this, the analyses were conducted by each vehicle type separately (passenger car, SUV/pickups, and trucks), and the combination of all the vehicle types.

Table 6 shows that the lowest brake failure occurrence in all vehicles (rows 17 to 19) was 9.3% (Cell N-19) among occupants while traveling uphill segments, with the highest rate (55.4%) among occupants traveling level segments. The p-value of less than 0.05 from the Chi-square test denotes the rejection of the null hypothesis and concludes that the brake failure occurrence is NOT the same across different grade types regardless of age and type of vehicle. These results were similar across various vehicle type categories, with exceptions for passenger car categories. It is reasonable that passenger car has relatively good braking performance because of their shorter braking distance and time, compared to SUV/pickups and large trucks resulted in comparatively less vulnerability to brake fades. Although the overall result indicated the highest brake failure occurrence on level grade segments, this was not the same when broken down to various vehicle types. For example, the highest brake failure occurrence for trucks was 58.1% (Cell K-18) among the truck occupants while traveling downhill segments. As mentioned earlier, trucks are typically more vulnerable to downgrade crashes due to their heavy loads and large sizes causing brake overheating, fade, and failure.

The hypothesis results in this study are also found to be consistent with other previous studies. It was indicated by several studies that the risk of crashes appears to be increased as the vehicles ages (Anderson & Searson, 2015; Liu & Subramanian, 2020; NHTSA, 2013; NHTSA, 2018). Yan and Xu (2018) reported trucks as more vulnerable to brake failures compared to other vehicle types as the brake temperature was found to be positively correlated to truck weight. Moreover, there are a number of studies that indicated the prolonging application of the brakes under long downhill condition as one of the critical factors behind brake failures (Dinh et al., 2020; Lu et al., 2020; Moomen et al., 2019, 2020; Yan & Xu, 2018).

5.4. Injury severity model

The factors affecting brake failure-related injury severities are discussed in this section, where 95% Bayesian credible interval (BCI) was applied to determine the significance of variables contributing to the occupant injury severity. The possible presence of multicollinearity was checked using the variance inflation factor (VIF), calculated for each predictor. Typically, a VIF value of greater than 10 is considered as an indication of the existence of multicollinearity (Kutner et al., 2005). However, no such issues were observed since the VIF value of all factors in the model fell below two, as shown in Table 7. The parameter estimates of the model along with the 95% BCI, standard error, and odds ratio (OR) of the variables are provided in Table 8. The table also includes the area under the curve (AUC) as a goodness of fit parameter. The prediction accuracy of the model can be measured using the AUC value, where the value close to 1 implies a better fit of the model. The result provided in Table 8 demonstrates an excellent overall model fit (0.814), which implies that 81.4% of the observations are in agreement with predictors when predicting occupant injury severity. The number of effective variables (pD) in the model was found closer to the actual number of significant variables, which implies the non-complexity of the models.

5.4.1. Measure of unobserved heterogeneity

Intra-class correlation (ICC) coefficient was computed to determine the proportion of variance in the occupant injury severity associated with the same crash, as provided in Table 8. The use of a hierarchical model was highly encouraged for the presence of unobserved heterogeneity associated with the model, although no threshold of ICC was found in the literature (McElreath, 2018). Based on Table 8, the ICC value was 0.41 for the model resulting from between-crash variance. This implies 41% of unexplained variations due to the presence of common crash-specific unobserved factors affecting the occupant injury severity in the same crash. Therefore, the use of the hierarchical model with random effects was justified to explain such unexplained factors within the same crash.

5.4.2. Significant factor analysis

The modeling results indicate the various vehicle, occupant, crash, and geometrical characteristics that significantly con-

Table 6
Brake failure occurrence by vehicle age and grade type.

Vehicle Age (years)	Row #	Grade Type	Brake Failure (PAS-CAR)			Brake Failure (SUV/PU)			Brake Failure (TRUCK)			Brake Failure (TOTAL)		
			Total	Percent	P-Value	Total	Percent	P-Value	Total	Percent	P-Value	Total	Percent	P-Value
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1–6	1	Level	16	88.9%	0.268	4	36.4%	<0.001	34	55.7%	<0.001	54	60.0%	<0.001
	2	Downhill	2	11.1%		3	27.3%		23	37.7%		28	31.1%	
	3	Uphill	0	0.0%		4	36.4%		4	6.6%		8	8.9%	
	4	Total	18	100.0%		11	100.0%		61	100.0%		90	100.0%	
7–11	5	Level	24	77.4%	0.249	31	54.4%	0.021	37	52.1%	< 0.001	92	57.9%	<0.001
	6	Downhill	3	9.7%		22	38.6%		26	36.6%		51	32.1%	
	7	Uphill	4	12.9%		4	7.0%		8	11.3%		16	10.1%	
	8	Total	31	100.0%		57	100.0%		71	100.0%		159	100.0%	
12–15	9	Level	49	70.0%	0.261	46	46.9%	<0.001	39	49.4%	<0.001	134	54.3%	<0.001
	10	Downhill	13	18.6%		46	46.9%		30	38.0%		89	36.0%	
	11	Uphill	8	11.4%		6	6.1%		10	12.7%		24	9.7%	
	12	Total	70	100.0%		98	100.0%		79	100.0%		247	100.0%	
> 15	13	Level	153	69.9%	0.781	311	67.6%	< 0.001	40	16.7%	<0.001	504	54.8%	<0.001
	14	Downhill	48	21.9%		101	22.0%		183	76.3%		332	36.1%	
	15	Uphill	18	8.2%		48	10.4%		17	7.1%		83	9.0%	
	16	Total	219	100.0%		460	100.0%		240	100.0%		919	100.0%	
TOTAL	17	Level	242	71.6%	<0.001	392	62.6%	<0.001	150	33.3%	<0.001	784	55.4%	<0.001
	18	Downhill	66	19.5%		172	27.5%		262	58.1%		500	35.3%	
	19	Uphill	30	8.9%		62	9.9%		39	8.6%		131	9.3%	
	20	Total	338	100.0%		626	100.0%		451	100.0%		1415	100.0%	

Table 7
Variance Inflation Factor (VIF) of the investigated variables used in this study.

Variables	VIF Values
Vehicle age (>15 years)	1.13
SUV/pickup (involved)	1.22
Truck (involved)	1.07
Driver (present)	1.31
Passenger (present)	1.01
Middle (25 ≤ age ≤ 55)	1.21
Gender (female)	1.12
Airbag (deployed)	1.06
Illegal drugs (involved)	1.05
License type (CDL)	1.45
Citation (at least one)	1.28
Season (summer)	1.08
Day of week (weekends)	1.03
Guardrail (yes)	1.09
Rollover (yes)	1.28
Fixed object (yes)	1.26
Horizontal alignment (curve)	1.45
Downhill grade (yes)	1.40

tributed to the severity of occupant injuries in brake failure-related crashes, as shown in Table 8. Vehicles older than 15 years were found to be involved in more fatal/injury-related brake failure crashes by estimated odds of 1.8 times as compared to vehicles aged less or equal to 15 years. This is attributed to the fact that the owner of an older vehicle does not pay more attention to the routine maintenance of the vehicle parts as compared to the owner of a newer vehicle. This tends to accelerate the vehicle defects, including brake deformation quicker than the normal condition. The result is consistent with the previous study that also found a significant association of vehicle age with severe injuries (Das et al., 2021). While investigating vehicle type in association with brake failures, it was found that trucks were more likely to experience brake failures compared to SUV/pickup based on the magnitude of the estimates, whereas the impact of car type was found insignificant. The estimated odds of fatal/injury resulting from brake failure occurrence increased by 1.5 and 6.9 times for the vehicles being SUV/pickup and trucks, respectively. The reason should be the fact that heavy vehicles carrying huge loads typically take more braking distance and time, resulting in faster degradation of brakes as compared to smaller vehicles. The brake failure occurrences of heavy vehicles while descending on downgrades are quite frequently reported by the previous study (Moomen, Rezapour, & Ksaibati, 2019).

While investigating occupant characteristics associated with brake failure-related crashes, the estimated odds of fatal/injury increased by 9.5, 11.8, 1.3, and 1.6 times when the occupant was driver, passenger, middle-aged, and female, respectively. It is reasonable to find higher injury severity of passengers compared to the driver because of the lower seat belt compliance rate by the passenger. Although Wyoming has a mandatory seat belt use law for all occupants, regardless of the sitting position, it seems possible that the passengers were less likely to wear a seatbelt. A possible explanation behind higher severe injuries for females could be greater physiological strength and injury-sustaining capability of males as compared to females, as argued by O'Donnell and Connor (1996). Similar results were also found in the previous studies (Sharmin et al., 2020; Hossain et al., 2022). The deployment of the airbag is supposed to reduce the injury severity. However, it was found to increase the estimated odds of fatal/injury by 7.9 times. This is attributed to the fact that airbags typically deploy with bigger crash impacts. This could be the probable reason behind higher injury severity. When the occupants have illegal

Table 8
Factors affecting Occupant Injury Severity in Brake Defects-related Crashes.

Variables	Estimates	Error	Credible Interval		ICC
			2.5%	97.5%	
Random Effects					
Between-crash variance	2.32	1.29	0.44	2.61	0.41
	Estimates	Error	2.5%	97.5%	OR
Intercept	−5.62	0.42	−6.45	−4.82	-
Vehicle Characteristics					
Vehicle age (>15 years)	0.57**	0.18	0.23	0.92	1.77
SUV/pickup (involved)	0.39**	0.17	0.05	0.73	1.48
Truck (involved)	1.94**	0.82	0.43	3.65	6.96
Occupant Characteristics					
Driver (present)	2.25**	0.26	1.74	2.76	9.49
Passenger (present)	2.47**	0.27	1.95	3.00	11.82
Middle (25 ≤ age ≤ 55)	0.29*	0.17	−0.05	0.63	1.34
Gender (female)	0.47**	0.18	0.11	0.81	1.60
Airbag (deployed)	2.06**	0.24	1.59	2.53	7.85
Illegal drugs (involved)	1.27**	0.53	0.24	2.33	3.56
License type (CDL)	0.47*	0.26	−0.04	0.98	1.60
Citation (at least one)	0.51**	0.21	0.09	0.91	1.67
Crash Characteristics					
Season (summer)	0.35**	0.16	0.04	0.66	1.42
Day of week (weekends)	0.28*	0.17	−0.04	0.60	1.32
Guardrail (yes)	0.68*	0.35	−0.01	1.35	1.97
Rollover (yes)	1.34**	0.21	0.94	1.75	3.82
Fixed object (yes)	−0.51**	0.22	−0.93	−0.10	0.60
Roadway Geometrics					
Horizontal alignment (curve)	0.38**	0.19	0.01	0.75	1.46
Downhill grade (yes)	0.30*	0.19	−0.05	0.65	1.35
Model Statistics					
Number of observations	1,415				
Effective number of variables, pD	17.7				
AUC	0.814				

** Variables significant at 95% credible interval.
* Variables significant at 90% credible interval.

drug involvement and previous citation records, they were more likely to experience fatal/injury by estimated odds of 3.6 and 1.7 times, respectively. The results were found in compliance with the previous studies (Haq et al., 2020b, 2021b; Lemp, Kockelman, & Unnikrishnan, 2011). Saeed, Nateghi, Hall, and Waldorf (2020) found that a township's population composition and its abundance of alcohol-related businesses influence the alcohol-related driving crash rates. Drivers with CDL license types were found to be involved in more fatal/injury crashes by the estimated odds of 1.6 times as compared to non-CDL drivers. Commercial motor vehicles usually carry huge loads, which require immense brake force to stop and thus increase the possibility of brake failures.

Among the crash characteristics, brake defects were more likely to be associated with fatal/injury in the summer season and weekends by estimated odds of 1.4 and 1.3 times, respectively. The possible explanations could be the generation of excessive heat during summer, which makes the brakes susceptible to faster degradation and significant stress on the brake pads. The result is consistent with a recent study by Assemi, Hickman, and Paz (2021), which also found a significant positive relationship of maximum temperature with vehicle defect-related crashes. Brake failures followed by hitting a guardrail and rollover increased the estimated odds of fatal/injury by 1.9 and 3.8 times, respectively. However, the involvement of a fixed object was found to decrease the estimated odds of occupant fatality or any injury. Such impacts of guardrail and rollover on severe injuries are commonly reported by previous studies (Alrejjal, Farid, & Ksaibati, 2021; Haq, Zlatkovic, & Ksaibati, 2021a). It seems possible that crashes involving fixed objects cause property damage only, resulting in no or minor injury. When it comes to roadway geometrics, the presence of curve and downgrade segments were found to increase the estimated odds of fatal/injury in brake failure-related crashes by 1.5 and 1.4 times, respectively. Several highways in Wyoming traverse over mountain passes featuring steep downgrades. The significant mountain passes in the state include US 14, US 16, Teton pass, South Pass, and US 14 Alternative. The trucks that traverse these mountain passes typically carry large loads, putting them in a perilous state. In fact, during certain periods of the year, some mountain passes only accommodate lower weight limits, whereas others are closed to traffic completely. Overheating of the brake assembly due to prolonging the application of the brakes while descending those downgrades combined with curves should be the main reason for brake failures to occur and the corresponding severe injuries. Previous studies also indicated the adverse impacts of such geometry on occupant injury severities (Sharmin et al., 2022; Haq, Zlatkovic, & Ksaibati, 2020a; Moomen, Rezapour, Raza, & Ksaibati, 2020; Rezapour et al., 2019). While comparing safety sensitivity of roadway characteristics to various highway classes, Chen, Saeed, Alinizzi, Lavrenz, and Labi (2019) found that crashes at higher-class highways (e.g., interstates) are more sensitive to: changes in traffic volume, average vertical grade, median width, inside shoulder width, and the pavement condition; but less sensitive to changes in lane width and pavement condition, as compared to the relatively lower-class highways (e.g., state roads).

6. Conclusions and recommendations

Although the braking system plays a key role in a safe and smooth vehicular operation, it has not been given proper attention and, hence, brake failures are still underrepresented in traffic safety. The current body of literature on brake failure-related crashes is very limited. Moreover, no previous study was found to extensively investigate the factors associated with brake failures and the corresponding injury severity. This study aims to fill this knowledge gap by examining brake failure-related crashes and assessing the impacts of brake failures on occupant injury severity. An extensive exploratory analysis was conducted using 10 years (2010–2019) of historical crash data along Wyoming highways.

The study first performed a Chi-square analysis to examine the relationship among brake failure, vehicle age, vehicle type, and grade type. Three hypotheses were formulated to investigate the associations between the variables. The result of the first hypothesis suggested that regardless of the type of vehicle (i.e., passenger cars, SUV/pickups, and trucks) and grade type (i.e., level, downhill, and uphill), people in older vehicles were more likely to experience brake failures than those who traveled in newer vehicles. The second hypothesis concluded that the brake failure occurrence is NOT the same across different vehicle types, regardless of the age of the vehicle and the grade type. This pattern was consistent across various vehicle age categories, with the exception of a few for uphill grade categories. Finally, the third hypothesis indicated that the brake failure occurrence is NOT the same across different grade types, regardless of age and type of vehicle. The statement was true across various vehicle types, with exceptions for passenger car categories. Based on the hypotheses tested for this study, vehicles aged more than 15 years, trucks, and downhill grade segments seemed to be highly associated with brake failure occurrences.

To investigate the impacts of brake defects on occupant injury severity, the study applied Binary logistic regression with the Bayesian inference approach. The ICC value showed 41% of unexplained variation resulted from between-crash variance, indicating a strong correlation in the injury propensities among occupants involved in the same brake failure-related crash. The modeling results quantified the significant impacts of brake failures on occupant injury severity and identified various vehicle, occupants, crash, and roadway geometrical characteristics. The incidence of brake failure in a crash significantly contributed to more severe injuries when combined with any of the following factors: Vehicle characteristics – age greater than 15 years, truck and SUV/pickup involvement; Occupant characteristics – being a driver, passenger, middle-aged, and female, airbag deployment, drugs use, commercial license type, and citation records; Crash characteristics: summer season, weekends, hitting a guardrail, and rollover; and Roadway geometrics: horizontal curves and vertical grades.

Based on the findings, several recommendations can be made to enhance vehicle safety. In the United States, 34 out of 50 states do not require annual or biennial vehicle safety inspections, and many of the states that currently require safety inspections are reconsidering the impact of vehicle inspection policies (Das et al., 2021). Wyoming does not have any periodic safety and emission inspection program. Therefore, a statewide vehicle inspection program concentrating on brakes, tires, and engines should be established to facilitate safety improvements. Driver education programs on the braking system should be introduced on when to check, maintain, and replace defective brakes. A more upgraded Grade Severity Rating System (GSRS) should be implemented to define safe descent speeds at Wyoming mountain passes since most of the commercial loaded trucks cannot handle such high speeds. WYDOT should take necessary actions to improve the crash reporting systems and emphasize the brake failure-related information on a challenging roadway with a high truck percentage.

Several data limitations and gaps were encountered in this study. The Wyoming crash database has an underreporting issue of detailed information associated with vehicle brakes along with a high discrepancy in brake failure reporting systems. This requires the demand for better brake defects-related crash data collection, since highway patrol are not trained to recognize brake defects. Ongoing and future research will explore more detailed information in terms of vehicular model, manufacturing company, emission system, safety system, advanced driver assistance system (ADAS), and other features to analyze brake defects.

Conflict of Interest

The authors report no potential conflict of interest.

Acknowledgement

The work presented in this paper is sponsored by the Mountain-Plains Consortium (MPC-540) and the Wyoming Department of Transportation (contract number RS08220).

References

Agresti, A. (2009). *An Introduction to Categorical Data Analysis* (second ed.). John Wiley & Sond Inc.

Alrejjal, A., Farid, A., & Ksaibati, K. (2021). A correlated random parameters approach to investigate large truck rollover crashes on mountainous interstates. *Accident Analysis & Prevention, 159*, 106233. https://doi.org/10.1016/j.aap.2021.106233.

Anderson, R. W. G., & Searson, D. J. (2015). Use of age-period-cohort models to estimate effects of vehicle age, year of crash and year of vehicle manufacture on driver injury and fatality rates in single vehicle crashes in New South Wales, 2003–2010. *Accident Analysis and Prevention, 75*, 202–210.

Assemi, B., Hickman, M., & Paz, A. (2021). Relationship between programmed heavy vehicle inspections and traffic safety. *Transportation Research Record*. https://doi.org/10.1177/03611981211016458.

Bürkner, P.-C. (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*, 1–28 https://www.uni-muenster.de/imperia/md/content/psyifp/ae_holling/brms_talk_26.02.16.pdf.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan : A probabilistic programming language. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v076.i01.

Chen, S., Saeed, T. U., Alinizzi, M., Lavrenz, S., & Labi, S. (2019). Safety sensitivity to roadway characteristics: A comparison across highway classes. *Accident Analysis & Prevention, 123*, 39–50. https://doi.org/10.1016/j.aap.2018.10.020.

Das, S., Dutta, A., & Geedipally, S. R. (2021). Applying Bayesian data mining to measure the effect of vehicular defects on crash severity. *Journal of Transportation Safety & Security, 13*(6), 605–621. https://doi.org/10.1080/19439962.2019.1658674.

DEKRA Automobil GmbH. (2015). *Road Safety Report 2015: A Future Based on Experience.* http://www.dekra-vision-zero.com/downloads/road_safety_report_2015.pdf.

Dinh, D. D., Vu, N. H., McIlroy, R. C., Plant, K. L., & Stanton, N. A. (2020). Applying the Systems Theoretic Accident Model and Process to Analyze a Downgrade-Truck Collision Caused by a Brake Failure in Vietnam. *Advances in Human Aspects of Transportation*. https://doi.org/10.1007/978-3-030-50943-9_41.

Eluru, N., Paleti, R., Pendyala, R. M., & Bhat, C. R. (2010). *Modeling multiple vehicle occupant injury severity: A copula-based multivariate approach*. Accepted by Transportation Research Record.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation using Multiple Sequences. *Statistical Science, 7*(4), 457–511.

Haq, M. T., Zlatkovic, M., & Ksaibati, K. (2020a). Occupant Injury severity in passenger car-truck collisions on interstate 80 in Wyoming: A Hamiltonian Monte Carlo Markov Chain Bayesian Inference Approach. *Journal of Transportation Safety & Security*. https://doi.org/10.1080/19439962.2020.1786872.

Haq, M. T., Zlatkovic, M., & Ksaibati, K. (2020b). Investigating occupant injury severity of tuck-involved crashes based on vehicle types on a mountainous Freeway: A hierarchical Bayesian random intercept approach. *Accident Analysis & Prevention, 144*. https://doi.org/10.1016/j.aap.2020.105654.

Haq, M. T., Zlatkovic, M., & Ksaibati, K. (2020c). Assessment of tire failure related crashes and injury severity on a mountainous freeway: Bayesian binary logit approach. *Accident Analysis & Prevention, 145*. https://doi.org/10.1016/j.aap.2020.105693.

Haq, M. T., Zlatkovic, M., & Ksaibati, K. (2021a). Assessment of commercial truck driver injury severity as a result of driving actions. *Transportation Research Record*. https://doi.org/10.1177/03611981211009880.

Haq, M. T., Zlatkovic, M., & Ksaibati, K. (2021b). Assessment of commercial truck driver injury severity based on truck configurations along mountainous roadway using hierarchical bayesian random intercept approach. *Accident Analysis & Prevention*. https://doi.org/10.1016/j.aap.2021.106392.

Haque, M. M., Chin, H. C., & Huang, H. (2010). Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Accident Analysis & Prevention*. https://doi.org/10.1016/j.aap.2009.07.022.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively setting path lengths in hamiltonian Monte Carlo. *Journal of Machine Learning Research, 15*(1), 1593–1623.

Hossain, M. J., Ivan, J. N., Zhao, S., Wang, K., Sharmin, S., Ravishanker, N., & Jackson, E. (2022). Considering demographics of other involved drivers in predicting the highest driver injury severity in multi-vehicle crashes on rural two-lane roads in California. *Journal of Transportation Safety and Security, 1–16*. https://doi.org/10.1080/19439962.2022.2033899.

Huang, H., Chin, H. C., & Haque, M. M. (2008). Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis. *Accident Analysis & Prevention, 40*(1), 45–54. https://doi.org/10.1016/j.aap.2007.04.002.

Huang, H., & Abdel-Aty, M. (2010). Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis & Prevention*. https://doi.org/10.1016/j.aap.2010.03.013.

Ibitayo, L. D., Mohammed, G. U., Rabiu, I., & Abdulrahman, A. (2016). Expert system for detecting and diagnosing car engine brake failure fault using dynamic control system (DCS). *Journal of Applied & Computational Mathematics*. https://doi.org/10.4172/2168-9679.1000281.

Jones, A. P., & Jørgensen, S. H. (2003). The use of multilevel models for the prediction of road accident outcomes. *Accident Analysis & Prevention, 35*(1), 59–69.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical model* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Lemp, J. D., Kockelman, K. M., & Unnikrishnan, A. (2011). Analysis of large truck crash severity using heteroskedastic ordered probit models. *Accident Analysis & Prevention*. https://doi.org/10.1016/j.aap.2010.09.006.

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*. https://doi.org/10.1111/oik.05985.

Liu, C., & Subramanian, R. (2020). The Relationship between Passenger Vehicle Occupant Injury Outcomes and Vehicle Age or Model Year in Police-Reported Crashes. Report No. DOT HS 812 937.

Lu, Y., Wang, F., & Zhang, G. (2020). Research on brake failure control of heavy commercial vehicles based on turning conditions. In *2020 4th CAA International Conference on Vehicular Control and Intelligence (CVCI)* (pp. 395–400). https://doi.org/10.1109/CVCI51460.2020.9338601.

Ma, J., & Kockelman, K. M. (2006). Bayesian multivariate Poisson regression for models of injury count, by severity. *Transportation Research Record., 1950*, 24–34.

McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC.

Mokhtar, W. A., & Pervez, N. (2012). Underbody drag for pickup trucks. In *30th AIAA Applied Aerodynamics Conference, 25–28*. https://doi.org/10.2514/6.2012-3210.

Moomen, M., Rezapour, M., & Ksaibati, K. (2019). An investigation of influential factors of downgrade truck crashes: A logistic regression approach. *Journal of Traffic and Transportation Engineering (English Edition)*. https://doi.org/10.1016/j.jtte.2018.03.005.

Moomen, M., Rezapour, M., Raza, M. N., & Ksaibati, K. (2020). Predicting injury severity and crash frequency: Insights into the impacts of geometric variables on downgrade crashes in Wyoming. *Journal of Traffic and Transportation Engineering (English Edition)*. https://doi.org/10.1016/j.jtte.2019.04.002.

Nambisan, S., Boakye, K., & Yu, D. (2021). Seatbelt Use among Vehicle Occupants in Fatal Crashes in the United States: Does Vehicle Type and Age Affect Injury Outcome? Presented in Transportation Research Board 100th Annual Meeting, Washington, D.C.

National Highway Traffic Safety Administration. (NHTSA). (2013). *How vehicle age and model year relate to driver injury severity in fatal crashes*. Report No. DOT HS 811 825.

National Highway Traffic Safety Administration. (NHTSA). (2018). *Passenger vehicle occupant injury severity by vehicle age and model year in fatal crashes*. Report No. DOT HS 812 528.

National Highway Traffic Safety Administration (NHTSA). (2019). *2017 State Traffic Data Traffic Safety Fact Sheet*. U.S. Department of Transportation. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812780.

O'Donnell, C. J., & Connor, D. H. (1996). Predicting the severity using models of ordered multiple choice. *Accident Analysis & Prevention, 28*(6), 739–753.

Oduru, S. D. (2012). Brake Failure and its Effect on Road Traffic Accident in Kumasi Metropolis, Ghana. *International Journal of Science and Technology (IJST), 1*(9).

Owusu-Ansah, P., Alhassan, T., Frimpong, A., & Agyemang, A. A. (2014). Survey of the Cause of Brake Failure in Commercial Mini-buses in Kumasi. *Research Journal of Applied Sciences, Engineering and Technology, 7*(23), 4877–4882.

Rezapour, M., Moomen, M., & Ksaibati, K. (2019). Ordered logistic models of influencing factors on crash injury severity of single and multiple-vehicle downgrade crashes: A case study in Wyoming. *Journal of Safety Research*. https://doi.org/10.1016/j.jsr.2018.12.006.

Saeed, T. U., Hall, T., Baroud, H., & Volovski, M. J. (2019). Analyzing road crash frequencies with uncorrelated and correlated random-parameters count models: An empirical assessment of multilane highways. *Analytic Methods in Accident Research, 23*, 100101. https://doi.org/10.1016/j.amar.2019.100101.

Saeed, T. U., Nateghi, R., Hall, T., & Waldorf, B. S. (2020). Statistical analysis of area-wide alcohol-related driving crashes: A spatial econometric approach. *Geographical Analysis, 52*(3), 394–417. https://doi.org/10.1111/gean.12216.

Schoor, O. V., Niekerk, J. L., & Grobbelaar, B. (2001). Mechanical failures as a contributing cause to motor vehicle accidents - South Africa. *Accident Analysis & Prevention., 33*(6), 713–721. https://doi.org/10.1016/S0001-4575(00)00083-X.

Shaheed, M. S., Gkritza, K., Carriquiry, A. L., & Hallmark, S. L. (2016). Analysis of occupant injury severity in winter weather crashes: A fully Bayesian multivariate approach. *Analytic Methods in Accident Research, 11*, 33–47. https://doi.org/10.1016/j.amar.2016.06.002.

Sharmin, S., Ivan, J. N., Zhao, S., Wang, K., Hossain, M. J., Ravishanker, N., & Jackson, E. (2020). Incorporating demographic proportions into crash count models by quasi-induced exposure method. *Transportation research record, 2674*(9), 548–560. https://doi.org/10.1177/0361198120930.

Sharmin, S., Ivan, J. N., Marsh, K. L., Paxton, A., & Tucker, A. (2022). Driver Psychology Latent Classes as Predictors of Traffic Incident Occurrence in Naturalistic Driving Study Data. *Transportation Research Record*. https://doi.org/10.1177/036119812211089.

Solah, M. S., Hamzah, A., Ariffin, A. H., Paiman, N. F., Abdul Hamid, I., Abdul Wahab, M. A. F., ... Osman, M. R. (2017). Private vehicle roadworthiness in Malaysia from the vehicle inspection perspective article history. *Journal of the Society of Automotive Engineers Malaysia, 1*(3), 262–271 www.journal.saemalaysia.org.my.

U.S. Department of Transportation (USDOT). (2018). *Critical reasons for crashes investigated in the national motor vehicle crash causation survey: A brief statistical summary*. Report No. DOT HS 812 506.

U.S. Department of Transportation (USDOT). (2020). *Risk-Based Processes for Safety Defect Analysis and Management of Recalls*. Report No. DOT HS 812 984.

Wang, Y., & Prato, C. G. (2019). Determinants of injury severity for truck crashes on mountain expressways in China: A case-study with a partial proportional odds model. *Safety Science, 117*, 100–107. https://doi.org/10.1016/j.ssci.2019.04.011.

Waseem, M., Ahmed, A., & Saeed, T. U. (2019). Factors affecting motorcyclists' injury severities: An empirical assessment using random parameters logit model with heterogeneity in means and variances. *Accident Analysis and Prevention, 123*, 12–19. https://doi.org/10.1016/j.aap.2018.10.022.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed. p. 252). Cambridge, Mass: MIT Press.

Xie, Y., Zhang, Y., & Liang, F. (2009). Crash injury severity analysis using Bayesian ordered probit models. *Journal of Transportation Engineering, 135*(1), 18–25. https://doi.org/10.1061/(ASCE)0733-947X(2009)135:1(18).

Xu, X., Xie, S., Wong, S. C., Xu, P., Huang, H., & Pei, X. (2016). Severity of pedestrian injuries due to traffic crashes at signalized intersections in Hong Kong: A Bayesian spatial logit model. *Journal of Advanced Transportation*. https://doi.org/10.1002/atr.1442.

Yan, M., & Xu, J. (2018). Prediction Model for Brake-Drum Temperature of Large Trucks on Consecutive Mountain Downgrade Routes Based on Energy Conservation Law. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2018/4587673.

Zhu, X., & Srinivasan, S. (2011). Modeling occupant-level injury severity: An application to large-truck crashes. *Accident Analysis & Prevention*. https://doi.org/10.1016/j.aap.2011.02.021.

Muhammad Tahmidul Haq, Ph.D., P.E. Dr. Muhammad Tahmidul Haq is a Post-doctoral Research Associate in the Wyoming Technology Transfer Center at the University of Wyoming. He joined the University of Wyoming transportation group in Fall 2016 as an MS student and earned his Ph.D. in 2020. Dr. Haq has a strong experience in writing research proposals for funding, writing several high-level scientific manuscripts and publishing those in different prestigious peer-reviewed journals. His primary research interest lies in traffic safety, traffic operation, transportation planning and Intelligent Transportation Systems (ITS). Dr. Haq is currently working as an Principal Investigator (PI) on three Federal funded projects of significant National Interest: Introducing Innovative New Interchanges to Improve the Performance of Failing Service Interchanges; Automating the Implementation of Updated Grade Severity Rating System (GSRS) for Mountain Passes; and Developing a Collision Warning and Collision Avoidance System for Snowplows and Other Maintenance Vehicles.

Vincent-Michael Kwesi Ampadu, Ph.D. Dr. Vincent Ampadu earned his Ph.D. from the Civil and Architectural and Construction Engineering Department of the University of Wyoming. He is currently working on a research project titled Grade Severity Rating System (GSRS) for plain (longitudinal grades only) and combined (longitudinal grades and horizontal curves) downgrades in Wyoming. In particular, he is focusing to automate the mathematical GSRS models by building software based on the algorithms and investigate the implementation of more accurate models to reduce the incidence of truck crashes on downgrades using experimental, statistical and machine learning techniques. Besides his specialization in transportation engineering with a focus on traffic safety, Dr. Ampadu has strong background in object oriented software development with Visual BASIC.net and MS-SQL database server. Earlier, he has also been to business school where he obtained MBA with a specialization in Finance and Strategic Management.

Khaled Ksaibati, PhD, PE Dr. Khaled Ksaibati is a Professor in the Department of Civil and Architectural and Construction Engineering at the University of Wyoming, and the Director of the Wyoming Technology Transfer Center. His specialization is in pavement design, maintenance and rehabilitation, pavement materials, traffic safety, operation and control. He has over 30 years of academic and professional experience in the field of transportation engineering. Dr. Ksaibati has managed and performed numerous research projects, conducted certification programs, participated in the development of the comprehensive safety plan for the state of Wyoming as a member of the Wyoming Safety Management System Committee, taught various courses in transportation and civil engineering and performed outreach. He has authored numerous journal and conference papers and technical reports.

Journal of Safety Research 84 (2023) 290–305

Contents lists available at ScienceDirect

Journal of Safety Research

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

A qualitative systematic review on the application of the normalization of deviance phenomenon within high-risk industries

Nejc Sedlar [a,*], Amy Irwin [a], Douglas Martin [a], Ruby Roberts [b]

[a] School of Psychology, William Guild Building, University of Aberdeen, Aberdeen AB24 2UB, United Kingdom
[b] Aberdeen Business School, Robert Gordon University (RGU), Aberdeen AB10 7AQ, United Kingdom

ARTICLE INFO

Article history:
Received 17 September 2021
Received in revised form 31 August 2022
Accepted 8 November 2022
Available online 18 November 2022

Keywords:
Normalization of deviance
Risk normalization
Systematic review
High-risk industry
Safety

ABSTRACT

Introduction: The concept of normalization of deviance describes the gradual acceptance of deviant observations and practices. It is founded upon the gradual desensitization to risk experienced by individuals or groups who recurrently deviate from standard operating procedures without encountering negative consequences. Since its inception, normalization of deviance has seen extensive, but segmented, application across numerous high-risk industrial contexts. The current paper describes a systematic review of the existing literature on the topic of normalization of deviance within high-risk industrial settings. *Method*: Four major databases were searched in order to identify relevant academic literature, with 33 academic papers meeting all inclusion criteria. Directed content analysis was used to analyze the texts. *Results:* Based on the review, an initial conceptual framework was developed to encapsulate identified themes and their interactions; key themes linked to the normalization of deviance included risk normalization, production pressure, culture, and a lack of negative consequences. *Conclusions:* While preliminary, the present framework offers relevant insights into the phenomenon that may help guide future analysis using primary data sources and aid in the development of intervention methods. *Practical Applications:* Normalization of deviance is an insidious phenomenon that has been noted in several high-profile disasters across a variety of industrial settings. A number of organizational factors allow for and/or propagate this process, and as such, the phenomenon should be considered as an aspect of safety evaluations and interventions.

1. Introduction

In January of 1986, after only 73 seconds of flight, Space Shuttle Challenger broke apart above the Atlantic Ocean. Following the incident, a Presidential Commission was established with the aim of uncovering the contributory factors and causes of the disaster. On a technical level, the vehicle's disintegration stemmed from the failure of eroded O-ring seals. This failure enabled the leakage of hot gas from the right booster rocket, culminating in structural collapse (NASA, 1986). Given the distinct and apparently avoidable nature of the failure, the question of why the issue had not been addressed at an earlier stage prompted an investigation into the broader context of the disaster, with a specific focus on the organizational factors that enabled the shuttle to be deemed safe for launch.

The nature of the disaster, coupled with revelations regarding NASA's organizational culture, led to the coining of the term 'Normalization of Deviance' (NoD) as a means of describing an individual/group's general acceptance of deviant actions or observations (Vaughan, 1996). Since its inception, the concept has seen extensive application across a broad range of industrial sectors and has been used to explain a number of other high-profile industrial incidents (e.g. Texas City Refinery [Dechy, Dien, Marsden, & Rousseau, 2018], Northwick Park drug trial [Hedgecoe, 2014]). To date, an extensive synthesis or compilation demonstrating the state of the literature has not been conducted. This is particularly noteworthy given that the category of high-risk industry is broad and highly varied, encompassing a diverse range of production aims, operating environments, and associated risks. As such, a systematic review across this category is needed to critically analyze and present how the concept of NoD has been applied within different settings and examine whether differences exist in the proposed theory, application, or intervention.

* Corresponding author at: APHF, School of Psychology, William Guild Building, University of Aberdeen, Aberdeen AB24 2UB, United Kingdom.
E-mail address: n.sedlar.20@abdn.ac.uk (N. Sedlar).

https://doi.org/10.1016/j.jsr.2022.11.005
0022-4375/© 2022 The Author(s). Published by the National Safety Council and Elsevier Ltd.

1.1. The Space Shuttle Challenger Disaster

A key finding outlined by the investigation into the Space Shuttle Challenger Disaster was that NASA and their engineers were in fact aware of the vehicle's structural weakness. Signs of erosion on the primary O-rings (rubber seals preventing the escape of hot gases between booster rocket segments) had been noted in 14 of the previous 24 missions across a period of 5 years (Starbuck & Milliken, 1988). In 9 of the final 10 flights prior to the disaster, engineers noted erosion on the primary O-rings, as well as evidence of gas leakage in most of these latter cases. The extent of the damage was further exemplified by evidence of erosion on the secondary O-rings, which represented a final safety mechanism and served as a redundant backup (NASA, 1986). These issues were highlighted by engineers on multiple occasions, however, NASA managers failed to implement corrective measures, deeming the risk of potential O-ring failure to be acceptable. Following the disaster, one of the managers responsible for the operations of the solid rocket boosters stated:

> *'Since the risk of O-ring erosion was accepted and indeed expected, it was no longer considered an anomaly to be resolved before the next flight ... the conclusion was, there was no significant difference in risk from previous launches. We'd be taking essentially the same risk on Jan. 28 that we have been ever since we first saw O-ring erosion.'* (Bell & Esch, 1987, p. 44, 47)

While the presence of the problem and its implications were acknowledged, the prior accumulation of successful launches fostered a tolerance towards the risk posed, enabling the issue to become relatively normalized. In spite of the increasing frequency and magnitude of erosion, as well as evidence of improper functioning, sub-contractor Thiokol suggested to NASA that the O-ring situation be considered 'closed' (Starbuck & Milliken, 1988). They presented the belief that it did not endanger flight safety and that the problem would not be resolved any time soon. This is particularly noteworthy given that the O-rings had previously been categorized as a "Criticality 1" component, wherein the component's failure is deemed likely to result in the loss of life or vehicle (NASA, 1986). Though the Criticality 1 of the O-rings was acknowledged as a launch constraint, it was consistently waived and rationalized as acceptable in light of prior mission successes (Starbuck & Milliken, 1988). Even on the eve of the launch, sub-contractor engineers who expressed concern over the potential for improper sealing under the low forecasted temperatures (−1 °C) were informed that they would need to provide evidence for their claims (Starbuck & Milliken, 1988). The engineers did not have enough data to determine the adequate functioning of the O-rings below 12 °C due to a lack of tests. This was not regarded by the leadership as an adequate cause for delaying the launch, a reluctance likely exacerbated the occurrence of multiple previous delays (Starbuck & Milliken, 1988).

1.2. Normalization of Deviance (NoD)

Within organizational contexts, safety culture describes an organization's collective underlying employee beliefs and values regarding personal and group responsibilities for safety and risk management (Everson, Wilbanks, & Boust, 2020). In reviewing the course of events preceding the Challenger disaster, it appears the O-ring failure merely represents the final fault within a sequence of issues on part of NASA's organizational system. Internal pressures stemming from financial costs, efficiency, political, and managerial demands, in concordance with increasing complacency and overconfidence, compromised the organization's safety culture and facilitated patterns of procedural deviations and risk acceptance (Vaughan, 1996). Diane Vaughan, a sociologist investigating the latent causes of the Challenger incident, coined the term 'Normalization of Deviance' (NoD) to describe how the compromised safety culture of NASA propagated itself to the point of disaster.

Vaughan (1996) defined NoD as the gradual process wherein, in the absence of perceived losses or harm, deviant practices become acceptable. A prominent feature of the phenomenon is the desensitization process, wherein frequent engagement in deviant practices facilitates the practice's normalization and perceived standardization within everyday operations. This normalized perception sets a new precedent for what is viewed as tolerable and routine, establishing a new normal from which further deviations may occur. In the absence of external intervention (e.g., external audits, change in procedures), this cycle of deviance is disrupted only when deviant behavior incurs an undesirable outcome.

According to Vaughan (1996), this process of normalized deviance provided the foundation for the Challenger disaster. The theory speculates that successes in the absence of overt negative consequences may cause an organization's members to develop overconfident perceptions of infallibility towards their existing programs, procedures, and leadership. In the case of the Challenger, risks associated with the shuttle's structural flaws, though likely a cause for concern to external observers, became imperceptible to many within the organization itself. Dillon, Rogers, Madsen, and Tinsley (2013) showcase this phenomenon in a temporal mapping of shuttle mission anomalies reported before and after each of the major disasters of the NASA program: Challenger in 1986, and Columbia in 2003. Data indicate a downward trend in reported anomalies over time, with initial missions displaying a far greater incidence of reporting by comparison to subsequent missions that preceded the disasters. The authors suggest the decrease in anomaly reporting likely resulted from anomaly normalization rather than resolution. With the accumulation of successful missions, some occurrences initially deemed anomalous became accepted as normal facets of operations and were no longer reported; implying that the more frequently an anomaly or near miss was observed without serious consequence, the greater the perception that no significant threat was being posed.

The progressive downgrading of anomaly importance was also discussed in the report published by the Columbia Accident Investigation Board (CAIB) (2003) following the Space Shuttle Columbia disaster. As with the Challenger, the downing of the Columbia resulted from a known issue; the shedding of insulation foam from one of the fuel tanks, previously observed within at least 30 prior missions (CAIB, 2003). While originally considered an in-flight anomaly, it does not appear to have been deemed a serious risk to flight safety. In fact, the frequency of observed shedding caused its significance to be downgraded from an in-flight anomaly to a so-called 'action item' only months prior to the disaster (CAIB, 2003). On the first of February 2003, a piece of foam debris hit the wing of Space Shuttle Columbia, puncturing a hole in the leading edge of the wing, and causing damage which proved terminal upon re-entry into the atmosphere.

1.3. System approach

Following the aftermath of the Challenger disaster, work by Vaughn proved a crucial contribution to the growing literature looking into accident causation as a product of complex systems. Banja (2010) notes that major disasters such as those of the space shuttles cannot be attributed to singular actions or individuals. They instead require the commission of numerous, often innocuous, mistakes that breach the organization's defenses. On this basis, it was understood that investigations and interventions should focus on systematic or latent errors, rather than attempt to pinpoint active individual errors. Reason (2000) describes how these latent errors foster an environment where error-provoking conditions (e.g., time pressure, inexperience) increase the likelihood of active failures (e.g., slips, procedural violations), whilst also undermining established safety measures that typically prevent hazards from resulting in losses (e.g., untrustworthy alarms, poorly designed procedures). The shuttle disasters, though physically speaking the product of technical failures, stemmed from issues relating to cognitive biases (i.e., the human vulnerability for systematic errors in information processing, perception and subsequent decision making; Kahneman, 2011). High-risk environments such as that of NASA, where technical problems and anomalies are part of the norm rather than an exception, are therefore particularly vulnerable to fostering desensitized perceptions of risk.

1.4. Industrial application

Following its inception within the aerospace industry, the concept of NoD has seen widespread application across numerous other high-risk industries, including oil and gas (Bogard, Ludwig, Staats, & Kretschmer, 2015), nuclear (Sanne, 2012), aviation (Paletz, Bearman, Orasanu, & Holbrook, 2009), and healthcare (Banja, 2010). As in the space shuttle disasters, the concept has been utilized to explain how deviant behaviors may become normalized within organizational contexts. Individuals engaging in deviant actions often appear largely unaware of their deviations or feel their deviance is justified; in either instance, their ability to accurately perceive and comprehend risk is compromised (Banja, 2010; Cavnor, 2018; Hase & Phin, 2015). Given the hazards, intrinsic safety concerns, and production pressures prevalent among high-risk industries and work environments, there is considerable interest in understanding the human mechanisms that may unknowingly propagate and facilitate unwanted outcomes.

Reviews of research into other phenomena such as teamwork and design characteristics have highlighted the significance of context-based variations with regards to industrial factors such as technology level, the focus of service, and the nature of production (Carter et al., 2019). To fully understand and utilize the NoD concept it is therefore important to synthesize research across a number of relevant high-risk domains to help ascertain the boundaries of the phenomenon and identify relevant commonalities, potential outliers, and general areas of interest that may help guide future research and intervention.

1.5. Aim

In recent years there has been a notable increase in the number of research papers on the topic of NoD from within various industry contexts. However, the majority of this research has been conducted independently and in isolation, with a lack of a defined overall theory. The present systematic review has the following objectives:

- Synthesize the existing literature in order to identify commonly discussed themes and components relevant to normalization of deviance.
- Determine the extent to which the central concept and associated factors can be generalized across high-risk industrial contexts.
- Identify gaps in the literature and develop suggestions for future research directions.
- Develop a preliminary conceptual model that would represent the manifestation and propagation of the NoD phenomenon within high-risk industry contexts.

2. Method

2.1. Search method

The literature search was conducted in February 2021. Four major databases were searched (Scopus, ProQuest, Web of Science, and Science Direct), using search terms: "normalization of deviance" OR "risk normalization" OR "normalization of risk" OR "deviance normalization" OR "normalization of deviance" OR "risk normalization" OR "normalization of risk" OR "deviance normalization." Risk normalization terms were included in the search criteria due to the concept's close association with NoD. All search results were then compiled, with all inter and intra database duplicates removed. The total number of unduplicated search results was 147.

2.2. Selection process

Based on the search criteria, 147 papers were identified. A two-step sifting process was then undertaken as seen in Fig. 1. Both sifting stages involved the application of exclusion criteria based on the title and abstract of the identified papers (as recommended in Siddaway, Wood, & Hedges, 2019). At the first stage, exclusion criteria related to the availability of the text, with four search results removed due to the unavailability of both the abstract and full text. A further 27 results were removed for being unrelated to NoD or risk normalization, as defined by Vaughan (1996). Specifically, these studies focused on biological normalization.

Of the remaining 106 search results, a further 40 were removed during the second sift where, based on the title and abstract, papers were excluded if they did not investigate NoD within high-risk industries. This choice of exclusion was due to the present review's focus on investigating safety-related deviations specifically within high-risk industries. While the NoD phenomenon is applicable across other industrial settings (e.g., finance, project management, retail) the motivations and consequences for deviating and risk normalization are likely to differ in the absence of overt physical safety concerns (Banja, 2010). High-risk industries, therefore, present a varied, but somewhat more homogenous, industry focus that more closely reflects the environment of NASA from which NoD originates. For the purposes of the present review, high-risk industries were defined as falling into categories such as transport (e.g., aviation and rail), healthcare, and process industries. As such, papers were excluded from further analysis if either the high-risk industrial setting was not apparent from the abstract, or if both industry and safety were not referenced in a relevant capacity.

At the final selection stage, the full text of the remaining studies was interrogated. Studies for which the full text was inaccessible or unavailable (20), were removed. The full texts of the remaining studies were then analyzed against the criteria from the initial sifts, with the further removal of studies that did not refer to NoD or risk normalization within the text. Four studies were removed due to a lack of clarity on the application of the concept, with insufficient detail available for meaningful analysis. To avoid repetition and maintain focus on the development of the phenomenon since its inception within the aerospace industry, a further four studies were excluded for solely discussing NoD with

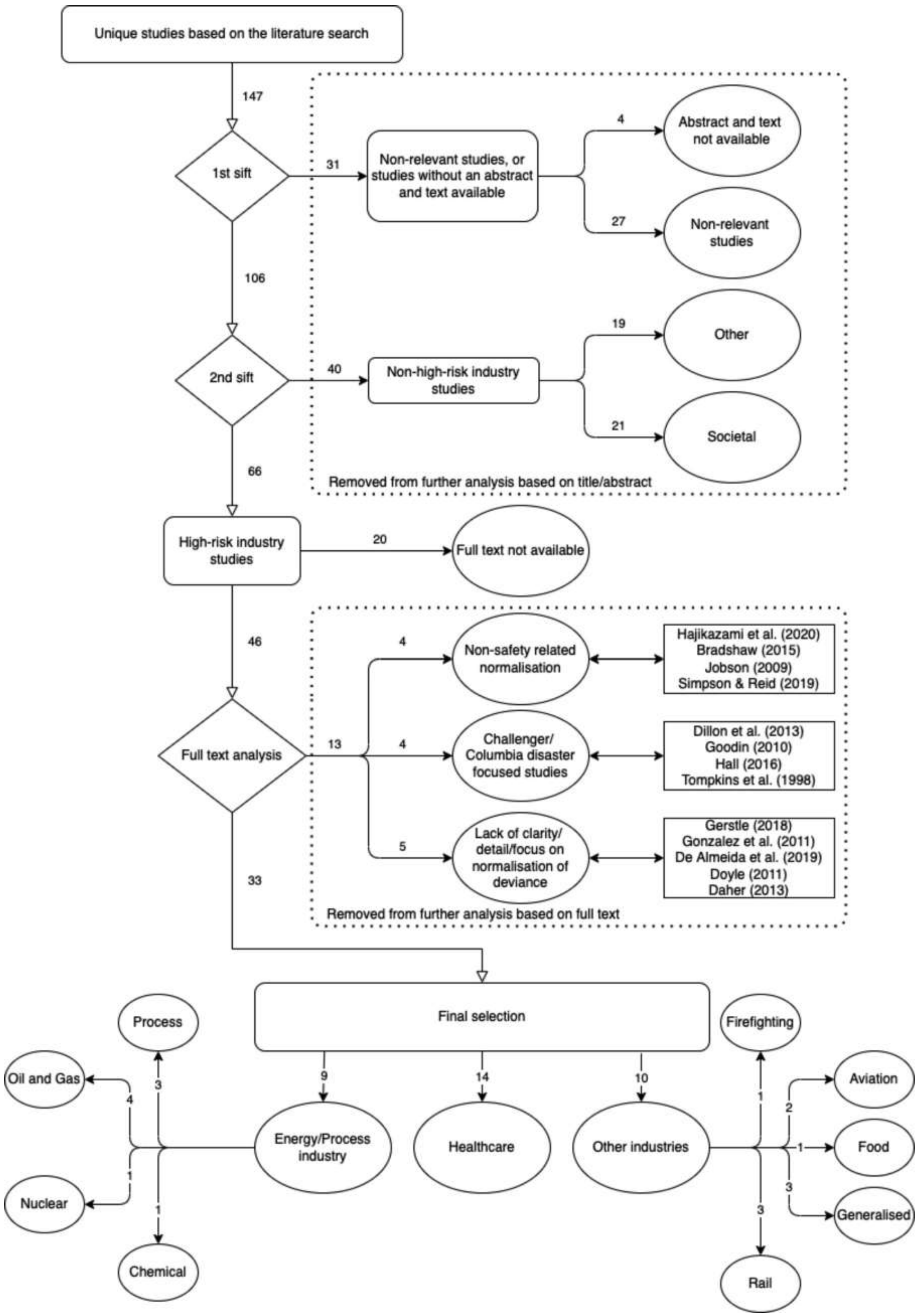


Fig. 1. Literature Selection Process Flow Chart.

reference to the space shuttle disasters. Finally, studies focusing on the normalization of deviance with no focus on safety were also excluded from further analysis.

2.3. Quality assessment

Out of the 33 articles meeting all of the above criteria, 27 were journal articles, 4 were articles from conference proceedings, 1 was a book chapter, and 1 was a master's thesis. Due to the nature of the existing literature on NoD being mostly conceptual in nature, as well as the aim of the present review being to understand the conceptualization of the phenomenon within the academic literature, no specific assessment tool of literature quality was used. These rely on evaluating the empirical integrity of studies based on factors relating to the research's validity and reliability (Siddaway et al., 2019); factors that are not applicable to conceptual papers or case studies. Instead of utilizing a quality assessment tool, presence within the aforementioned scientific databases (Scopus, ProQuest, Web of Science, and Science Direct) was used as a criterion of academic quality and therefore academic literature. Information on the publication and evidence type of each included study is displayed in Table 1.

3. Analysis

To comprehend the complex internal dynamics of high-risk industries, analysis required that the literature be broken down into comprehensive conceptual categories/components. As suggested by Hsieh and Shannon (2005) a directed content analysis approach (a method for summarizing large quantities of text via fewer content categories [Weber, 1990]) was used. This theoretical conceptualization of the phenomenon was used as a guide for the initial identification, coding, and categorization of data, as well as the subsequent development of an initial conceptual framework intended to encapsulate the reported interactions between the identified components.

Given that the majority of the identified literature did not solely focus on the phenomenon of NoD, the coding strategy within the present review required the initial identification of relevant text extracts from within each paper (as suggested in Hsieh & Shannon, 2005). These were identified by reading through the entire text and extracting sections which, directly or implicitly, referenced and/or discussed the NoD phenomenon. Sections were gathered and organized in a Microsoft Word document and were then coded by the first author on the basis of their semantic meaning, relevance, and relationship to NoD. Extract coding and subsequent categorization followed an inductive approach, with each code being generated on the basis of the content of the identified extracts (n = 25). Extracts and initial codes were discussed with the research team to explore the potential higher-order categories (n = 10), which were developed through the amalgamation of semantically/categorically similar codes (Elo & Kyngäs, 2008). Through the process of abstraction (Elo, Kääriäinen, Kanste, Pölkki, Utriainen, & Kyngäs, 2014), these categories were further refined until representative overarching categories encompassing the phenomenon as described and discussed across the identified literature were developed (n = 7). Individual category names were determined by conventional terminology used within the texts (e.g., production pressure, leadership), or were generated using phraseology intended to describe the category's subject matter (e.g., lack of negative consequences). Table 2 presents an overview of the components identified across the included studies. All components were represented across the main industrial sectors; however, some variations in component frequency across industries did emerge. These are discussed in section *4.2 Industry Comparison.*

To encapsulate the identified components from the current review and portray the nature of their interactions as illustrated across the identified literature a conceptual framework was developed (as seen in Fig. 2). The showcased component interactions within the framework were developed inductively through the re-reading of coded excerpts and the identification of reported links and interactivity.

The following excerpt from Arendt and Manton (2015) offers an example of the type of content that informed this identification:

> *"In this case, a senior operating manager put extreme pressure on his staff and workforce to generate production and numerous decisions were evident that put safety behind economics. This resulted in a low sense of vulnerability in operating staff due to the apparent priority of safety behind production. The low sense of vulnerability led to a "superman complex" on the part of some operations staff that encouraged workarounds...*"

The example excerpt portrays the components of leadership, production pressure, and risk normalization, and indicates their interactions. In this instance, the authors report how leadership actions were directly associated with increased production pressure and a low sense of vulnerability (amalgamated into risk normalization), resulting in subsequent workarounds among operating staff (deviances). All of these reported links can be noted within the present framework.

Four of the identified components (production pressure, procedure/environment design, leadership, and culture) displayed a notable number of interactions with one another and were reported to have similarly influential relationships on other elements within the framework, acting as moderating factors. Consequently, while maintained and discussed individually in terms of their features, relevance, and influence on NoD, these were grouped under the broader label of 'Organizational Factors.'

4. Discussion

The aim of the present systematic review was to synthesize the existing literature on the topic of safety-related NoD within high-risk industrial settings. It is made evident throughout the literature that the nature of deviance and NoD is highly complex within industry contexts, wherein a multitude of factors pertaining to organizational, social, and technical processes contribute to the phenomenon (Cavnor, 2018). These are influential to the development and propagation of NoD across its different components. Factors such as production pressure have the potential to influence a range of outcomes, including the likelihood of normalizing risk, the likelihood of deviating from set procedures, and the likelihood of initiating a pre-emptive response following a deviation. Within the present review, we have represented these interactions through the use of an initial conceptual framework which expands upon previous models of NoD by integrating the phenomenon of risk normalization. While these findings are only preliminary, and somewhat limited by the scope and nature of the phenomenon's academic literature, the framework may help in guiding further analysis with primary data sources.

4.1. Conceptual framework

The conceptual NoD framework (Fig. 2) offers a visual representation of the flow path an organization or a group may take from normal operations to the onset of a loss event as illustrated across the identified literature. As within previous models (Hajikazemi, Aaltonen, Ahola, Aarseth, & Andersen, 2020; Heimann, 2005), the present framework illustrates a cyclical progression, where the propagation of NoD is essentially self-sustaining. The cycle is main-

Table 1
Normalisation of Deviance Literature Categorised by Industry Sector and Evidence Type.

Study	Title	Industry Sector	Evidence Type
Arendt and Manton (2015)	Understanding Process Safety Culture Disease Pathologies - How to Prevent, Mitigate and Recover From Safety Culture Accidents	Process Industry*	Conference proceedings - Summary of 3 case studies evaluating process safety culture
Banja (2010)	The Normalization of Deviance in Healthcare Delivery	Healthcare	Journal article - Conceptual article
Bloch and Williams (2004)	Normalize Deviance at Your Peril	Oil and Gas	Journal article - Case study of condenser failure at a major refinery
Bogard et al. (2015)	An Industry's Call to Understand the Contingencies Involved in Process Safety: Normalization of Deviance	Oil and Gas	Journal article - Conceptual article
Cavnor (2018)	Fighting the Fire in Our Own House: How Poor Decisions are Smoldering Within the U.S. Fire Service	Firefighting	Thesis – Policy and incident analysis
Creedy (2011)	Quantitative Risk Assessment: How Realistic are Those Frequency Assumptions?	Process Industry*	Journal article - Conceptual article
Dechy et al. (2018)	Learning Failures as the Ultimate Root Causes of Accidents	Generalised Industries**	Book chapter - Conceptual article
Everson et al. (2020)	Exploring Production Pressure and Normalization of Deviance and Their Relationship to Poor Patient Outcomes	Healthcare	Journal article - Meta-synthesis of 7 qualitative closed claims studies from anesthetise database
Furey and Rixon (2018)	When Abnormal Becomes Normal: How Altered Perceptions Contributed to the Ocean Ranger Oil Rig Disaster	Oil and Gas	Journal article - Case study of the Ocean Ranger disaster
Geisz-Everson et al. (2019)	Cardiovascular Complications in Patients Undergoing Noncardiac Surgery: A Cardiac Closed Claims Thematic Analysis	Healthcare	Journal article - Incident report analysis (34 malpractice claims)
Golinski and Hranchook (2018)	Adverse Events During Cosmetic Surgery: A Thematic Analysis of Closed Claims	Healthcare	Journal article - Incident report analysis (25 incident claims)
Hase and Phin (2015)	The Normalisation of Deviance in the Oil and Gas Industry: The Role of Rig Leadership in Success and Failure	Oil and Gas	Conference proceedings - Conceptual article
Hedgecoe (2014)	A Deviation From Standard Design? Clinical Trials, Research Ethics Committees and the Regulatory Co-construction of Organizational Deviance	Healthcare	Journal article - Case study into a failed UK drug clinical trial
Heimann (2005)	Repeated Failures in the Management of High Risk Technologies	Generalised Industry**	Journal article - Conceptual article
King (2010)	To Err is Human, to Drift is Normalization of Deviance	Healthcare	Journal article - Conceptual article
Mast (2018)	Summary of the King County, Washington, West Point WWTP Flood of 2017	Process Industry*	Conference proceedings - Case study into a major failure at a wastewater treatment plant
McNamara (2011)	The Normalization of Deviance: What are the Perioperative Risks?	Healthcare	Journal article - Conceptual article
Mize (2019)	The Roundabout Way to Disaster: Recognizing and Responding to Normalization of Deviance	Chemical	Journal article – A collection of case studies illustrating NoD within chemical industries
Naweed et al. (2015)	Are You Fit to Continue? Approaching Rail Systems Thinking at the Cusp of Safety and the Apex of Performance	Rail	Journal article - Observation of driving and interviews, focus group interviews, scenario simulation exercise (28 participants)
Naweed and Rose (2015)	It's a Frightful Scenario: A Study of Tram Collisions on a Mixed-Traffic Environment in an Australian Metropolitan Setting	Rail	Journal article - Accident report review, observation, focus group exercise, interview (23 participants)
Odom-Forren (2011)	The Normalization of Deviance: A Threat to Patient Safety	Healthcare	Journal article - Conceptual article
Paletz et al. (2009)	Socializing the Human Factors Analysis and Classification System: Incorporating Social Psychological Phenomena into a Human Factors Error Classification System	Aviation	Journal article - Interviews (28 participants)
Pannick et al. (2017)	Translating Concerns Into action: A detailed Qualitative Evaluation of an Interdisciplinary Intervention on Medical Wards	Healthcare	Journal article - Qualitative evaluation of an intervention (ethnography and 2 focus groups)
Price and Williams (2018)	When Doing Wrong Feels so Right: Normalization of Deviance	Healthcare	Journal article - Conceptual article
Prielipp et al. 2010)	The Normalization of Deviance: Do We (Un)Knowingly Accept Doing the Wrong Thing?	Healthcare	Journal article - Conceptual article
Quinn (2018)	When "SOP" Fails: Disseminating Risk Assessment in Aviation Case Studies and Analysis	Aviation	Journal article - Conceptual article
Ruault et al. (2013)	Sociotechnical Systems Resilience: A Dissonance Engineering Point of View	Rail	Conference proceedings - Case study of a railway accident
Sanne (2012)	Learning From Adverse Events in the Nuclear Power Industry: Organizational Learning, Policy Making and Normalization	Nuclear	Journal article - Conceptual article
Scott et al. (2017)	Countering Cognitive Biases in Minimising Low Value Care	Healthcare	Journal article - Narrative review of PubMed original articles on cognitive biases in clinical decision making
Simmons et al. (2011)	Tubing Misconnections: Normalization of Deviance	Healthcare	Journal article - Review of 116 case studies within 34 reports

(continued on next page)

Table 1 (*continued*)

Study	Title	Industry Sector	Evidence Type
Stave and Törner (2007)	Exploring the Organisational Preconditions for Occupational Accidents in Food Industry: A Qualitative Approach	Food Industry	Journal article - Qualitative investigation of 54 accidents, including 24 interviews
Stergiou-Kita et al. (2015)	Danger Zone: Men, Masculinity and Occupational Health and Safety in High Risk Occupations	Generalised Industry**	Journal article - Review of 96 articles
Wilbanks et al. (2018)	Transfer of Care in Perioperative Settings: A Descriptive Qualitative Study	Healthcare	Journal article - Incident report analysis (19 transfer of care claims)

Note. Industry sector represents the papers industrial focus. Evidence type gives information on the paper's publication type, study type, and additional detail where appropriate.
* Industrial sector identified solely as process industry.
** Study either has no specific industrial focus, or the focus is not stated.

tained by the factors and conditions present within a given system, in this instance the high-risk industry context. In the absence of losses or negative consequences, and without adequate preemptive response to near-miss events, deviations and their associated risks become normalized through a feedback loop influenced by prevailing organizational factors (e.g., procedural shortcuts/-corner cutting repeatedly carried out in order to benefit production outputs). In this regard, individual instances of deviations may not be explicitly harmful, rather, it is the cumulative degradation of operating procedure that increases the likelihood of a major loss event.

Each of the identified framework components is defined and explored in relation to the relevant literature. These components should be understood as largely non-linear in their interactions, wherein the degrees of overlap and cumulative contribution is likely to vary depending on the specific industry contexts. For theoretical purposes, it should be assumed that the initial development of NoD within organizations begins when a pattern of deviating from an initial procedural baseline is first sustained.

4.1.1. Risk normalization

Existing literature typically uses the term risk normalization to describe the desensitization to risks present within one's environment, and in broader contexts offers an explanation for how societies come to accept known risks in order to remain operational. Schweitzer and Mix (2018), for example, discuss how risks associated with nuclear energy were largely normalized within French mainstream media in response to the 2011 Fukushima disaster. Public support for nuclear energy was generally unfazed following the incident, which Schweitzer and Mix rationalize to be largely due to the nation's heavy dependence on nuclear energy. Similarly, Luís et al. (2015) observed that increased awareness of coastal hazards appeared to inversely correlate with perceptions of risk regarding the phenomena; an effect that was particularly strong among permanent coastal residents. In this regard, normalization of risk may be largely seen as an adaptive response, facilitating functionality in the presence of circumstances outside one's control (Stave & Törner, 2007). In the industrial context, Stave and Törner refer to several organizational preconditions that aid in normalizing the presence of risk, citing, for example, how operators are often assigned high levels of personal responsibility despite possessing low levels of actual control over their environments and performance of tasks.

A core feature of the present theoretical framework is its integration of risk normalization within the NoD phenomenon, with risk normalization being accounted for as a contributory precursor to the initiation and subsequent acceptance of deviances. Though deviances may occur in the absence of risk normalization, it is unlikely that behaviors will be repeated if their associated risks are continuously perceived to be high. Risk normalization thus requires that individuals develop an increased risk threshold/tolerance wherein they lose the ability to accurately perceive vulnerabilities within their physical or procedural operating systems.

Periods of perceived successes, or at a minimum, periods absent of negative events may further encourage a loss of perceived vulnerability by increasing complacency and overconfidence in the safety of operations and the environment (Hase & Phin, 2015; Mast, 2018). Organizations that maintain a history of success may come to be perceived as "too big to fail" (Hedgecoe, 2014). Arendt and Manton (2015) describe this as a "superman complex," wherein a lack of attention to risk and safety prevents workers from perceiving vulnerabilities within themselves and their environment. Banja (2010) clarifies this illusion of invulnerability by pointing out that inherent system deviations, flaws, and weaknesses are generally inevitable, it is however the unpredictability

Table 2
Distribution of NoD components across the identified literature.

Industry	Sector	Study		Organisational Factors					
			Risk Normalisation	Production Pressure	Procedure/ Environment Design	Leadership	Culture	Lack of Negative Consequences	Pre-emptive Response
Energy/ Process	Chemical	Mize (2019)	X	X	X	X			X
	Nuclear	Sanne (2012)	X					X	X
	Oil and Gas	Bloch and Williams (2004)	X					X	X
		Bogard et al. (2015)		X	X	X	X	X	
		Furey and Rixon (2018)	X	X			X	X	X
		Hase and Phin (2015)	X			X	X		
	Process Industry*	Arendt and Manton (2015)	X	X		X	X		X
		Creedy (2011)	X					X	X
		Mast (2018)	X					X	
Healthcare		Banja (2010)	X	X	X	X	X		X
		Everson et al. (2020)		X		X	X		
		Geisz-Everson et al. (2019)		X					X
		Golinski and Hranchook (2018)		X			X		
		Hedgecoe (2014)	X	X	X		X		X
		King (2010)					X		
		McNamara (2011)		X		X			X
		Odom-Forren (2011)		X		X	X		X
		Pannick et al. (2017)							X
		Price and Williams (2018)		X	X	X	X	X	X
		Prielipp et al. (2010)	X	X	X			X	
		Scott et al. (2017)		X	X				
		Simmons et al. (2011)			X				X
		Wilbanks et al. (2018)		X					
Other	Aviation	Paletz et al. (2009)	X					X	
		Quinn (2018)			X				
	Firefighting	Cavnor (2018)		X	X	X	X		X
	Food Processing	Stave and Törner (2007)	X		X				
	Generalised Industries**	Dechy et al. (2018)		X			X		X
		Heimann (2005)	X	X				X	
		Stergiou-Kita et al. (2015)	X	X			X		X
	Rail	Naweed et al. (2015)	X	X	X				
		Naweed and Rose (2015)		X			X		
		Ruault et al. (2013)			X				

Note. X denotes the component(s) that were identified within each study, and which contributed to the conceptual framework.
* Industrial sector identified solely as process industry.
** Study either has no specific industrial focus, or the focus is not stated.

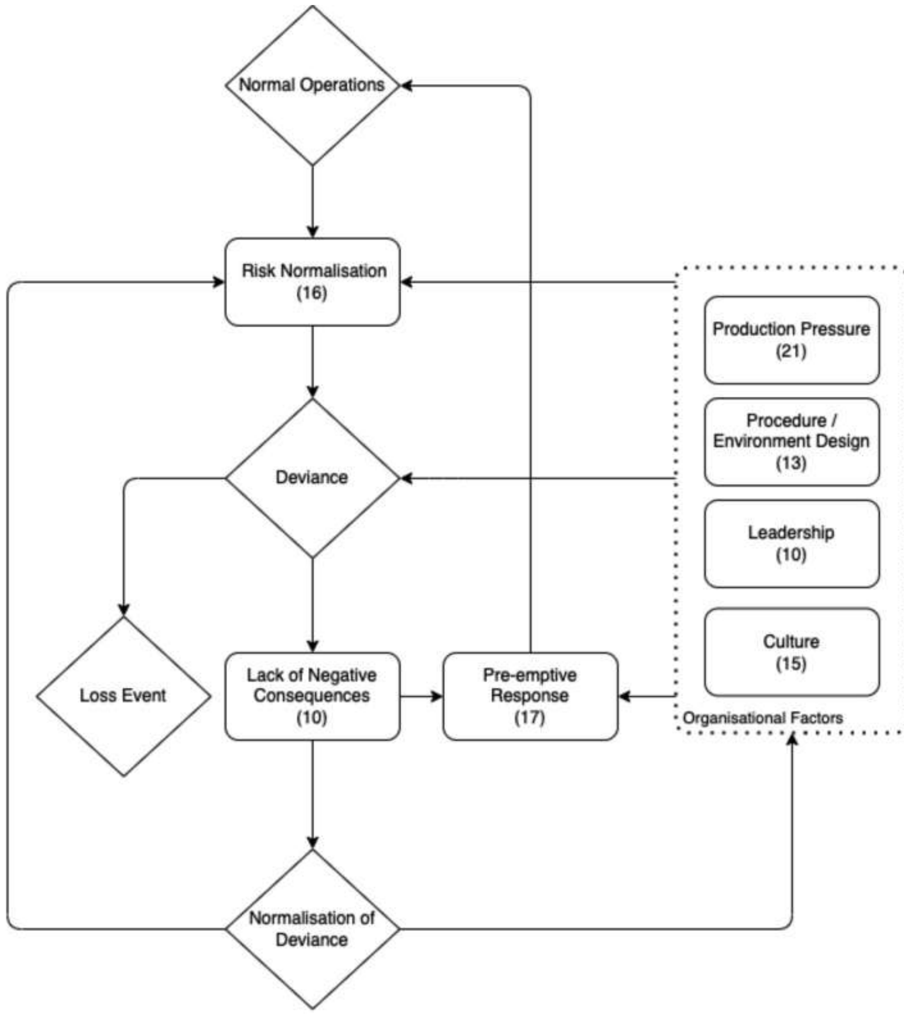


Fig. 2. Conceptual Framework of NoD Based on the Present Systematic Review. *Note.* (n) represents the number of individual studies within which the category was identified.

and infrequency with which these result in serious incidents that encourages complacency.

Under such circumstances, desensitization to hazards can lead to the acceptance of increasing levels of risk. Hase and Phin (2015) describe this process as relatively mundane, innocuous, and largely imperceptible, given the gradual manner in which it develops. Creedy (2011) moreover highlights the temporal nature of the phenomenon in observing how deviations in standard operating procedure often parallel the time elapsed following a past incident. Paletz et al. (2009) similarly outline the dangers of complacency among experienced pilots, who report becoming accustomed to the risks of flying in bad weather conditions, and demonstrate greater engagement in risky behavior than their less experienced counterparts.

An additional variable that has been noted to impact perceptions of risk is the introduction of new protective measures or system safety barriers. These represent the physical and non-physical initiatives used to enhance the safety of operations and mitigate unwanted outcomes. The introduction of a new protective measure generally increases perceived safety, which may unwittingly encourage employee perceptions of system invulnerability (Mize, 2019; Prielipp, Magro, Morell, & Brull, 2010). In other words, new protective measures may be viewed as solutions rather than fail-safes to known problems. Their introduction may therefore incentivize deviations in an attempt to bypass prior safety demands and maximize production efficiency (Banja, 2010; Mize, 2019; Prielipp et al., 2010).

4.1.2. Organizational factors

For the purposes of the present model, several of the identified components are encapsulated under the category of organizational factors; specifically, the components of production pressure, procedure/environment design, leadership, and culture. These components, and their relevance within the organizational context, were often discussed in tandem, and as interconnected facets which are influential on one another. From the organizational standpoint, it is the accumulation of these organizational components that contributes to the normalization of risk, propagation of deviance, and failure to respond adequately to early warning signs (i.e., pre-emptive response).

4.1.2.1. Production pressure. Broadly speaking, production pressure refers to both overt and covert organizational demands and emphasis on output efficiency (Everson et al., 2020). Issues with production pressure typically arise due to conflicts between the demands of safety and production. This conflict is complex and well documented within the realm of high-risk industries where production pressure is commonly discussed as a key contributory factor in industry accidents (Goh, Love, Brown, & Spickett, 2012; Mohammadi & Tavakolan, 2019; Probst & Graso, 2013).

The consensus across the high-risk industry safety literature is that.

production pressure and safety are akin to antagonist agents, whereby increased attention to one often causes detriment to the other (Cavnor, 2018). This idea has been discussed further by Heimann (2005) with reference to type I and type II errors. In principle, high-risk industries are generally cited as being averse to committing Type I errors (active errors of commission), where implementing an incorrect policy or course of action results in failure. Heimann (2005) notes that Type I error aversion is indeed often present initially within organizations, which typically begin operating with low thresholds of risk tolerance so as to create the impression of a functionally safe system. Under such conditions, accidents are generally infrequent and less severe, which encourages focus to shift towards the elimination of Type II errors of omission (e.g., the use of unnecessary measures that are costly to efficiency and productivity). This desire for increased productivity and efficiency acts as the driving force for deviations and shortcuts to be undertaken by operators (Dechy et al., 2018).

In the absence of immediate negative outcomes, organizations and individuals may become susceptible to the aforementioned influence of risk normalization and may feel justified in re-evaluating and altering their potentially costly and overly 'conservative' thresholds. As a result, a so-called "cycle of failure" is propagated, wherein continued deviation from initial standards in pursuit of efficiency ultimately culminates in major failure (Heimann, 2005).

Naweed, Rainbird, and Dance (2015) and Naweed and Rose (2015) reference how organizations within the rail industry emphasize punctuality and 'on-time performance,' describing the heightened pressure experienced by operators running behind schedule as a condition under which they report greater susceptibility to taking shortcuts and violating procedures to recover lost time. Specifically, Naweed et al. (2015) note that driver interpretation of signals has shifted over time in order to facilitate faster train movement. This behavior has increased the likelihood of 'signal passed at danger' (SPAD) events, wherein a train passes a stop signal without explicit allowance to do so; a practice which, when performed frequently, is associated with an increased risk of derailment or collision.

Pressures associated with having to accomplish more with less are exemplified in a number of other cases throughout the literature, such as Mize (2019) who outlines a case of operators within a chemical plant violating standard procedure to meet increasing production targets, and Cavnor (2018) who notes evidence of firefighters skipping safety checks prior to entering compromised structures to achieve tactical goals more efficiently. Within healthcare, McNamara (2011) and Arendt and Manton (2015) cite managerial and institutional pressures on productivity and maintenance of the operating room on schedule as factors typically accountable for the introduction of deviations. Clinicians may, for example, disconnect vitality monitors prior to the end of a procedure, or before a patient has fully emerged from anesthesia, in order to speed up the turnover process (Prielipp et al., 2010). However, Bogard et al. (2015) state that these shortcuts and deviations rarely result in serious process safety issues and often directly facilitate the organization's target progression.

It is important to also acknowledge, however, that the relationships between production pressures, safety, and Type I and II errors vary across individual industries. Specifically, it is somewhat more complicated in occupations such as healthcare and firefighting where circumstances may cause production pressures to be explicitly tied to physical safety. In these contexts, both Type I and Type II errors may result in harm or loss of life, either through the initiation of incorrect/unsafe treatment, or the withholding of correct treatment (Price & Williams, 2018). In this regard, motivations for deviating may differ in some respects from traditional process industries given that production demands are directly concerned with minimizing the harm done. Insights from clinician reports regarding their rationale for procedural deviations reflect this, with individuals often citing a desire to minimize patient discomfort and eliminate unnecessary or counterproductive measures as being justification for procedural deviations (Banja, 2010; Scott, Soon, Elshaug, & Lindner, 2017).

Deviations guided by a patient-centric or 'greater good' approach may provide justification for the normalization of shortcuts, given the perception that these might offer a means of attending to more patients, or provide the opportunity to prioritize those with more serious conditions (Price & Williams, 2018). Cavnor (2018) similarly notes a form of 'melioration bias' (a tendency towards alternatives seen as preferable in the short-term) in regard to certain operating procedures; namely the correct wearing of PPE, which firefighters have claimed hinders movement and impedes life-saving action.

Among process industries, common generalized instances of justified deviance may be observed in shortcuts performed by operators seeking to improve productivity; not for explicit and immediate personal gain, but rather as a means of satisfying broader organizational demands (Mize, 2019). These deviations, intended to maximize productivity, may be further compounded by a pre-existing rule ambiguity and unfamiliarity, particularly for tasks that do not involve standardized checklists (Banja, 2010; Mize, 2019; Stergiou-Kita et al., 2015).

4.1.2.2. Procedure/Environment Design. Within many high-risk industries, special considerations must be made for the design of both the physical work environment and the nature of processes and procedures in order to facilitate productivity and reduce risk (Gambatese & Hinze, 1999; Marsden & Green, 1996; Park & Jung, 2003; Reuter & Camba, 2017). These considerations may include placing emphasis on computerization and automation to streamline processes and reduce workload (Marsden & Green, 1996; Park & Jung, 2003; Wang & Ruxton, 1997), standardizing operating procedures (Kurt, Arslan, Comrie, Khalid, & Turan, 2016), and evaluating and making provisions for fail-safes that will mitigate unintentional error or sudden failure (Garrick & Morey, 2015).

Procedures are agreed-upon methods of work, intended to ensure that tasks are performed in an efficient, controlled, and safe manner (Marsden & Green, 1996). Issues with procedures generally arise when these are deficient in designating activities or enabling the successful accomplishment of tasks (i.e., due to being inaccurate, outdated, incomplete, or overly complex and demanding; Park & Jung, 2003).

Throughout the identified literature, inappropriate implementation of procedures and poor environmental designs were frequently cited as contributory to the initiation and maintenance of deviant behavior. The reasoning provided was that under time and production constraints, procedural or environmental limitations often provide justification for deviances and violations (Mize, 2019; Price & Williams, 2018); with some operators arguing that perfect compliance to rules and standards makes it impossible to achieve productivity demands (Banja, 2010).

Price and Williams (2018) state that the very presence of deviance inherently signals potential flaws within a system's environment or work process. In reference to healthcare, they illustrate how factors such as inconveniently placed hand hygiene stations decrease hygiene compliance, and even minor obstacles such as malfunctioning barcode scanners disrupt entire workflows and prompt the skipping of the scanning process in order to achieve on-time administration of medication.

In some organizations, Quinn (2018) argues that rather than amending poor procedure and environmental design, deviances become a normalized and expected practice intended to "fill in the gaps" of standard operating procedures. In other instances, there may be an initial lack of overt procedural rules or adequate resources that precipitates compensatory individual and team solutions (Cavnor, 2018; Hedgecoe, 2014; Stave & Törner, 2007).

A further weakness explicitly referenced within the literature is that of maladaptive alarm/warning system design resulting in the experience of alarm fatigue. Bogard et al. (2015) highlight how overexposure to alarms causes desensitization and loss of vulnerability towards these. Frequent alarm exposure, particularly when false, normalizes the alarm presence as routine, prompting a lack of response. Poor implementation of an alarm system may also encourage procedural deviations intended to circumvent system activation, as evidenced in the railway industry where cautionary signals have been largely devalued by drivers. Naweed et al. (2015) report that on some journeys it is routine to operate in a continuous "alarmed" state without ever being clear of cautionary signals.

4.1.2.3. Leadership. Within organizational contexts, leadership describes a variety of multifaceted management roles that encompass a range of responsibilities, styles, and behaviors depending on the context and the leader's respective level of responsibility (Denis, Langley, & Rouleau, 2010; Pilbeam, Doherty, Davidson, & Denyer, 2016). Senior management and leadership are responsible for a range of decision-making directly associated with safety, including training and resource allocation and investment, oversight, scheduling, and maintenance of equipment (Kelloway, Nielsen, & Dimoff, 2017; Reason, 2000), as well as role modeling and influencing worker attitudes and behavior (Flin & Yule, 2004; Pilbeam et al., 2016).

Reason (2000) has been particularly critical of the role of leadership, identifying decision makers and line management as a core element of any productive system. Reason further argues that many organizational accidents can be traced back to deficiencies in managerial decision-making. Similarly, within the identified literature, Everson et al. (2020) describe the nature of an organization's safety culture to be largely determined by the approaches taken by executive leadership. Mize (2019) notes that it is the leadership of an organization that is responsible for setting expectations for employee attitudes and behavior, with the responsibility of providing sufficient training and reinforcement of operational discipline. In this regard, leadership failures in the maintenance of a system's risk mitigation often play a crucial role in facilitating NoD (Bogard et al., 2015). Actions by leadership are generally perceived as having top-down consequences, wherein poor leadership decisions are filtered through the various levels of an organization, causing damage to an organization's operational safety and general safety culture (Hase & Phin, 2015).

Supervisors may, for example, avoid or choose not to discipline operators who engage in shortcuts and deviations in order to simplify processes, reduce workloads and increase production speed (Bogard et al., 2015). To conserve resources, some organizations may also fail to provide adequate training by limiting the amount of time available for operators to familiarize themselves with new tools or procedures (Geisz-Everson, Jordan, Nicely, & McElhone, 2019), or in some cases, through the active teaching of already normalized shortcuts and deviations (Banja, 2010; Odom-Forren, 2011). A key issue here is that in such instances deviations performed by authority figures typically go unchallenged (McNamara, 2011).

Actions such as these facilitate NoD by instilling a "production over safety mindset" when led by the example of decision-making authority figures (Cavnor, 2018). When an organization places excessive demand on economics, leadership may fail to uphold process safety as a core value, resulting in the dismissal of warning signs and the encouragement of workarounds in the interest of production (Arendt & Manton, 2015; Dechy et al., 2018). Younger, and more inexperienced employees are particularly vulnerable to production demands given their limitations in power, agency, and inability to accurately comprehend or question safety procedures (Banja, 2010; Stergiou-Kita et al., 2015). Furthermore, it is suggested that observations of issues and weaknesses may be minimized when reported to supervisors/higher authorities due to a fear of repercussion or punitive action from leadership and/or a general lack of confidence that voicing concerns would lead to actual change (Banja, 2010; Furey & Rixon, 2018; Odom-Forren, 2011).

4.1.2.4. Culture. Culture describes the collective nature of an organization's underlying values, beliefs, expectations, and perceptions that guide and inform individual and group behaviors and practices (Everson et al., 2020; Van den Berg & Wilderom, 2004). Van den Berg and Wilderom (2004) describe organizational culture as the "glue" which binds together an organization. When it comes to NoD, the significance of culture is pertinent with regard to understanding how formal and informal attitudes and decision-making processes enable deviances to take place and be normalized. As previously mentioned, within Vaughan's investigation, understanding the culture within NASA as a social organization was crucial to helping identify the rationale and motivations, particularly from a managerial standpoint, behind the decision-making that took place prior to the disaster. Vaughan specifically outlined how NASA's culture was one with a "major preoccupation" with bureaucracy, which failed to realistically account for safety, cost, efficiency, and productivity demands (Vaughan, 1996).

Throughout the identified literature, organizations were cited as possessing individual identities that shaped the nature of group dynamics within work settings (Cavnor, 2018; Price & Williams, 2018; Stergiou-Kita et al., 2015). These social identities, while influenced by organizational demands, were said to also exist independently as products of an organization's history, projected image, and working environment. Cavnor (2018) for example, extensively discusses the cultural and social implications of firefighting, describing how beliefs shared among firefighter groups often encourage behaviors that favor risk acceptance. As a result, authors frequently identified the importance of understanding culture as a variable that may inadvertently sustain unhelpful practices (Everson et al., 2020; Hase & Phin, 2015; Price & Williams, 2018; Stergiou-Kita et al., 2015).

An organization's history, externally projected image, and working environment, were said to be of particular significance to culture, as these often become integrated with the individual identities of work personnel, fostering traditions and operational practices that may be both adaptive and maladaptive (Cavnor, 2018; Stergiou-Kita et al., 2015). Price and Williams (2018), note how healthcare workers traditionally promote a standard of individual perfection that ultimately distracts from addressing wider underlying issues relating to equipment, systems, or procedure. Similarly, the distinct social image of firefighters may promote mutual trust, courage, and concern for the safety of others, however, it may also encourage excessive and unreasonable risk-taking (Cavnor, 2018; Stergiou-Kita et al., 2015). In this regard, Hedgecoe (2014) notes that the everyday culture of work groups may often inadvertently accommodate and normalize risk; leading organizational cultures to foster environments where normalized deviances are mundane occurrences rather than exceptions (Hase & Phin, 2015). Stave and Törner (2007) similarly describe the working practices of a team as the product of continuous internal negotiations, which may lead to risk acceptance within work cultures that do not prioritize safety.

Alternatively, some organizations were said to also manifest a 'silo effect,' characterized by a lack of cohesion and interaction between workgroups and departments. These experience fragmented individual group cultures, operating on independent standards so as to meet their own needs rather than a common shared agenda across the organization (Golinski & Hranchook, 2018). This may result in inconsistent practices across an organization, wherein a lack of communication perpetuates rule unfamiliarity and deviations in practice. Thus, despite the aforementioned potential for unwanted consequences, a shared social identity among employees is typically seen as desirable within the organizational context (Golinski & Hranchook, 2018).

Helmreich and Merritt (2001) described how organizational culture represents a 'complex framework' composed of national, organizational, and professional attitudes and values. It should therefore be noted that while frequently referenced, given its breadth and complexity, the concept of organizational culture is not always clearly defined. This has also been pointed out within wider literature where the notion of organizational culture has been criticized for lacking clarity and definition (Van den Berg & Wilderom, 2004). Moreover, there is debate as to whether an organization may truly be defined under a singular overarching cultural identity, or whether its culture should be understood as the product of several collective subcultures and group identities across various departments and chains of command (Willcoxson & Millett, 2000). The present review does distinguish the component of culture as somewhat independent of leadership and production pressure, which may traditionally be considered subsets of the organizational culture. While, as with all the themes discussed, there is likely to be overlap in the actual manifestation of components within real-world settings, culture as it pertains to NoD was in many instances flagged as a unique contributor to the phenomenon, particularly with regards to the organizational culture surrounding safety (Arendt & Manton, 2015; Cavnor, 2018; Everson et al., 2020; Stergiou-Kita et al., 2015).

4.1.3. Lack of negative consequences

In general literature, the relevance of perceived negative consequences has been explored primarily within the realms of human risk perception, specifically with regard to the human evaluations and management of risk on individual and societal levels (Creyer, Ross, & Evers, 2003; Johnson & Tversky, 1984; Sitkin & Pablo, 1992). The perceived lack of negative consequence works in tandem with the previously discussed issue of unnoticed, latent errors/failures that accumulate over lengths of time (Dekker & Pruchnicki, 2014). Similarly, Rasmussen (1997) highlights the issue of reliability being mistaken as an indicator of safety (i.e., that something is good enough simply by virtue of its past successes). As with risk normalization, the absence of consequence fosters a 'presumption of safety' that impairs the collective and individual abilities to detect risk (Hedgecoe, 2014).

Unsurprisingly, the absence of negative consequences is consistently cited as an integral element of NoD and is discussed extensively in relation to deviance and risk perception desensitization (Price & Williams, 2018). It is widely understood that perceptions of risk and risky behavior are subjective and may be positively or negatively evaluated depending on the framing and evaluative points of reference used (Kahneman & Tversky, 1979; Tversky & Kahneman, 1985). With regards to NoD, when a deviation fails to result in an apparent adverse outcome, it may be seen as an indication that initial standards or procedures are over-conservative (Creedy, 2011). This perception, or framing, justifies deviations as acceptable evolutions of the productive process, wherein behavior is merely adapting to maximize efficiency; a notion that is parallel and complimentary to Rasmussen's "migration model" of the adaptive processes undertaken by organizations attempting to maximize productivity and profitability. Rasmussen notes that this behavior is typical of sociotechnical systems given the pressures and constraints under which they operate (Rasmussen, 1997). These pressures encourage deviations in attitude and action, which in the absence of consequence, are highly prone to repetition (Paletz et al., 2009), and acceptance by both workers and management throughout the organization (Bogard et al., 2015); with perceived benefits to production additionally de-incentivizing intervention and enforcement of discipline (Bogard et al., 2015).

4.1.4. Pre-Emptive Response

Perrow (1999) famously argued that "normal accidents" or failures within highly complex systems, such as those found within high-risk industries, are likely to be unavoidable given the complexity of the system's components (machinery/equipment, operators/employees, procedures etc.) and the manifold possibilities for these components to interact and result in failure. Turner and Pidgeon (1997), however, denotes that incidents are nearly always preceded by warning signs and claims that major accidents require preconditions to be present, often for extended periods of time. Turner argues that accidents can be prevented if these are identified and appropriately dealt with. Reason (1990) describes these preconditions as "resident pathogens," that is, latent failures which may combine with any number of factors such as active failures (human error and violations) or system faults to produce an adverse outcome. Reason, in agreement with Perrow, states that highly complex systems do contain a greater number of resident pathogens, and will thus be more susceptible to failure; however, he also asserts that these can be monitored, assessed, and understood with adequate system knowledge (Reason, 1990).

Pre-emptive responses to risks have in more recent years been discussed in terms related to organizational resilience (i.e., the ability of an organization to identify, cope with, and learn from incidents and failures and adjust positively under challenging conditions; Hutter, 2010; Vogus & Sutcliffe, 2007). A well-known general approach for pre-emptively dealing with hazards within high-risk work environments involves the implementation of a hierarchy of controls framework, intended to identify and prioritize hazards and their respective intervention strategies (Barnett, 2020; Hopkins, 2006; Morris & Cannady, 2019). Depending on the hazards present, a range of control measures with various levels of efficacy can be implemented. These typically include elimination (physical removal of a hazard), substitution (replacement of a hazard with a less dangerous alternative), engineering controls (isolating a hazard from workers, often through technology), administrative controls (changes in work practices) and use of PPE (use of personal protective equipment; Morris & Cannady, 2019).

With regards to NoD pre-emptive response refers to measures taken to anticipate, identify, and prevent the propagation of maladaptive deviance. This encompasses both the nature of proactive measures used to detect and respond to near-misses/signals, as well as the quality of retroactive learning following an incident or near-miss (Cavnor, 2018). The importance of identification and learning is particularly relevant given that pre- and post- investigation processes are both susceptible to normalization biases. Initial signals normalized in advance of an incident may be subject to the same framing after an accident, often in an attempt to cover up wrong-doings and minimize responsibility (Furey & Rixon, 2018; Sanne, 2012). Moreover, signals of potential disaster can manifest at various time intervals and across varied locations, which may cause individuals within an organization to view pre- and post-events from a detached personal level (Simmons, Symes, Guenter, & Graves, 2011).

Ideally, behavioral deviances, warning signals, and near-misses should always be accounted for. However, in light of the potential associated effort and costs, individuals may be biased towards discounting originally proposed risks when there is a lack of incentive for reporting/speaking up (Banja, 2010; Cavnor, 2018; Sanne, 2012). Furthermore, while some organizations outline policies regarding what events need to be reported, criteria are often subjective and dependent upon voluntary input (Dechy et al., 2018).

Another component detrimental to pre-emptive learning is that of inappropriate safety reporting systems. Pannick et al. (2017) describe a healthcare setting wherein the formal mechanism for recording incidents was an online reporting system that was difficult to use and poorly suited for this purpose, with long delays in the processing of even relatively simple issues; resulting in common/recurrent problems being left unreported and normalized within everyday practice. Failure to document warning signs or procedural changes, even those perceived as positive workarounds and innovations, enables these to remain unchallenged, and set precedents for procedural ambiguity and shifting norms (Mize, 2019). When incident analysis does take place, Price and Williams (2018) specifically outline the importance of appropriate system/process investigation in order to avoid simply blaming individual behaviors or components. They cite how patient safety literature demonstrates the efficacy of addressing issues from a system, rather than a human, perspective.

4.2. Industry comparison

While the healthcare industry represents the largest single industrial sector among the identified literature, many of the core components and patterns of NoD appear generally consistent across the industries accounted for within this review. Production pressure was among the most consistently referenced and discussed components across the industry literature, however, its prevalence in healthcare (11/14 healthcare papers) is particularly noteworthy. Another distinction between healthcare and other industries can be seen in the apparent lack of reference to risk normalization within the identified healthcare literature (3/14 healthcare papers, by contrast to 8/9 papers within the process industry and 5/10 within other industries).

These differences may be due to a number of reasons however a comparative analysis of healthcare to other high-risk industries by Gaba (2000) extensively discusses several key structural differences between healthcare and other high-risk industries; including a lack of centralization, regulation, investigation, and reporting by contrast to other high-risk industries such as aviation, oil and gas, nuclear, and chemical manufacturing (Gaba, 2000; Hudson, 2003). While issues of production demands may be more openly vocalized, issues surrounding the conscious or unconscious normalization of risky behaviors or malpractice may be more covert within healthcare, potentially due to the more explicit medical attitudes regarding individual responsibility and blame (Gaba, 2000; Hudson, 2003; Price & Williams, 2018). Gaba also describes how healthcare systems may often enable "structural secrecy," wherein problems can be "defensively encapsulated" within respective units or departments and blame may be shifted elsewhere.

Depending on the industry, the nature of risk and risk management will also vary, given the variations in potential outcomes associated with hazards and risky behavior, and whether these are likely to only affect workers themselves or have consequences for others (Banja, 2010; Cavnor, 2018; Hudson, 2003). In this regard, healthcare, while conscious of medical dangers, may be said to have a more reactive focus to managing dangers, with some proactive considerations; given that medical personnel manage a wide range of unpredictable dangers and hazards experienced by others but rarely themselves. Other high-risk industries, such as oil and gas and nuclear, may be described as having more proactive approaches, given that their workers must, by contrast, contend with an array of potential risks that have the potential to be hazardous to themselves, their colleagues, and the wider society (Hudson, 2003).

Furthermore, as previously highlighted, the production outcomes and demands for service industries, particularly public service industries such as healthcare and firefighting, by comparison to process industries, should be accounted for; specifically with regard to understanding the nature of industry outputs (i.e., minimizing harm and saving lives vs. maximizing physical productivity and efficiency). This difference is not undermined within the present model, as 'production pressure' does not specify the type of motivation it describes, but rather refers to any form of medium or motivation by which perceived output demands are prioritized and likely to encourage deviations in practice. Arguably, these descriptive differences may also fundamentally be merely a simple case of categorization and semantics; however, given the complexity of organizational contexts and individual experiences, the importance of understanding and accounting for the unique variables within individual organizations should not be understated.

4.3. Theoretical contribution

The current academic literature on the topic of NoD indicates that research has been largely independent and fragmented across a variety of sectors. The present systematic review synthesizes literature from a variety of high-risk industries in an attempt to ascertain common components of the phenomenon and introduce a new conceptual framework that seeks to encapsulate the manner in which the phenomenon has been presented and discussed. One of the main theoretical contributions of the present paper is the integration of risk normalization within a model of NoD as an integral component in the development and maintenance of the phenomenon. While entirely conceptual at present, this model suggests that intervention methods for the prevention of harmful NoD may need to focus on the initial normalization of risk; more specifically, ensuring that operator perceptions of risk do not degrade over time. Furthermore, the present review highlights the impact of organizational factors on the propagation of NoD. While these are likely to be context-specific and variable, they point toward factors that should be considered when investigating NoD within the high-risk industry context (e.g., Is production pressure encouraging deviations and short-cuts? Does culture within the work environment discourage the reporting of near misses?).

4.4. Limitations

While the present review and conceptual model are based on the current academic literature on the topic of NoD, there are some notable limitations that should be considered. Namely, the model is preliminary and untested and based on a relatively small sample of academic literature. While a systematic method of analysis was utilized for its creation, with directed content analysis often being used to develop conceptual models (Elo et al., 2014), an inherent level of subjectivity and potential bias exist both in the initial coding and subsequent categorization and model mapping. This may have been particularly pronounced within the present review where only one coder was used (first author). However, the coder remained open to new and alternative codes or potential categorizations. Future research will address this limitation through the testing of the preliminary framework in real-world settings (e.g., case studies).

A further consideration that should be addressed is that NoD with regards to the high-risk industry has so far been mostly discussed within the confines of conceptual articles or on the basis of accident reports and case studies. Of the 33 identified studies within the present review, 21 utilized secondary data, and of the remaining 13, seven were based on case studies. As such, the majority of the reviewed literature consisted of studies that did not present novel data or findings, but rather built upon and discussed the relevant topics from a number of industry perspectives. Though these offer valuable insights and points of consideration, the presence of primary data within this review has been severely limited. The review is therefore confined to focusing on the phenomenon from a largely conceptual and observational standpoint. The lack of applied research on the topic is an issue highlighted by several authors, who acknowledge that many of the observations and speculations, though theoretically reasonable, are yet to be actively quantified in terms of real-world intervention and risk reduction (Arendt & Manton, 2015; Bogard et al., 2015; Cavnor, 2018; Creedy, 2011).

4.5. Future research

These limitations suggest testing of the present model is required prior to any serious or consequential application. Specifically, the framework should be tested and quantified with respect to real-life settings, individual industries, and primary data, in order to make further refinements and provide validity. This could be accomplished through the analysis of incident reports, or by using primary data obtained from interviews or direct observations. Additionally, applied methods of analysis could be used in order to test specific components of the framework.

Of specific interest would be investigations into the development of risk normalization at the individual level, the specific factors accelerating this normalization, and the examination of potential interventions intended to reduce the likelihood that normalization of risk will lead to the initiation and normalization of deviance. Furthermore, the effect of the absence of negative consequences following a deviation could be investigated, with specific reference to the subsequent likelihood of engaging in said deviation. As suggested by Bogard et al. (2015), behavioral research is desperately needed to support the mostly conceptual nature of the academic literature investigating the present phenomenon. Based on the present review, we argue that more general empirical and experimental research would not only aid in the understanding of NoD but may further provide insights into potential interventions through the investigation of the aforementioned causal relationships. The use of an experimental vignette method (EVM) in particular could lend itself to further investigations, wherein fictitious scenarios may be manipulated to investigate the impact of specific factors on the attitudes and perceptions of participants (Aguinis & Bradley, 2014). These scenarios can be designed to replicate the working environments of individual workforces, allowing for the assessment of specific predispositions to normalization of risk/deviance. Such investigation may also be of particular importance for investigating elements of risk normalization within healthcare settings where the concept has thus far not been explored or considered in as much depth as in other high-risk industries.

Additionally, the further conceptualization of the role of Type I and II errors in the development of NoD within organizational systems appears warranted, especially with respect to their efficiency, safety, and cost/benefit interactions/trade-offs, as highlighted by Heimann (2005); in addition to their potentially varied presentations and implications within different industrial sectors (e.g., healthcare versus process industries). Understanding the priorities of an industry and its workers with regards to type I and II errors may also be supplemental to NoD investigations by illuminating where an organization stands from a "cycles of failure" perspective (Heimann, 2005). The identification of patterns of deviance in tandem with type I and II prioritization may prove to be particularly important in the recognition of otherwise overlooked system risks and may help inform on appropriate preventive measures and/or beneficial system changes.

5. Conclusion

The study of NoD is theoretically based on the systems approach to accident causation, wherein emphasis is placed on understanding how dynamic components of a system enable a given phenomenon to manifest and propagate. An important facet of this approach is its emphasis on understanding the impact of latent failures, framing active failures as by-products of a flawed system rather than vice-versa. The benefit of this perspective is that it enables the development of interventions and improvements that can be applicable and generalized across a range of contexts that accommodate similar system dynamics. The present review, which aimed to synthesize the existing literature on the phenomenon of NoD from a range of high-risk industrial sectors, may represent an initial step toward such interventions with regard to the NoD phenomenon and high-risk industry. Using a directed content analysis approach, the present systematic review of 33 articles synthesizes the existing literature and presents its findings within a conceptual framework. The framework seeks to encapsulate the reported interactions between identified industry components and NoD, while building upon prior examples through the incorporation of risk normalization. While unable to offer specific interventions, the present paper provides foundations for future applied research on the topic and offers a common framework for the phenomenon that is applicable across a range of industrial sectors.

Funding

This review was conducted as part of a PhD programme, funded by the Energy Institute. No financial or personal conflicts of interest exist.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The present review was undertaken as part of a funded PhD by student Nejc Sedlar, with funding provided by the Energy Institute through the University of Aberdeen. The funding organisation had no influence on the study design, and was not involved in the collection, analysis, or interpretation of data; in the writing of the report; and in the decision to submit the article for publication. There are no further conflicts of interest to disclose from any of the authors of the present review paper.

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods, 17*(4), 351–371. https://doi.org/10.1177/1094428114547952.

Arendt, S., & Manton, M. (2015, May 13-15). *Understanding Process Safety Culture Disease Pathologies - How to Prevent, Mitigate and Recover from Safety Culture Accidents* [Paper presentation]. Hazards 25.

Banja, J. (2010). The normalization of deviance in healthcare delivery. *Business Horizons, 53*(2), 139–148. https://doi.org/10.1016/j.bushor.2009.10.006.

Barnett, R. L. (2020). On the Safety Hierarchy and Hierarchy of Controls. *American Journal of Mechanical Engineering, 8*(2), 61–68 http://pubs.sciepub.com/ajme/8/2/3/.

Bell, T. E., & Esch, K. (1987). The fatal flaw in flight 51-L. *IEEE Spectrum, 24*(2), 36–51.

Bloch, K. P., & Williams, S. A. (2004). Normalize deviance at your peril: Do not let longtime incident-free operation justify a design or procedure that is not justifiable. *Chemical Engineering, 111*(5), 52–57.

Bogard, K., Ludwig, T. D., Staats, C., & Kretschmer, D. (2015). An industry's call to understand the contingencies involved in process safety: Normalization of deviance. *Journal of Organizational Behavior Management, 35*(1–2), 70–80. https://doi.org/10.1080/01608061.2015.1031429.

CAIB (Columbia Accident Investigation Board) (2003). *Report of Columbia Accident Investigation Board, Volume I.* National Aeronautics and Space Administration (NASA). http://s3.amazonaws.com/akamai.netstorage/anon.nasaglobal/CAIB/CAIB_lowres_full.pdf.

Carter, K. M., Mead, B. A., Stewart, G. L., Nielsen, J. D., & Solimeo, S. L. (2019). Reviewing work team design characteristics across industries: Combining meta-analysis and comprehensive synthesis. *Small Group Research, 50*(1), 138–188. https://doi.org/10.1177/1046496418797431.

Cavnor, C. D. (2018). *Fighting the fire in our own house: how poor decisions are smoldering within the US fire service* [Unpublished master's thesis]. Naval Postgraduate School.

Creedy, G. D. (2011). Quantitative risk assessment: How realistic are those frequency assumptions? *Journal of Loss Prevention in the Process Industries, 24*(3), 203–207. https://doi.org/10.1016/j.jlp.2010.08.013.

Creyer, E., Ross, W., & Evers, D. (2003). Risky recreation: An exploration of factors influencing the likelihood of participation and the effects of experience. *Leisure Studies, 22*(3), 239–253. https://doi.org/10.1080/026143603200068000.

Dechy, N., Dien, Y., Marsden, E., & Rousseau, J.-M. (2018). Learning failures as the ultimate root causes of accidents. In J. U. Hagen (Ed.), *How Could This Happen?* (pp 105–128). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-76403-0_6.

Dekker, S., & Pruchnicki, S. (2014). Drifting into failure: Theorising the dynamics of disaster incubation. *Theoretical Issues in Ergonomics Science, 15*(6), 534–544. https://doi.org/10.1080/1463922x.2013.856495.

Denis, J. L., Langley, A., & Rouleau, L. (2010). The practice of leadership in the messy world of organizations. *Leadership, 6*(1), 67–88. https://doi.org/10.1177/1742715009354233.

Dillon, R. L., Rogers, E. W., Madsen, P., & Tinsley, C. H. (2013, March 2-9). *Improving the recognition of near-miss events on NASA missions.* [Paper presentation] 2013 IEEE Aerospace Conference, Big Sky, Montana, USA. https://doi.org/10.1109/aero.2013.6496846.

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE open, 4*(1), 2158244014522633. http://doi.org/10.1177/2158244014522633.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x.

Everson, M. G., Wilbanks, B. A., & Boust, R. R. (2020). Exploring production pressure and normalization of deviance and their relationship to poor patient outcomes. *AANA Journal, 88*(5), 365–371.

Flin, R., & Yule, S. (2004). Leadership for safety: Industrial experience. *BMJ Quality & Safety, 13*, ii45-ii51. https://doi.org/10.1136/qshc.2003.009555.

Furey, M., & Rixon, D. (2018). When abnormal becomes normal: How altered perceptions contributed to the ocean ranger oil rig disaster. *Journal of Management Policy and Practice, 19*(3), 106–116. https://doi.org/10.33423/jmpp.v19i3.51.

Gaba, D. M. (2000). Structural and organizational issues in patient safety: A comparison of health care to other high-hazard industries. *California Management Review, 43*(1), 83–102. https://doi.org/10.2307/41166067.

Gambatese, J., & Hinze, J. (1999). Addressing construction worker safety in the design phase: Designing for construction worker safety. *Automation in Construction, 8*(6), 643–649. https://doi.org/10.1016/s0926-5805(98)00109-5.

Garrick, R., & Morey, R. (2015). Dialysis facility safety: Processes and opportunities. *Seminars in Dialysis, 28*(5), 514–524. https://doi.org/10.1111/sdi.12395.

Geisz-Everson, M., Jordan, L., Nicely, K. W., & McElhone, P. (2019). Cardiovascular complications in patients undergoing noncardiac surgery: A cardiac closed claims thematic analysis. *AANA Journal, 87*(2), 124–130.

Goh, Y. M., Love, P. E., Brown, H., & Spickett, J. (2012). Organizational accidents: A systemic model of production versus protection. *Journal of Management Studies, 49*(1), 52–76. https://doi.org/10.1111/j.1467-6486.2010.00959.x.

Golinski, M., & Hranchook, A. M. (2018). Adverse events during cosmetic surgery: A thematic analysis of closed claims. *AANA Journal, 86*(2), 127–136.

Hajikazemi, S., Aaltonen, K., Ahola, T., Aarseth, W., & Andersen, B. (2020). Normalising deviance in construction project organizations: A case study on the collapse of Carillion. *Construction Management and Economics, 38*(12), 1122–1138. https://doi.org/10.1080/01446193.2020.1804069.

Hase, S., & Phin, S. (2015, November 9-12). *The normalization of deviance in the oil and gas industry: The role of rig leadership in success and failure.* [Paper presentation] Abu Dhabi International Petroleum Exhibition and Conference. https://doi.org/10.2118/177777-ms.

Hedgecoe, A. (2014). A deviation from standard design? Clinical trials, research ethics committees, and the regulatory co-construction of organizational deviance. *Social Studies of Science, 44*(1), 59–81. https://doi.org/10.1038/sj.bdj.2013.1116.

Heimann, L. (2005). Repeated failures in the management of high risk Technologies. *European Management Journal, 23*(1), 105–117. https://doi.org/10.1016/j.emj.2004.12.015.

Helmreich, R. L., & Merritt, A. C. (2001). Culture at work in aviation and medicine: National, organizational and professional influences. *Routledge*. https://doi.org/10.4324/9781315258690.

Hopkins, A. (2006). What are we to make of safe behaviour programs? *Safety Science, 44*(7), 583–597. https://doi.org/10.1016/j.ssci.2006.01.001.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687.

Hudson, P. (2003). Applying the lessons of high risk industries to health care. *BMJ Quality & Safety, 12*(1), i7–i12. https://doi.org/10.1136/qhc.12.suppl_1.i7.

Hutter, B. M. (2010). *Anticipating risks and organising risk regulation*. Cambridge University Press. https://doi.org/10.1017/cbo9780511761553.

Johnson, E. J., & Tversky, A. (1984). Representations of perceptions of risks. *Journal of Experimental Psychology: General, 113*(1), 55–70. https://doi.org/10.1037/0096-3445.113.1.55.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291. https://doi.org/10.2307/1914185.

Kelloway, E. K., Nielsen, K., & Dimoff, J. K. (Eds.). (2017). *Leading to occupational health and safety: How leadership behaviours impact organizational safety and well-being*. John Wiley & Sons.

King, C. A. (2010). To err is human, to drift is normalization of deviance. *AORN Journal, 2*(91), 284–286. https://doi.org/10.1016/j.aorn.2009.10.020.

Kurt, R. E., Arslan, V., Comrie, E., Khalid, H., & Turan, O. (2016, November 28-30). *SEAHORSE procedure improvement system*. 6th Conference on Design for Safety Brooktorkai, Hamburg, Germany. https://strathprints.strath.ac.uk/58983/.

Luís, S., Pinho, L., Lima, M. L., Roseta-Palma, C., Martins, F. C., & Betâmio de Almeida, A. (2015). Is it all about awareness? The normalization of coastal risk. *Journal of Risk Research, 19*(6), 810–826. https://doi.org/10.1080/13669877.2015.1042507.

Marsden, P., & Green, M. (1996). Optimising procedures in manufacturing systems. *International Journal of Industrial Ergonomics, 17*(1), 43–51. https://doi.org/10.1016/0169-8141(94)00102-2.

Mast, B. L. (2018, August 6-7). *Summary of the King County, Washington, West Point WWTP Flood of 2017* [Paper presentation]. ISA Water/Wastewater and Automatic Controls Symposium.

McNamara, S. A. (2011). The normalization of deviance: What are the perioperative risks? *AORN Journal, 93*(6), 796–801. https://doi.org/10.1016/j.aorn.2011.02.009.

Mize, J. F. (2019). The roundabout way to disaster: Recognizing and responding to normalization of deviance. *Process Safety Progress, 38*(2), e12014.

Mohammadi, A., & Tavakolan, M. (2019). Modeling the effects of production pressure on safety performance in construction projects using system dynamics. *Journal of Safety Research, 71*, 273–284. https://doi.org/10.1016/j.jsr.2019.10.004.

Morris, G. A., & Cannady, R. (2019). Proper use of the hierarchy of controls. *Professional Safety, 64*(08), 37–40 https://onepetro.org/PS/article/64/08/37/33643/Proper-Use-of-the-Hierarchy-of-Controls.

NASA (National Aeronautics and Space Administration). (1986). *Report to the Presidential Commission on the Space Shuttle Challenger Accident.* https://sma.nasa.gov/SignificantIncidents/assets/rogers_commission_report.pdf.

Naweed, A., Rainbird, S., & Dance, C. (2015). Are you fit to continue? Approaching rail systems thinking at the cusp of safety and the apex of performance. *Safety Science, 76*, 101–110. https://doi.org/10.1016/j.ssci.2015.02.016.

Naweed, A., & Rose, J. (2015). "It's a frightful scenario": A study of tram collisions on a mixed-traffic environment in an Australian metropolitan setting. *Procedia Manufacturing, 3*, 2706–2713. https://doi.org/10.1016/j.promfg.2015.07.666.

Odom-Forren, J. (2011). The normalization of deviance: A threat to patient safety. *Journal of PeriAnesthesia Nursing, 26*(3), 216–219. https://doi.org/10.1016/j.jopan.2011.05.002.

Paletz, S. B., Bearman, C., Orasanu, J., & Holbrook, J. (2009). Socializing the human factors analysis and classification system: Incorporating social psychological phenomena into a human factors error classification system. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 51*(4), 435–445. https://doi.org/10.1177/0018720809343588.

Pannick, S., Archer, S., Johnston, M. J., Beveridge, I., Long, S. J., Athanasiou, T., & Sevdalis, N. (2017). Translating concerns into action: A detailed qualitative evaluation of an interdisciplinary intervention on medical wards. *BMJ Open, 7*(4), e014401.

Park, J., & Jung, W. (2003). The operators' non-compliance behavior to conduct emergency operating procedures—comparing with the work experience and the complexity of procedural steps. *Reliability Engineering & System Safety, 82*(2), 115–131. https://doi.org/10.1016/s0951-8320(03)00123-6.

Perrow, C. (1999). *Normal accidents: Living with high-risk technologies*. Princeton University Press. https://doi.org/10.1515/9781400828494.

Pilbeam, C., Doherty, N., Davidson, R., & Denyer, D. (2016). Safety leadership practices for organizational safety compliance: Developing a research agenda from a review of the literature. *Safety Science, 86*, 110–121. https://doi.org/10.1016/j.ssci.2016.02.015.

Price, M. R., & Williams, T. C. (2018). When doing wrong feels so right: Normalization of deviance. *Journal of Patient Safety, 14*(1), 1–2. https://doi.org/10.1097/pts.0000000000000157.

Prielipp, R. C., Magro, M., Morell, R. C., & Brull, S. J. (2010). The normalization of deviance: Do we (un) knowingly accept doing the wrong thing? *Anesthesia & Analgesia, 110*(5), 1499–1502. https://doi.org/10.1213/ane.0b013e3181d5adc5.

Probst, T. M., & Graso, M. (2013). Pressure to produce=pressure to reduce accident reporting? *Accident Analysis & Prevention, 59*, 580–587. https://doi.org/10.1016/j.aap.2013.07.020.

Quinn, R. E. (2018). When "SOP" fails: Disseminating risk assessment in aviation case studies and analysis. *The Collegiate Aviation Review International, 35*(2). https://doi.org/10.22488/okstate.18.100479.

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science, 27*(2–3), 183–213. https://doi.org/10.1016/s0925-7535(97)00052-0.

Reason, J. (1990). The contribution of latent human failures to the breakdown of complex systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 327*(1241), 475–484. https://doi.org/10.1098/rstb.1990.0090.

Reason, J. (2000). Human error: models and management. *BMJ (Clinical research ed.), 320*(7237), 768–770. https://doi.org/10.1136/bmj.320.7237.768.

Reuter, E., & Camba, J. D. (2017). Understanding emergency workers' behavior and perspectives on design and safety in the workplace. *Applied Ergonomics, 59*, 73–83. https://doi.org/10.1016/j.apergo.2016.08.023.

Ruault, J. R., Vanderhaegen, F., & Kolski, C. (2013). Sociotechnical systems resilience: A dissonance engineering point of view. *IFAC Proceedings Volumes, 46*(15), 149–156. https://doi.org/10.3182/20130811-5-us-2037.00042.

Sanne, J. M. (2012). Learning from adverse events in the nuclear power industry: Organizational learning, policy making and normalization. *Technology in Society, 34*(3), 239–250. https://doi.org/10.1016/j.techsoc.2012.07.002.

Schweitzer, J., & Mix, T. L. (2018). Risk normalization in a nuclear reliant society: The construction of risk in French media narratives pre-and post-Fukushima. *Environmental Sociology, 4*(2), 264–274. https://doi.org/10.1080/23251042.2017.1393856.

Scott, I. A., Soon, J., Elshaug, A. G., & Lindner, R. (2017). Countering cognitive biases in minimising low value care. *Medical Journal of Australia, 206*(9), 407–411. https://doi.org/10.5694/mja16.00999.

Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology, 70*, 747–770. https://doi.org/10.1146/annurev-psych-010418-102803.

Simmons, D., Symes, L., Guenter, P., & Graves, K. (2011). Tubing misconnections: Normalization of deviance. *Nutrition in Clinical Practice, 26*(3), 286–293. https://doi.org/10.1177/0884533611406134.

Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the determinants of risk behavior. *Academy of Management Review, 17*(1), 9–38. https://doi.org/10.5465/amr.1992.4279564.

Starbuck, W. H., & Milliken, F. J. (1988). Challenger: Fine-tuning the odds until something breaks. *Journal of Management Studies, 25*(4), 319–340. https://doi.org/10.1111/j.1467-6486.1988.tb00040.x.

Stave, C., & Törner, M. (2007). Exploring the organizational preconditions for occupational accidents in food industry: A qualitative approach. *Safety Science, 45*(3), 355–371. https://doi.org/10.1016/j.ssci.2006.07.001.

Stergiou-Kita, M., Mansfield, E., Bezo, R., Colantonio, A., Garritano, E., Lafrance, M., ... Travers, K. (2015). Danger zone: Men, masculinity and occupational health and safety in high risk occupations. *Safety Science, 80*, 213–220. https://doi.org/10.1016/j.ssci.2015.07.029.

Tversky, A., & Kahneman, D. (1985). The Framing of Decisions and the Psychology of Choice. In G. Wright (Ed.), *Behavioral Decision Making* (pp. 25–41). Springer. https://doi.org/10.1007/978-1-4613-2391-4_2.

Turner, B. A., & Pidgeon, N. F. (1997). *Man-made disasters* ((2nd ed.).). Butterworth-Heinemann.

Van den Berg, P. T., & Wilderom, C. P. (2004). Defining, measuring, and comparing organizational cultures. *Applied Psychology, 53*(4), 570–582. https://doi.org/10.1111/j.1464-0597.2004.00189.x.

Vaughan, D. (1996). *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. The University of Chicago Press.

Vogus, T. J., & Sutcliffe, K. M. (2007). Organizational resilience: Towards a theory and research agenda. In *2007 IEEE International Conference on Systems, Man and Cybernetics* (pp. 3418–3422). https://doi.org/10.1109/icsmc.2007.4414160.

Wang, J., & Ruxton, T. (1997). Design for safety. *Professional Safety, 42*(1), 24–29 https://www.proquest.com/scholarly-journals/design-safety/docview/200423293/se-2?accountid=8155.

Weber, R. P. (1990). *Basic content analysis*. Sage. https://doi.org/10.4135/9781412983488.

Wilbanks, B. A., Geisz-Everson, M., Clayton, B. A., & Boust, R. R. (2018). Transfer of care in perioperative settings: A descriptive qualitative study. *AANA Journal, 86*(5), 401–407.

Willcoxson, L., & Millett, B. (2000). The management of organizational culture. *Australian Journal of Management and Organizational Behaviour, 3*(2), 91–99.

Nejc Sedlar: Nejc Sedlar is a Psychology PhD student at the University of Aberdeen who is currently working on a project funded by the Energy Institute investigating the impact of cognitive biases and risk normalization on safety within the energy industry. He has been involved in human factors research since the undergraduate level and has worked on a number of projects investigating the role of human factors and non-technical skills within general aviation, offshore/search and rescue helicopters, and farming, helping in the development of both the HeliNOTS and FLINTS behavioural marker systems.

Dr Amy Irwin: Dr Irwin is a senior lecturer within the School of Psychology, University of Aberdeen. She attained her PhD in 2008 at the University of Nottingham, and shortly after moved to the University of Aberdeen to conduct applied research. She now leads the Applied Psychology and Human Factors research group and conducts applied research across a variety of domains including agriculture, aviation and healthcare.

Dr Douglas Martin: Doug is an experimental social psychologist who is primarily interested in social cognition. Doug received both an MA (2000) and PhD (2005) in Psychology from the University of Aberdeen. Much of Doug's research addresses two broad themes: 1. How people extract social category information from faces; 2. How stereotypes form, are maintained, and change via cumulative cultural evolution.

Dr Ruby Roberts: Dr Ruby Roberts is Research Fellow of Industrial Psychology at Aberdeen Business School, Robert Gordon University. She specialises in human factors and safety in high hazard environments, working in a range of industries including oil and gas, and healthcare. More recently she has been working to understand decisions around innovation and corporate technology adoption.

Journal of Safety Research 84 (2023) 204–211

Contents lists available at ScienceDirect

Journal of Safety Research

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

A review of occupational safety and health research for American Indians and Alaska Natives

Kaitlin Wingate, Elizabeth Dalsey *, Deborah Poling Scott

National Institute for Occupational Safety and Health, Western States Division, United States

ARTICLE INFO

Article history:
Received 7 February 2022
Received in revised form 12 July 2022
Accepted 25 October 2022
Available online 3 November 2022

Keywords:
AI
AN
Occupational safety and health
Worker safety

ABSTRACT

Introduction: To better understand what is known about issues affecting American Indian and Alaska Native (AI/AN) workers, authors conducted a literature review of publications specific to AI/AN and occupational safety and health. *Methods:* Search criteria included: (a) American Indian tribes and Alaska Native villages in the United States; (b) First Nations and aboriginals in Canada; and (c) occupational safety and health. *Results:* Results of two identical searches in 2017 and 2019 identified 119 articles and 26 articles respectively, with references to AI/AN people and occupation. Of the 145 total articles, only 11 articles met the search criteria for addressing occupational safety and health research among AI/AN workers. Information from each article was abstracted and categorized according to National Occupational Research Agenda (NORA) sector, resulting in: four articles related to agriculture, forestry, and fishing; three related to mining; one related to manufacturing; and one related to services. Two articles reported on AI/AN people and occupational well-being in general. *Conclusions:* The review was limited by the small number and age of relevant articles, reflecting the likelihood that findings could be out of date. General themes across the reviewed articles point to the need for increased overall awareness and education regarding injury prevention and risks associated with occupational injuries and fatalities among AI/AN workers. Similarly, increased use of personal protective equipment (PPE) is recommended for the agriculture, forestry, and fishing industries, as well as for workers exposed to metals dust. *Practical Applications:* The lack of research in most NORA sectors indicates the need for heightened research efforts directed toward AI/AN workers.

1. Introduction

American Indian and Alaska Native (AI/AN) people accounted for 2% (5.8 million) of the total U.S. population in 2020 (U.S. Census Bureau, 2020). Little is known about the occupational safety and health of AI/AN workers. AI/AN people face a disproportionate burden of illness as well as a lower life expectancy than the total U.S. population. The leading causes of death for AI/AN people include heart disease, cancer, unintentional injuries, and diabetes. AI/AN people experience higher rates of these and other causes of death compared to the total U.S. population (Indian Health Service, 2019).

As sovereign nations, AI/AN tribes maintain a government-to-government relationship with the U.S. federal government (§ 4.01[1][a], 2017).Of the 5.8 million AI/AN, 2.7 million individuals report AI/AN as their only race, of which, over half (1,548,549) are active in the workforce (Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 2022; U.S. Census Bureau, 2020). Approximately 28% of people who identify as AI/AN lived on federal or state reservations or trust lands from 2016 to 2018. Of those, 52.0% were active in the labor force, compared to 63.6% of AI/AN people who did not live in AI/AN areas (Bureau of Labor Statistics, 2019). Table 1 describes AI/AN employment in the U.S. by National Occupational Research Agenda (NORA) sector, based upon the North American Industry Classification System (NAICS). These sectors include: agriculture, forestry, and fishing; construction; healthcare and social assistance; manufacturing; mining; oil and gas extraction; public safety; services; transportation, warehousing, and utilities; and wholesale and retail trade (National Occupational Research Agenda, 2018). The services sector has the highest number of AI/AN workers, followed by wholesale and retail trade and healthcare and social assistance.

Detailed data on AI/AN workers is scarce for several reasons. Race and ethnicity are not commonly collected in traditional occupational safety and health data sources. For example, the Occupa-

* Corresponding author at: Western States Division, P.O. Box 25226, Denver, CO 80225, United States.

E-mail address: edalsey@cdc.gov (E. Dalsey).

https://doi.org/10.1016/j.jsr.2022.10.020
0022-4375/

Table 1
American Indian and Alaska native employment by national occupational research agenda sector, 2020.

National Occupational Research Agenda (NORA) Sectors	AI/AN People in Workforce (n = 1,548,549)
Agriculture, Forestry, and Fishing	28,877
Construction	144,847
Healthcare and Social Assistance	211,576
Manufacturing	132,012
Mining (includes Oil and Gas Extraction)	16,408
Services (includes Public Safety)	695,309
Transportation, Warehousing, and Utilities	82,896
Wholesale and Retail Trade	236,624

Note: The Employed Labor Force Query System is based on data from the Bureau of Labor Statistics Current Population Survey, which uses Bureau of Census Industry Codes. These codes can not be directly matched with the North American Industry Classification System, which organizes the current NORA sectors. As a result, the table includes the population for the NORA sector, Oil and Gas Extraction, in the Mining category. Similarly, it includes the population for the Public Safety sector in the Services category.
Source: Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health. (2020). Retrieved from Employed Labor Force Query System: https://wwwn.cdc.gov/wisards/cps/default.aspx.

tional Safety and Health Administration's (OSHA) injury and illness recording requirements do not collect information on race or ethnicity (National Academies of Sciences, Engineering, and Medicine, 2018). Sources that do collect race and ethnicity, such as the Survey of Occupational Injuries and Illnesses, often do not require a response to these questions (Bureau of Labor Statistics, U.S. Department of Labor, 2021). Therefore, race and ethnicity data are frequently missing. Additionally, AI/AN workers may choose to not report or self-identify their race and ethnicity. Furthermore, no tribal surveillance systems or data sources collect this information.

AI/AN occupational fatalities are likely underestimated, as they are often not reported to OSHA when they occur on tribal lands and reservations (Hill, Reyes, & Dalsey, 2013). Despite the existence of tribal occupational safety and health programs and codes and laws, the scientific literature on AI/AN occupational health and safety is limited (Center for State, Tribal, Local, and Territorial Support, Centers for Disease Control and Prevention, 2017). An additional challenge when working with AI/AN data is that sources do not always specify whether AI/AN is reported as an individual's only race or a race in combination with another. The purpose of this review is to summarize what is documented in the peer-reviewed literature about AI/AN worker safety and health.

AI/AN workers are often employed in hazardous occupations and were 42% more likely to be employed in high risk occupations than non-Hispanic White workers in 2010 (Steege, Baron, Marsh, Menendez, & Myers, 2014). High risk occupations are those whose "days away from work" illness or injury rates are more than twice the national average, such as construction workers and miners (Council of State and Territorial Epidemiologists, 2017). In 2019, 24% of AI/AN employees worked in occupations classified as having high morbidity risk, while only 16% of all U.S. employees worked in these occupations (Council of State and Territorial Epidemiologists, 2017; Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 2022). In 2020, 5,340 AI/AN workers suffered a nonfatal injury resulting in one or more days away from work, and 32 AI/AN workers were killed on the job (Bureau of Labor Statistics, U.S. Department of Labor, 2020; Bureau of Labor Statistics, U.S. Department of Labor, 2020). The occupational fatality rate for all U.S. workers in 2020 was 3.5 deaths per 100,000 workers, compared to 3.6 deaths per 100,000 AI/AN workers (Bureau of Labor Statistics, U.S. Department of Labor, 2020; Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 2022). It appears that AI/AN workers may face a high risk of occupational injury, and the nature of these risks may differ from that of workers of other races and ethnicities. AI/AN workers are an underserved worker population, and research on AI/AN people and occupational safety and health is limited.

2. Methods

Authors submitted a literature search request in 2017 and again in 2019 to the Centers for Disease Control and Prevention's Public Health Library and Information Center Reference Team. Both searches included online databases MedLine, Embase, PsycInfo, CINAHL, and Scopus. To identify all possible articles, no date limits were placed on the search activities. Search criteria included: (a) American Indian tribes and Alaska Native villages in the United States; (b) First Nations/Aboriginal groups located in Canada; and (c) occupational health and safety. Search criteria excluded articles on other indigenous groups in Australia, Mexico, Central and South America, and migrant workers. The review also excluded articles related to environmental exposures, substance abuse, and child or civilian seatbelt safety and traffic laws, unless related to work activities. Search terms included: American Indians, Alaska Natives, Inuit, Native American, Native Alaskan, Tribal Nation, First Nations, Indian reservation, occupational safety, occupational health, occupational injuries, occupational fatalities, occupational health services, occupational medicine, occupational illnesses, workers' compensation, epidemiology, population surveillance, participatory research, chemical exposures, hazardous chemicals, ergonomics, psychosocial, falls, accidents, construction, agriculture, farming, mining, fishing, motor vehicle, collisions, asthma, noise, National Institute for Occupational Safety and Health, and violence.

Results of the initial search in 2017 identified 119 articles, reports, and theses. An identical search limited to articles published since the initial search was requested in 2019 and identified 26 additional articles. Further inclusion criteria was used to focus on peer-reviewed journals, published after 2000, addressing AI/AN occupational safety and health. Fig. 1 illustrates the process used to select relevant articles for the literature review. Authors abstracted information from each relevant article and categorized them according to NORA sector. Although many articles pertained to AI/AN people, most documented research unrelated to occupational safety and health.

3. Results

This review comprises 11 articles categorized by NORA sector including: four articles categorized into the agriculture, forestry, and fishing sector; one to manufacturing; three to mining; and one article to services. Two of the 11 articles did not fit within one specific sector but described research relating to the overall well-being of AI/AN workers. Authors categorized the article related to rodeo competitors to the agriculture, forestry, and fishing sector, even though the specific rodeo NAICS code falls under the Arts, Entertainment, and Recreation sector. Although Canada was included in the original search criteria, only articles about AI/AN workers in the United States met the inclusion criteria. The findings and themes presented in these articles are discussed by NORA sector below and referenced in Table 2.

3.1. Agriculture, forestry and, fishing

Four of the 11 articles found in the literature search were related to agriculture, forestry, and fishing. Of these four, two dis-

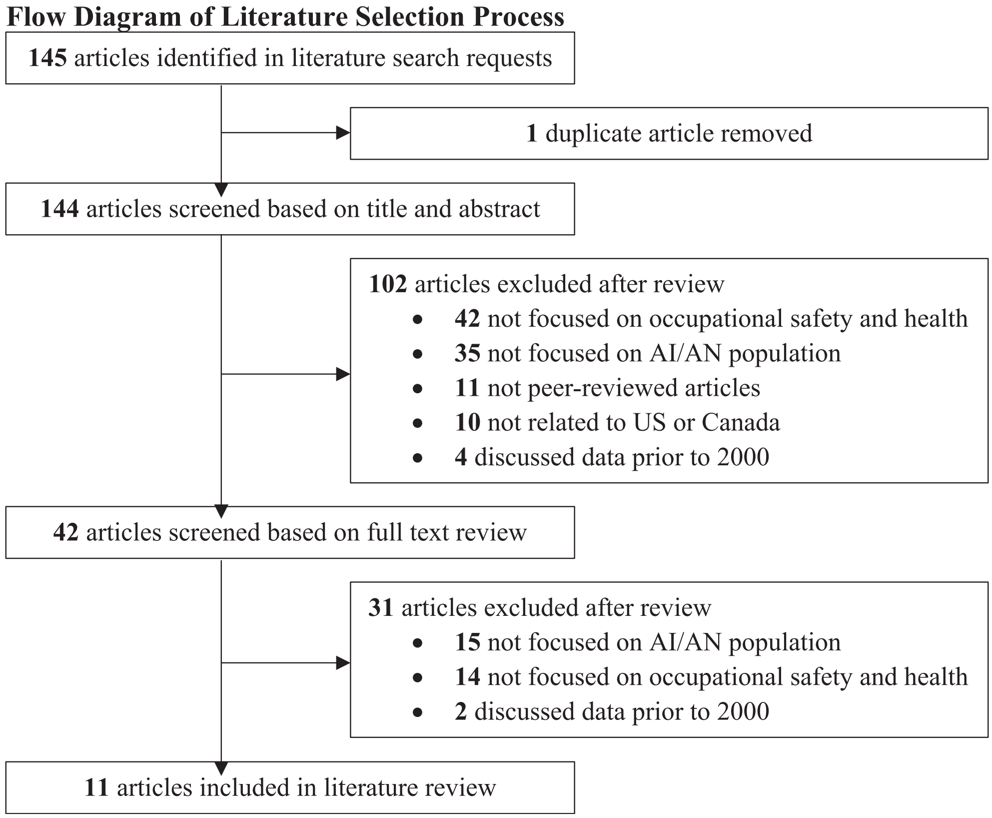


Fig. 1. Flow diagram of literature selection process.

cussed farmworkers, one investigated professional rodeo competitors, and one focused on bison workers.

3.1.1. Objectives and methods

Among the four articles, three included discussion on work-related injuries among farmworkers and rodeo competitors. Goldcamp, Hendricks, Layne, and Myers analyzed previously collected data from the Minority Farm Operator Childhood Agricultural Injury Survey (M-CAIS), conducted in 2000, to examine nonfatal injuries among youth farmworkers (Goldcamp, Hendricks, Layne, & Myers, 2006). Crichlow, Williamson, Geurin, and Heggem administered a self-reported survey to 180 professional rodeo competitors to evaluate injuries and use of protective equipment (Crichlow, Williamson, Geurin, & Heggem, 2006). Lastly, Duysen et al. conducted observational audits and analyzed convenience surveys completed by herd managers in order to investigate the hazards of bison handling (Duysen, et al., 2017).

In the fourth article related to agriculture, forestry, and fishing, Helitzer, Hathorn, Benally, and Ortega conducted a five-year agricultural intervention based on a theoretical framework, the *Diffusion of Innovations Theory,* to develop a culturally relevant model to train American Indian farmers in New Mexico on the safe use of pesticides. The intervention trained six "model farmers" on safe pesticide use, and later implemented an intervention and delayed intervention across 830 Navajo family farms (Helitzer, Hathorn, Benally, & Ortega, 2014).

3.1.2. Findings

The three articles discussing work-related injuries first identified estimates for injuries at various work sites. The nonfatal injury rate for work-related injuries on Native American farms was 17.7 per 1,000 household youth, compared to a non-work injury rate of 13.8 per 1,000 household youth. This rate was even higher for youth under 10 years of age at 21.8 injuries per 1,000 working youth. Lacerations and fractures, generally to the arm and leg, were the most common work-related injuries reported among household youth working on Native American-operated farms (Goldcamp, Hendricks, Layne, & Myers, 2006). Among rodeo competitors responding to a self-reported survey, the most common injuries were among the lower extremities, rather than the body core. The self-reported injury rate among competitors was 14 injuries per 100 rodeos. Injury history varied among rodeo events, ranging from 24% of tie-down ropers to 100% of bull riders. Additionally, 26% of the competitors had a history of rodeo injury that prevented them from working, leading to an average of 3.2 months away from competition (Crichlow, Williamson, Geurin, & Heggem, 2006). Injuries were also commonly observed at bison handling sites, which may be due to the high-stress handling that exists during bison roundups. These injuries were reportedly caused by equipment, tools, weather, bison, and ATV use. Additional safety concerns found through the observational audits include obsolete or broken equipment, poor facility designs, and inadequate barriers in the chutes (Duysen, et al., 2017).

A common theme among these articles was the lack of personal protective equipment (PPE) among workers in this sector. Among rodeo competitors, although past injuries were common, only 40% of athletes reported using PPE. The most widely used protective equipment were vests (Crichlow, Williamson, Geurin, & Heggem, 2006). Additionally, research has shown that safety equipement is rarely used at AI/AN bison worksites as well as at agriculture sites in general. Specifically, inappropriate footwear was reported by 27% of herd managers. PPE use was reportedly very low and deficiencies in the equipment such as dust masks and safety goggles were also reported (Duysen, et al., 2017). Authors recommended increased safety training, such as proper PPE, including targeted interventions related to the demographics of

Table 2
Summary of articles by NORA sector.

Sector	Author(s) and Title	Objective	Study Type	Findings
Agriculture, Forestry, and Fishing	**Goldcamp, E. M., et al. (2006)** Nonfatal injuries to household youth on Native American operated farms in the U.S., 2000.	Identify characteristics of work and non-work-related farming injuries to Native American youth	Cross-sectional survey; self-administered by 9,556 racial minority-operated farms	There were an estimated 177 nonfatal injuries of youth living on Native American operated farms in 2000. The injury rate was 17.7 per 1,000 Native American working youth, compared to 13.8 per 1,000 non-working youth.
Agriculture, Forestry, and Fishing	**Crichlow, R. et al. (2006)** Self-reported injury history in Native American professional rodeo competitors.	Evaluate rodeo injury and use of protective equipment.	Cross-sectional survey; self-administered to 180 competitors	14 injuries per 100 rodeos were reported, ranging from 100% of bull riders to 24% of tie-down ropers. Only 40% reported use of protective equipment.
Agriculture, Forestry, and Fishing	**Duysen, E., et al. (2017)** Assessment of tribal bison worker hazards using trusted research facilitators.	Identify hazards of bison handling in American Indians	Cross-sectional survey; collaborative pilot research project with observational audits and convenience surveys	A lack of safety equipment was observed at worksites. Bison injuries occurred at 9 out of 10 sites, and worker injuries occurred at 3 out of 10 sites.
Agriculture, Forestry, and Fishing	**Helitzer, D. L., et al. (2014)** Culturally relevant model program to prevent and reduce agricultural injuries.	Describe a pesticide injury prevention program among AI farmers based on Diffusion of Innovations theory	Longitudinal study of injury prevention intervention using train the trainer for two groups: intervention and delayed intervention	There was an increase in pesticides stored out of reach of children ($G = 15.5$, $p < 0.001$; $G = 7.7$, $p < 0.05$)[1] as well as owning safety equipment ($G = 64.8$, $p < 0.001$; $G = 12.5$, $p < 0.005$)[1] and knowledge of safe pesticide application ($t = 5.479$, $p < 0.001$; $t = 8.559$, $p < 0.05$)[2].
Manufacturing	**Gonzales, M. et al. (2004)** Concentrations of surface-dust metals in Native American jewelry-making homes in Zuni Pueblo, New Mexico.	Identify and quantify metals used by home-based AI jewelers	Cross-sectional exposure survey; in-person interviews with 40 participants	Metal dust concentrations were significantly higher in jewelers' homes compared to homes of non-jewelers ($p < 0.02$).
Mining	**Schubauer-Berigan, M. K., Daniels, R. D., and Pinkerton, L. E. (2009)** Radon exposure and mortality among white and American Indian uranium miners: an update of the Colorado Plateau cohort.	Follow-up study of 4,137 uranium miners on the U.S. Colorado Plateau	Prospective cohort study; analysis of linked databases	Silicosis, tuberculosis and other lung diseases remained highly elevated among American Indian miners compared to white miners.
Mining	**Jones, B. (2017)** The social costs of uranium mining in the US Colorado Plateau cohort, 1960–2005.	Estimated health costs associated with mining in Colorado Plateau	Cost analysis of prospective cohort study	Over $2 billion in health costs over 1960–2005 existed due to uranium mining. Native Americans had larger costs per elevated death.
Mining	**Mulloy, K. B., et al. (2001)** Lung cancer in a nonsmoking underground uranium miner.	Illustrate the effects of increased risk of lung cancer among uranium miners	Case report of 72-year-old Navajo male who worked as uranium miner for 17 years	The uranium miner's radon exposure was estimated at 506 months of work with no other exposures. His risk of lung cancer was 100 times greater than if he never mined. He was treated for pneumonia and died from respiratory failure.
Service	**Klepeis, N. E., et al. (2016)** Measuring indoor air quality and engaging California Indian stakeholders at the Win-River Resort and Casino: collaborative smoke-free policy development.	Describe efforts to institute smoke-free policies in AI casino	Formative evaluation using air testing, surveys, and focus groups	Increased exposures to secondhand smoke existed in the casino. 100% Smoke-free policies were implemented in 2014 and amended to permit smoking on 30% of the casino floor in 2015.
Other	**Redwood, D., et al. (2012)** Occupational and environmental exposures among Alaska Native and American Indian people living in Alaska and the Southwest United States.	Report on prevalence of self-reported exposure to 9 occupational and environmental hazards in a large cohort of AI/AN living in AK and Southwest U.S.	Prospective cohort study; self- and interviewer- administered questionnaire to 11,326	28% of participants reported exposure to 1 to 2 hazards and 8% were exposed to 3 or more. Exposures were higher for men, those ages 40–59, and those living in the Southwest ($p < 0.05$).
Other	**Christiansen, K., et al. (2017)** Work, worksites, and wellbeing among North American Indian women: a qualitative study.	Identify factors contributing to work-family balance and health behaviors among American Indian women	Cross-sectional study; interviews and focus groups with 89 women in 4 tribal communities	Shift and seasonal work make healthy lifestyles difficult for AI/AN women. Men have to travel farther for work, leaving women to take care of the home.

[1] G = test statistics from log-likelihood goodness-of-fit test.
[2] t = test statistic from paired-sample t-test.

the worker population (Goldcamp, Hendricks, Layne, & Myers, 2006).

The fourth article analyzing a five-year agricultural intervention identified significant improvements in safe pesticide use, storage behaviors, and safety and pesticide application ownership. These knowledge improvements were maintained over time by the study group who received the immediate intervention, while the group who received the delayed intervention demonstrated a greater improvement in attitudes about pest management. Including the agricultural workers on the training development as well as a face-to-face training method and a culturally appropriate foundation led to success in this intervention (Helitzer, Hathorn, Benally, & Ortega, 2014).

3.2. Manufacturing

3.2.1. Objective and methods

Only one article related to manufacturing was identified. Gonzales et al. conducted a pilot study to inventory the materials used by Native American home-based jewelry makers in western New Mexico. Researchers compared surface concentrations of metals between 20 jewelry-making households and 20 control households where jewelry was not made (Gonzales, et al., 2004).

3.2.2. Findings

Within the jewelers' homes, metal dust concentrations were significantly higher in work areas than in living areas. Concentrations were also higher in living areas of jewelers' homes compared with the homes of non-jewelers. Additionally, use of ventilation varied depending on the metal. When ventilation was present in jewelry work areas, metal concentrations for a few metals were significantly reduced compared to concentrations where no ventilation was present (Gonzales, et al., 2004).

Gonzales et al. also reported that PPE was not commonly used among jewelry makers. For example, less than a quarter of participants reported using safety glasses, dust masks, or gloves. Even fewer jewelry makers reported using coveralls, and no participants reported ever using a face shield. Similarly, mechanical ventilation such as exhaust fans were used by less than half of jewelry makers. While natural ventilation like open doors or windows were used more frequently, many jewelry makers used no ventilation at all (Gonzales, et al., 2004).

3.3. Mining

3.3.1. Objectives and methods

Three articles highlighting mining were identified in the literature searches. Two of these articles discussed the Colorado Plateau cohort, a group of miners who were originally evaluated from 1950 through 1960 to investigate the association between radon and lung cancer as well as the risks of other diseases among miners (Schubauer-Berigan, Daniels, & Pinkerton, 2009). Schubauer-Berigan et al. used mortality data gathered from 1990 through 2005 that was linked to the National Death Index and the Social Security Administration's mortality file databases as well as the Renal Management Information System database. Jones also used these data, from 1960 through 2005, to conduct an economic cost analysis in order to understand the health costs associated with uranium mining. The value of a statistical life-year was calculated as $213,000 per year of life lost (Jones, 2017). The third article presented a case study of a 72-year-old Navajo male who spent 17 years working as an underground uranium miner (Mulloy, James, Mohs, & Kornfeld, 2001).

3.3.2. Findings

The Colorado Plateau cohort included 3,358 white uranium miners and 779 miners who were American Indian. Forty-five percent of the American Indian miners were current or former smokers, compared to 84% of white miners. Among all miners, lung cancer accounted for one in five deaths. The mortality data revealed that during the 1990 to 2005 time period, American Indian miners had a lung cancer standardized mortality ratio of 3.27 compared with the regional population; white miners had a ratio of 3.99 (Schubauer-Berigan, Daniels, & Pinkerton, 2009).

In Jones' analysis of the Colorado Plateau cohort, the median years of life lost was 13.3 for white uranium miners in the Colorado Plateau cohort and 13.9 for American Indian miners. Jones reported that $1.24 billion in health costs was created due to lung cancer mortality, and $127.9 million of the cost was for American Indian miners. The cost analysis further revealed that lung cancer accounted for 60% of total health costs for American Indian miners. Overall, $213.7 million of total mortality health costs were associated with American Indians. Additionally, a 6.9% larger excess death mortality health cost existed for American Indian miners compared to white miners. Jones argues that American Indian miners faced a disproportionate share of the social costs of mining in the Colorado Plateau, compared to white miners (Jones, 2017).

The third article presented a case study of a Navajo uranium miner and estimated his exposure to radon progeny at 506 working level months. The miner did not have any other significant occupational or environmental exposures, such as smoking. He developed lung cancer 22 years after leaving mining and died from pneumonia and respiratory failure. Unfortunately, lung function prediction equations, which use standards of physical characteristics such as age, sex, and height to determine a predicted lung function value, often lack specific variables for ethnicities or races. This can produce bias against Hispanic and American Indian miners. Additional difficulties include the lack of diagnostic resources for disease recognition and distance for primary care for American Indians in the Navajo Nation (Mulloy, James, Mohs, & Kornfeld, 2001).

3.4. Services

3.4.1. Objective and methods

One article about workers in the services sector was identified. Klepeis (2016) reported on efforts of a coalition of public health professionals working with casino management to conduct on-site studies of secondhand smoke (SHS) in a casino on a tribal reservation. The goal was to help provide guidance for future efforts of adopting smoke-free policies in casinos. During a seven-year period, the coalition members conducted air quality testing, collected surveys from casino employees and patrons, and held staff and community focus groups (Klepeis, et al., 2016).

3.4.2. Findings

Air quality evaluations revealed a range of only 8 to 12% of active smokers amongst all patrons throughout the casino; however, despite the small percentage of smoking patrons, elevated levels of urinary cotinine and airborne nicotine confirmed evidence of high levels of SHS. Survey responses indicated that over half of patrons would visit the casino about the same or more often if a smoke-free policy were enacted. Additionally, most of the casino employees preferred to work in a smoke-free environment. Based on these findings, Klepeis et al. reported that a 100% smoke-free policy be implemented in the casino. After the smoking ban was in effect, particle levels dropped by 98% in main smoking areas throughout the casino and by 51% in previously designated non-smoking areas. After a reduction in revenue and complaints from smoking patrons, the smoke-free policy was amended to restrict smoking on 70% of the casino floor (Klepeis, et al., 2016).

3.5. Other

3.5.1. Objectives and methods

Two articles were identified through the literature searches that did not focus on a specific industry sector, but rather on AI/AN worker exposures and overall well-being. Redwood et al. analyzed previously collected data from over 11,000 participants in the Education and Research Towards Health (EARTH) Study. They investigated self-reported occupational and environmental hazard exposures among participants in Alaska, of which 95% were Alaska Native, and Navajo participants living in the Southwest United States. Nine environmental hazards of concern were identified by tribal leadership and a study advisory board: petroleum, pesticides, welding/silversmithing, asbestos, military chemicals, mining dust, heavy metals, lead, and radioactive materials. Study participants were asked to indicate possible exposures to these hazards (Redwood, et al., 2012). Christiansen et al. conducted in-person interviews and focus groups with 89 females living in tribal communities in the Southwest and Upper Midwest to investigate work-related themes including structural characteristics, role stressors, and the influence of social support (Christiansen, Gadhoke, Pardilla, & Gittelsohn, 2019).

3.5.2. Findings

Among Navajo workers, 64% of all study participants reported no hazardous exposure to the nine environmental hazards of concern, while 28% reported exposure to one to two hazards, and 8% reported exposure to three or more hazards. Among AI/AN participants living in Alaska, the top three most commonly reported hazards were petroleum products, military chemicals, and asbestos. Among Navajo participants living in the Southwest United States, the top three most commonly reported hazards were pesticides, petroleum, and welding/silversmithing. Among all study participants, reported exposures were higher among male participants, participants aged 40 through 59, and individuals living in the Southwest compared to Alaska. Additionally, Redwood et al. identified a higher likelihood of reported hazard exposure for participants who spoke an AI/AN language at home compared to participants who spoke only English at home, as well as for participants with lower educational attainment (Redwood, et al., 2012).

For females living in tribal communities, Christiansen et al. identified common issues related to structural characteristics, including unemployment, seasonal employment, and low-wage work opportunities. The proximity to worksites as well as social and medical services was also commonly discussed among females as males were more likely to have to travel long distances for work. As a result, females experienced the burden of being sole caregiver for their family. Some females also described being both the caregiver and breadwinner for their family when their partners were incarcerated or involved with drugs or alcohol. Workplace wellness programs and incentives as well as relationships with coworkers highlighted some ways that employed females engaged in healthy activities in the workplace. Conversely, some females noted that these opportunities were not always available or their work duties made it impossible to participate (Christiansen, Gadhoke, Pardilla, & Gittelsohn, 2019).

4. Conclusion

This review captures the available literature on occupational health and safety among AI/AN workers. Out of 145 articles, only 11 were peer-reviewed and dealt specifically with occupational health and safety among AI/AN people in the past 20 years. Of the 11 articles: four addressed agriculture, forestry, and fishing; three addressed mining; one addressed each of manufacturing and services. Lastly, two focused on AI/AN people and occupational exposures or well-being in general. It is interesting to note that the two smallest workforce sectors (agriculture, forestry, and fishing, 28,877; and mining, 16,408) have been the subject of the most research, likely attributable to high risk rates as well as congressionally appropriated funds for mining research. Other sectors with much larger workforces, such as healthcare and social assistance (211,576), services (695,309), and wholesale and retail trade (236,624) are virtually unexplored. The lack of representative research across NORA sectors underscores gaps in knowledge and the need for additional research.

Limitations of this review were the small number of articles identified and in some cases the age of the article, reflecting the likelihood that findings could be out of date. The absence of relevant literature may also be due to other priorities or lack of resources for tribes to collect data or conduct research themselves. Tribes also may choose to operationalize their sovereign status and safeguard their data and stories, which may have been mishandled, misinterpreted, and even misused in the past. For example, tribes may choose to not publish in mainstream publications or may not allow non-tribal entities to conduct research on AI/AN workers.

Additional limitations include the inability to compare injury rates across various publications, as the populations included and methodologies used varied widely. Specifically, it is difficult at times to compare findings across AI/AN communities as conclusions for one tribe may not be generalizable to all tribal communities. Finally, it is difficult to classify exposures as occupationally-related, which may lead to many occupational health and safety risks being underreported or ignored.

5. Practical Applications

General themes across these publications point to the need for increased overall awareness and education regarding injury prevention and risks associated with occupational injuries and fatalities. It is important that recommendations and interventions related to education be culturally appropriate and tailored specifically to AI/AN communities at risk. Including AI/AN workers in the development of worker safety and health materials and trainings would ensure that these interventions are culturally appropriate. Differentiating between AI/AN workers who live on tribal lands and those who live in the general population is also important in order to understand the unique needs of those who live and work on tribal lands. It is also important to acknowledge and honor tribal uniqueness and practices.

Increased use of PPE is recommended specifically for agriculture and fishing, as well as those exposed to metal dust. Addressing adherence to safety protocols, including the use of PPE, is also necessary in order to reduce occupational injury risk. According to the hierarchy of controls, PPE is considered the least effective at protecting workers. Therefore, it is important that additional research investigates interventions and evaluations for elimination, engineering, and administrative controls for AI/AN workers, as well as all workers in these industries.

Improving occupational health and safety surveillance systems to accurately identify worker race and ethnicity is vital to better understanding the safety and health trends among AI/AN workers. The inclusion of AI/AN as a race category is necessary in all data collection methods, not just those related to occupational health and safety. Requiring responses for race and ethnicity in surveillance systems is also important to help identify vulnerable worker populations (National Academies of Sciences, Engineering, and

Medicine, 2018). Furthermore, collecting tribal affiliation could be extremely valuable with 574 federally recognized tribes and non-federally recognized tribes in the United States (National Conference of State Legislatures, 2020; Bureau of Labor Statistics, 2019). Future research could utilize community-based participatory research strategies in order to identify additional issues in worker safety and health and approaches that may be unique to AI/AN communities. Utilizing already available data, such as Web-Cident, My Tribal Data, and state health departments, beyond traditional worker data sources to further explore AI/AN worker safety and health trends is also necessary.

Lastly, additional research is needed to fully understand the socioeconomic inequities experienced by AI/AN people resulting from structural racism, such as living in poverty, lacking a high school degree, and lacking access to health insurance (Sequist, 2021). These disparities related to social determinants of health may also contribute to inequalities AI/AN people face in the workforce, such as larger numbers of young workers and workers in the service sector, that are too complicated to be addressed through this literature review analysis. Considering the role of social determinants of health as well as AI/AN culture and social contexts in occupational health research and prevention measures is one way to examine work-related risks and outcomes unique to AI/AN workers (Flynn, Check, Steege, Siven, & Syron, 2022). Gaps in occupational safety and health research for AI/AN workers indicate the need for additional and renewed efforts to identify workplace disparities and to educate workers and employers regarding effective interventions and prevention strategies.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

Conflict of Interest

The authors have no conflicts of interests to disclose.

References

§ 4.01[1][a]. (2017). In N. J. Newton (Ed.), *Cohen's Handbook of Federal Indian Law.*

Bureau of Labor Statistics. (2019). *American Indians and Alaska Natives in the U.S. labor force*. Retrieved from Monthly Labor Review: https://www.bls.gov/opub/mlr/2019/article/american-indians-and-alaska-natives-in-the-u-s-labor-force.htm.

Bureau of Labor Statistics, U.S. Department of Labor. (2020). *Census of Fatal Occupational Injuries, Table A-7, Fatal Occupational Injuries by Worker Characteristics and Event or Exposure, All United States, 2020.* Retrieved from https://www.bls.gov/iif/oshwc/cfoi/cftb0342.htm.

Bureau of Labor Statistics, U.S. Department of Labor. (2020). *Fatal occupational injuries, total hours worked, and rates of fatal occupational injuries by selected worker characteristics, occupations, and industries, civilian workers, 2019.* Retrieved from Current Population Survey, Census of Fatal Occupational Injuries: https://www.bls.gov/iif/oshwc/cfoi/cfoi_rates_2019hb.xlsx.

Bureau of Labor Statistics, U.S. Department of Labor. (2020). *Number of nonfatal occupational injuries and illnesses involving days away from work by selected worker and case characteristics and race or ethnicity, All U.S., all ownerships, 2020.* Retrieved from Occupational Injuries/Illnesses and Fatal Injuries Profiles: https://data.bls.gov/gqt/InitialPage.

Bureau of Labor Statistics, U.S. Department of Labor. (2021). *Survey of Occupational Injuries, 2021*. Retrieved from https://www.bls.gov/respondents/iif/forms/soii-sy2021-long-form-omb.pdf.

Center for State, Tribal, Local, and Territorial Support, Centers for Disease Control and Prevention. (20170). Public Health Law. Retrieved from Selected Tribal Laws Related to Occupational Safety and Health: https://www.cdc.gov/phlp/docs/menu-tribalosh.pdf.

Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health. (2022). Retrieved from Employed Labor Force Query System: https://wwwn.cdc.gov/wisards/cps/default.aspx.

Christiansen, K., Gadhoke, P., Pardilla, M., & Gittelsohn, J. (2019). Work, worksites, and wellbeing among North American Indian women: A qualitative study. *Ethnicity & Health, 24*(1), 24–43. https://doi.org/10.1080/13557858.2017.1313964.

Council of State and Territorial Epidemiologists. (2017). Occupational health indicators: A guide for tracking occupational health conditions and their determinants. Retrieved from https://cdn.ymaws.com/www.cste.org/resource/resmgr/pdfs/pdfs2/2017_OHI_Guidance_Manual_201.pdf.

Crichlow, R., Williamson, S., Geurin, M., & Heggem, H. (2006). Self-reported injury history in Native American professional rodeo competitors. *Clinical Journal of Sport Medicine, 16*(4), 152–154. https://doi.org/10.1097/00042752-200607000-00012.

Duysen, E., Irvine, K., Yoder, A., Topliff, C., Kelling, C., & Rajaram, S. (2017). Assessment of tribal bison worker hazards using trusted research facilitators. *Journal of Agromedicine, 22*(4), 337–346. https://doi.org/10.1080/1059924X.2017.1353937.

Flynn, M. A., Check, P., Steege, A. L., Siven, J. M., & Syron, L. N. (2022). Health Equity and a Paradigm Shift in Occupational Safety and Health. *International Journal of Environmental Research and Public Health, 19*(1), 349. https://doi.org/10.3390/ijerph19010349.

Goldcamp, E. M., Hendricks, K. J., Layne, L. A., & Myers, J. R. (2006). Nonfatal injuries to household youth on Native American operated farms in the U.S., 2000. *Journal of Agromedicine, 11*(3–4), 61–69. https://doi.org/10.1300/J096v11n03_07.

Gonzales, M., Shah, V., Bobelu, A., Qualls, C., Natachu, K., Bobelu, J., ... Zager, P. (2004). Concentrations of Surface-Dust Metals in Native American Jewelry-Making Homes in Zuni Pueblo, New Mexico. *Archives of Environmental and Occupational Health, 59*(5), 245–249. https://doi.org/10.3200/AEOH.59.5.245-249.

Helitzer, D. L., Hathorn, G., Benally, J., & Ortega, C. (2014). Culturally relevant model program to prevent and reduce agricultural injuries. *Journal of Agricultural Safety and Health, 20*(3), 175–198. https://doi.org/10.13031/jash.20.10333.

Hill, R. D., Reyes, E. D., & Dalsey, E. J. (2013, December). NIOSH Launches Worker Safety and Health Initiative for American Indians and Alaska Natives. *The IHS Primary Care Provider, 38*(12). Retrieved from https://www.ihs.gov/sites/provider/themes/responsive2017/display_objects/documents/2010_2019/PROV1213.pdf.

Indian Health Service. (2019, October). *Disparities*. Retrieved from Indian Health Service: https://www.ihs.gov/newsroom/factsheets/disparities/.

Jones, B. A. (2017). The social costs of uranium mining in the US Colorado Plateau cohort, 1960–2005. *International Journal of Public Health, 62*(4), 471–478. https://doi.org/10.1007/s00038-017-0943-z.

Klepeis, N. E., Dhaliwal, N., Hayward, G., Acevedo-Bolton, V., Ott, W. R., Read, N., ... Moore, R. S. (2016). Measuring indoor air quality and engaging California Indian stakeholders at the Win-River Resort and Casino: Collaborative Smoke-Free Policy Development. *International Journal of Environmental Research and Public Health, 13*(1), 143. https://doi.org/10.3390/ijerph13010143.

Mulloy, K. B., James, D. S., Mohs, K., & Kornfeld, M. (2001). Lung cancer in a nonsmoking underground uranium miner Retrieved from. *Environmental Health Perspectives, 109*(3), 305–309 https://www.jstor.org/stable/3434701.

National Academies of Sciences, Engineering, and Medicine. (2018). Appendix A, Recommendations. In *A Smarter National Surveillance System for Occupational Safety and Health in the 21st Century*. Washington, DC: National Academies Press (US).

National Conference of State Legislatures. (2020, March). *Federal and State Recognized Tribes*. Retrieved from National Conference of State Legislatures: https://www.ncsl.org/research/state-tribal-institute/list-of-federal-and-state-recognized-tribes.aspx.

National Occupational Research Agenda. (2018). *Sector-Based Approach - Sectors*. Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/nora/sectorapproach.html.

Redwood, D., Lanier, A. P., Brubaker, M., Orell, L., Tom-Orme, L., George, C., ... Slattery, M. (2012). Occupational and environmental exposures among Alaska Native and American Indian people living in Alaska and the Southwest United States. *Journal of Environmental Health, 74*(9), 22–28.

Schubauer-Berigan, M. K., Daniels, R. D., & Pinkerton, L. E. (2009). Radon exposure and mortality among white and American Indian uranium miners: An update of the Colorado Plateau cohort. *American Journal of Epidemiology, 169*(6), 718–730. https://doi.org/10.1093/aje/kwn406.

Sequist, T. D. (2021). Improving the Health of the American Indian and Alaska Native Population. *JAMA, 325*(11), 1035–1036. https://doi.org/10.1001/jama.2021.0521.

Steege, A. L., Baron, S. L., Marsh, S. M., Menendez, C. C., & Myers, J. R. (2014). Examining occupational health and safety disparities using national data: A cause for continuing concern. *American Journal of Industrial Medicine, 57*, 527–538. https://doi.org/10.1002/ajim.22297.

U.S. Census Bureau. (2020). *2016-2020 American Community Survey 5-Year Demographic and Housing Estimates*. Retrieved from https://data.census.gov/cedsci/table?q=DP05&y=2020.

Kait Wingate is an epidemiologist at the National Institute for Occupational Safety and Health (NIOSH). She works on data analysis and manuscript writing for research pertaining to the NIOSH Oil and Gas Sector Extraction program. She also helps to manage the NIOSH Fatalities in Oil and Gas database. She has a Master of Public Health degree in Global Epidemiology from the Emory University Rollins School of Public Health.

Elizabeth Dalsey is a Health Communication Specialist in the NIOSH Western States

Division in Denver, CO. She conducts communication activities and outreach for the Division to promote NIOSH knowledge, interventions, and technologies. Elizabeth also manages the NIOSH American Indian/Alaska Native Initiative. She began work at NIOSH in 2006 after receiving her Master of Arts in Communication from Michigan State University.

Deborah Scott received a Master of Public Health in Health Promotion and Evaluation from the University of Texas, School of Public Health. As the President of Sage Associates, Inc, Deborah is a founding officer of the Native American Health Coalition, a community-based organization serving American Indians and Alaska Natives living in the southeast Texas area. She has also served as a consultant for the NIOSH, the National Native American AIDS Prevention, and the National Indian Women's Health Resource Center.

Journal of Safety Research 84 (2023) 232–242

Contents lists available at ScienceDirect

Journal of Safety Research

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

Assessment of different pedestrian communication strategies for improving driver behavior at marked crosswalks on free channelized right turns

Abdulkarim Almukdad [a], Deepti Muley [b,*], Radwan Alfahel [c], Firas Alkadour [c], Reem Ismail [c], Wael K.M. Alhajyaseen [b]

[a] Qatar Transportation and Traffic Safety Center, College of Engineering, Qatar University, Qatar
[b] Qatar Transportation and Traffic Safety Center, Department of Civil & Architectural Engineering, College of Engineering, Qatar University, PO Box 2713, Doha, Qatar
[c] Department of Civil & Architectural Engineering, College of Engineering, Qatar University, Qatar

ARTICLE INFO

Article history:
Received 11 September 2021
Received in revised form 1 May 2022
Accepted 27 October 2022
Available online 4 November 2022

Keywords:
Driver behavior
Yielding behavior
Pedestrian safety
Unsignalized crossings
Logistic regression

ABSTRACT

Introduction: Previous studies have indicated low driver yielding rates to pedestrians in various countries. This study analyzed four different strategies to improve driver yielding rates at marked crosswalks on channelized right turn lanes at signalized intersections. *Method:* A sample of 5,419 drivers was collected for four gestures using field experiments for males and females in the State of Qatar. The experiments were conducted in daytime and nighttime on weekends at three different locations; two sites are located in an urban area and the third is located in non-urban area. The effect of pedestrians' and drivers' demographic characteristics, gestures, approach speed, time of the day, location of the intersection, car type, and driver distractions on yielding behavior is investigated using logistic regression analysis. *Results:* It was found that for the base gesture, only 2.00% of drivers yielded to the pedestrians, while for hand, attempt, and vest-attempt gestures the yielding percentages were considerably higher, 12.81%, 19.59%, and 24.60%, respectively. The results also showed that females received significantly higher yielding rates compared to males. In addition, the probability of a driver yielding increased 2.8 times when drivers approached at slower speed compared to a higher speed. Further, drivers' age group, accompanied, and distractions were not significant in determining drivers' probability of yielding.

1. Introduction

According to the World Health Organization (WHO), approximately 1.35 million fatalities around the world are caused by traffic accidents. Furthermore, traffic accidents are responsible for considerable economic losses to individuals and nations as a whole. Among the total road deaths, vulnerable road users (i.e., pedestrians, cyclists, and motorcyclists) contribute to more than 50% of total fatalities (World Health Organization, 2020). In the State of Qatar, pedestrian fatalities contributed to 28.6% of total road fatalities; slightly higher compared to 25% in United Arab Emirates and 17% in the United States (de Albuquerque & Awadalla, 2020; Gulf Times, 2020; National Center for Statistics and Analysis, 2019). Moreover, statistical data from 2005 to 2018 showed that pedestrian fatalities contributed to an average of one third of total fatalities in the State of Qatar (National Road Safety Strategy, 2018). Although detailed crash data showing exact pedestrian-vehicle crash location is not available for the State of Qatar, the issues with pedestrian safety and presence of a higher proportion of residents from varied background/nationalities emphasizes the need for this research (Timmermans et al., 2019, 2020). Generally, pedestrian crossings are facilitated by signalized crosswalks at intersections and midblock, or by marked crosswalks at channelized right turn lanes. At signalized crossings, pedestrians have dedicated time for completing crossing. Typically, in the State of Qatar, right turn lanes at signalized intersections are channelized and right turning traffic do not enter the intersections. Usually, marked crosswalks are provided on these dedicated right turn lanes to designate crossing area for pedestrians. The pedestrian crossing area is designated by standard white stripes (zebra markings) with a warning sign fixed ahead to alert drivers that they are approaching the pedestrian crossing. These crosswalks

* Corresponding author.
E-mail addresses: aa1304575@student.qu.edu.qa (A. Almukdad), deepti@qu.edu.qa (D. Muley), ra1404482@student.qu.edu.qa (R. Alfahel), fa098462@student.qu.edu.qa (F. Alkadour), ri1607496@student.qu.edu.qa (R. Ismail), wyaseen@qu.edu.qa (W.K.M. Alhajyaseen).

https://doi.org/10.1016/j.jsr.2022.10.023
0022-4375/

are not controlled by traffic signals and operate on a priority basis. Globally, pedestrians have the right of way and drivers should yield to pedestrians (Muley et al., 2019). Similarly, in accordance with the State of Qatar's traffic law, drivers should stop prior to the crosswalk to ensure the safe passage of pedestrians (2007) and a fine of 300 Qatari Riyal (around 82.5 USD) is charged to the drivers who fail to yield at pedestrian crossings (MOI, 2007). However, such fines are not implemented and generally drivers tend not to yield to pedestrians, but rather compete over the right of way (Bella et al., 2017; Hirun, 2016; Malenje et al., 2019; Muley et al., 2019). Consequently, pedestrians are forced to wait for an appropriate gap to be able to cross at their designated crosswalks, which leads to failure of the intended function of the marked crosswalks. On the other hand, due to this, the pedestrians do not feel the necessity to cross only at their designated locations, which may encourage them to jaywalk, causing unsafe conditions that may lead to crashes. In order to reduce or minimize the risks and potential crashes, Pedestrian-Vehicle Interaction (PVI) must be properly understood (Iryo-Asano & Alhajyaseen, 2017b). The non-verbal communications between pedestrians and vehicles, which is highly subjected to misinterpretation, have received significant interest from researchers, government, and concerned authorities to study their impacts on the PVI aiming at developing effective mitigations and policies to minimize pedestrian fatalities (Schroeder & Rouphail, 2011). The PVI is controlled by two aspects, pedestrians' walking behavior and drivers' yielding behavior (Alhajyaseen & Iryo-Asano, 2017; Alhajyaseen et al., 2013). The study of the driver yielding behavior is considered as more relevant to road safety since pedestrians are at higher risk of injury as they have no protection, unlike drivers. Further, local field experiments are considered vital to realistically capture the impact of the various variables on driver yielding behavior as road users with different backgrounds and diverse cultures may perceive situations in different ways, and also react differently based on their own risk assessment and risk taking. Subsequently, this study aims to investigate strategies using field experiments to improve the driver yielding behavior at marked crosswalks located at exclusive right turn lanes.

2. Strategies to improve driver yielding rates

Driver's yielding behavior is influenced by many factors such as environmental factors (e.g., road geometry, time of day, weather, and legal obligations), pedestrian's characteristics (e.g., number of pedestrians, gender, and utilized gesture), and driver's characteristics (e.g., age, gender, distraction, attitude, approach speed, and temperament; Ferenchak, 2016; Hirun, 2016; Iryo-Asano & Alhajyaseen, 2017; Stapleton et al., 2017). The impact of using advance yield markings on drivers' yielding behavior was investigated by Samuel et al. (2013). The markings were provided at about 6.1 m to 15.24 m ahead of the crosswalks to give the drivers more time to check and stop before reaching the crosswalks. The results of the before and after study showed an 8.2% increase in driver's yielding at the crosswalk. Other studies found that ensuring the visibility of pedestrians was proven to be effective in enhancing yielding rates and preventing pedestrian crashes (Clark et al., 2019; Retting et al., 2003; Schroeder & Rouphail et al., 2014). Clark et al. (2019) examined the effect of using Pedestrian Crossing Flags (PCFs) at marked crosswalks and found a significant increase in driver yielding behavior. In a similar study by Turner et al. (2006), high yielding rates reaching an average of 65% were reported when PCFs were being used. In order to further increase driver awareness, there should be a common way of communication between drivers and pedestrians as interface between them have endured increased conflicts globally over the years. Consequently, the method of communication can play a major role on the percentage of yielded vehicles (Hu & Cicchino, 2018; Ravishankar & Nair, 2018; Zhuang & Wu, 2011). Guéguen, Eyssartier, and Meineri (2016) investigated the impacts of pedestrians' smile on driver's yielding behavior. It was found that the pedestrian's smile increased number of drivers yielding to pedestrians when they crossed at crosswalk or outside pedestrian crossing. This positive effect was observed regardless of gender of driver and the pedestrian. Other ways of communication that pedestrians can use to show their intent at crosswalks are facial expressions and direct gaze. Direct gaze is an emotional and social cue that can be utilized by pedestrians as a way to get the attention of drivers prior to crossing (Böckler et al., 2014). Field experiments were conducted to assess the effect of pedestrians staring at drivers yielding at crosswalks (Guéguen et al., 2015). It was found that the overall yielding percent improved significantly (67.7% vs 55.1%) when the pedestrian stared at the drivers. Further, drivers stopped more for a females rather than males. In addition, male drivers were more influenced by pedestrian's stare compared to female drivers. The reported yielding rates were promising as there was a considerable increase in the percentage of yielding vehicles through use of communication skills by pedestrians without installing any device or providing additional notifications to drivers before reaching crosswalks.

Furthermore, surveys and field experiments were carried out in China to find which gesture out of 11 studied gestures would have the highest impact on drivers yielding rate (Zhuang & Wu, 2014). From the surveys, four gestures were selected by the drivers for their good clarity, familiarity, and visibility. The gestures that were selected were 'L-bent-level,' 'R-bent-erect,' 'T gesture,' and 'L-straight-erect.' For the site experiments, only 'L-bent-level' gesture was found to have a significant impact on yielding rates and decrease in speed of vehicles prior to crosswalks. The yielding rate was increased from 1.2% to 8.2% when 'L-bent-level' gesture was used in comparison with the baseline scenario with no gesture. Also, this gesture had no impact on the drivers' comfort, which was observed by monitoring the use of horn or shift in lanes. Since the side effects are negligible with positive impact on the driver's yielding, it was recommended for pedestrians to utilize the 'L-bent-level' gesture (Zhuang & Wu, 2014). Moreover, results to field experiments on the drivers behavior at crosswalks using two gestures (raising hand and extended arm) were provided by Crowley-Koch et al. (2011). The experiments were conducted at 10 different locations in the United States. The average increase in the yielding rate percent for all locations were 18.67% and 33.59% in the case of extended arm and raising hand, respectively, in comparison with baseline yielding rate (i.e., no gestures). It was further recommended to combine those gestures with other interventions like engineering changes to further enhance the yielding rate.

Another way of communication that was studied is gratitude (Nasar, 2003). Experiments were conducted for a period of three weeks. In the first and third weeks, the data were collected without treatment. However, at the second week the treatment was implemented. Two signs were used for treatment; one sign thanking the driver for stopping if the driver stopped, and a second sign saying please stop next time if the driver did not stop. The second sign was held by the person standing downstream of the road to ensure that drivers who did not yield had seen the sign. The results showed that the yielding rate increased to 50.9%, while base yielding rate were 46% and 37.3% for week 1 and week 3, respectively. Furthermore, it was also shown that during the second week, the yielding rate of vehicles on the downstream road was 44%, while for other weeks the yielding rates were 38% and 42%. It was concluded that hand-held signs enhanced the yielding rate of drivers at crosswalks. Several other variables impacting the drivers' yielding behavior were studied by researchers. Interestingly the yielding rate was enhanced by 40.27% (reaching 94.11%) when flashing devices were installed at the crossings (Lantieri et al., 2020). The adopted system consisted of in-curb LED flashing white strips, backlit 'yield here to pedestrian' vertical signs, flashing orange beacons, and enhanced lighting.

Malenje et al. (2019) investigated the effect of six environmental factors on the driver's yielding behavior at 13 uncontrolled crosswalks in Shanghai, China. It was found that the temporal gap size and number of traffic lanes had the largest impact on the driver yielding behavior. Also, higher yielding rates were observed in the presence of police. Another significant variable impacting the yielding behavior was the approach speed of the vehicle. Several studies noted that a driver tends to yield only if a reasonable reaction can be made depending on the travel speed, distance to conflict area, and maximum deceleration rate that the driver feels comfortable making (Bertulis & Dulaski, 2014; Chen et al., 2016; Dutta & Ahmed, 2018; Fricker & Zhang, 2019; Lu et al., 2016; Schneider et al., 2018; Wang et al., 2016).

A survey conducted by Hirun (2016) indicated that more than 50% of the drivers did not have knowledge about the right of way of pedestrians at zebra crossings in Thailand. The impact of public enforcement campaigns regarding the right of way at pedestrian crossings was studied and found to have a positive influence on the increase of yielding rates (Van Houten et al., 2013, 2017). Further, a study was conducted in three different locations with different types of non-signalized crossings to determine driver yielding rates (Fu et al., 2018) in three different situations: (1) the driver cannot stop completely; (2) the drivers' stopping is based on their reaction time; and (3) the drivers can stop completely. It was noted that the highest yielding rate was associated with the site containing painted crosswalk and drivers having sufficient time to stop (82.4% for situation 3 compared to 52.0% for situation 2 and 0% for situation 1. However, considering the applied safety measures, the site containing a stop sign showed higher time for vehicles to reach the crosswalk (TC) and lower vehicle deceleration rate (DRS) than the other sites. Likewise, the direction of crossing was also found to be significantly affecting yielding behavior in the State of Qatar (Muley et al., 2017). The study concluded that the pedestrians' attempting to cross from the sidewalk toward the intersection had a higher yielding rate.

In summary, several countermeasures and approaches have been studied to improve yielding rates at crosswalks. Remarkably, the main focus from previous studies was set on the effect of gestures on drivers yielding behavior without taking into account the other factors including time, location type, demographic characteristics, car type, and distractions. To the best of our knowledge, the effect of gestures in combination with the other parameters has not been investigated before. Moreover, the effect of gestures on the drivers' yielding behavior at marked (unsignalized) crossing, has not been studied in the State of Qatar and other GCC countries, which are characterized with a very heterogeneous population with very diverse cultural backgrounds and habits (Soliman et al., 2018; Timmermans et al., 2019). Therefore, this experimental study was conducted to collect field data to address this gap. The detailed research objectives are provided in the following section.

3. Study objectives

This study aims to investigate the yielding behavior of drivers at marked crosswalks located on the channelized right-turn lanes at signalized intersections. The underlying hypothesis of the study is that the drivers' decision to yield depends on multiple factors including driver's characteristics (e.g., gender, age), pedestrian characteristics (e.g., gender, gesture), vehicle characteristics (car type, and approach speed), and environment characteristics (time of the day, location of the intersection [Urban vs Non-urban]). Two objectives will be tested: (a) first is to determine the factors affecting driver yielding behavior under given circumstance and (b) second is to assess best strategy to improve driver yielding rates by testing different gestures.

4. Experimental setup

4.1. Participants

Three male and two female postgraduate students participated in this study as subjects and observers. Accordingly, the study considered two different situations considering males and females. Males were dressed similarly in casual clothes (T-shirt, jeans, and

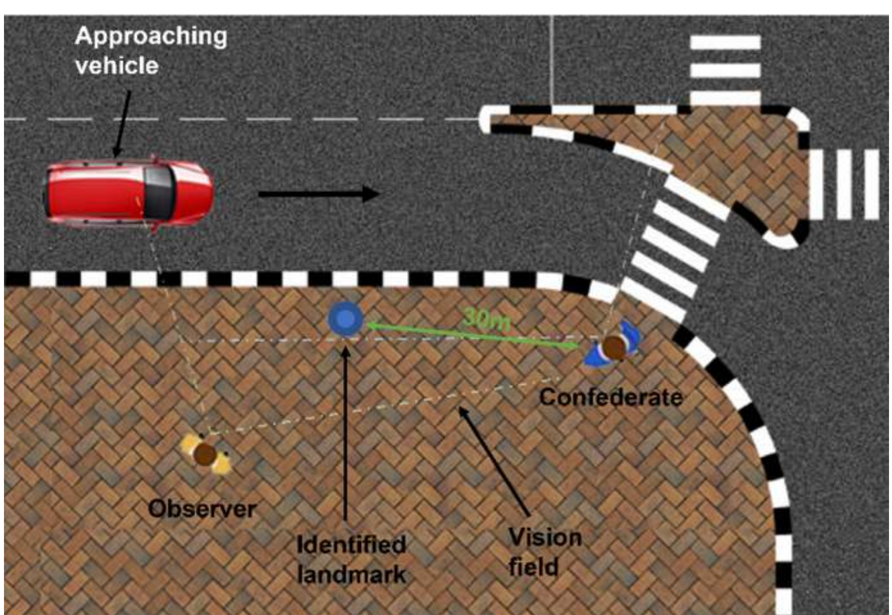


Fig. 1. Typical conditions for data collection.

sneakers) and the females were wearing traditional Arabic clothing (Abaya and Hijab). Other participants were drivers on the roads during the observation period. To ensure consistency, several meetings were held with the participating students to agree on the steps followed during experimental sessions. In addition, observers were trained to adequately view, analyze, and record the driver's yielding behavior uniformly. Pilot studies were conducted inside the Qatar University campus by mentors to ensure uniformity and accuracy in data collection.

4.2. Study sites

The experiments were conducted at three marked crosswalks located at channelized right-turn lanes at signalized intersections in the State of Qatar. It should be noted that all three sites have similar geometric layout and same type of control at the free right turn. Further, two sites were located in an urban area, in Doha city (Tawar intersection and The Mall intersections), and the third was located in non-urban area, in Al-Khor city (Al-Khor intersection). The pedestrian crossing area is defined by standard white stripes (zebra markings) with a warning sign placed fixed ahead to warn drivers approaching the pedestrian crossing. Fig. 1 presents a typical road layout at the selected crossings.

As shown in Fig. 2**,** the Tawar intersection and The Mall intersections were selected as representatives of urban area as the two intersections were nearby shopping malls and residential area. The Al-khor intersection was selected as representative of non-urban area due to its low population density and presence of fewer commercial establishments. The Tawar intersection is a major intersection located close to a large shopping center and business area. Similarly, The Mall intersection is located in front of a large and famous shopping center in Doha city and with many commercial establishments in the vicinity. The Al-Khor intersection is located in a small coastal city in the State of Qatar, 50 kilometers north of the capital, Doha. Few shops and residential buildings are located in the vicinity of this intersection.

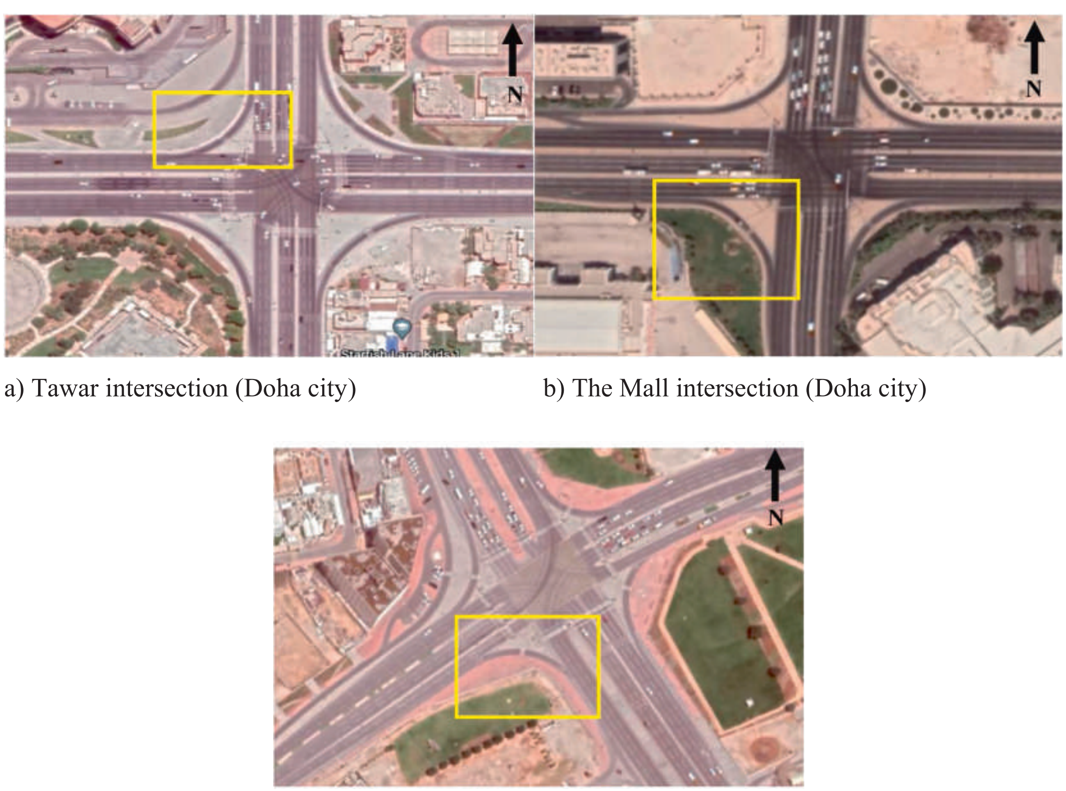
Table 1 provides a summary of the site characteristics. All three sites had similar characteristics in terms of approach speed of 80 km/hr, and the availability of entry deceleration lane. Slight variations were noted in the width of the crosswalk (varied from 5.00 m to 5.50 m) and the number of approach lanes (from four lanes for Al-Khor intersection to six lanes for Tawar and The Mall intersections).

4.3. Conditions

This study was conducted on weekends (Fridays and Saturdays) twice a day; daytime and nighttime. The data were collected in clear weather so that the visibility and stopping sight distances were not impacted. Further, the data were only recorded when no other pedestrian, except the subject, was present on the approach. Consequently, the driver's reaction to the subject crossing the marked crosswalk was recorded. Additionally, in the case of traffic jam or queues of vehicles waiting to merge on the intended approach, the experiment was stopped and resumed once the free flow of traffic was resumed. Moreover, this study only considered drivers in sedan cars and Sport Utility Vehicles (SUV). The drivers of trucks and motorcycles were excluded from the study as they were only 0.03% of the total vehicles encountered during the experiment.

4.4. Gestures

Four different gestures were used by the subjects while performing the experiments at the crosswalks, as shown in Fig. 3. For all the gestures, when the vehicle reached a predefined land-



a) Tawar intersection (Doha city)

b) The Mall intersection (Doha city)

c) Alkhor intersection (Alkhor City)

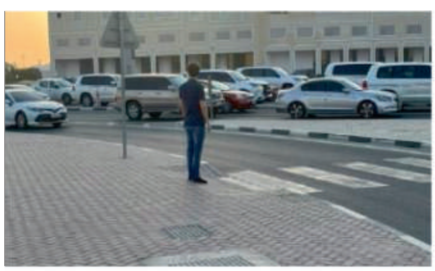
Fig. 2. Three marked uncontrolled crosswalks located at channelized right-turn lanes in Qatar.

mark located approximately 30 meters away from the pedestrian crossing, the subject indicated his/her intent to cross by making direct contact with the driver. It was ensured that there are no obstacles between the vehicle and pedestrian jeopardizing the driver's vision. In the first gesture, the pedestrian maintained a neutral facial expression and approached the crosswalk without any other interaction with the driver (Fig. 3(a)). This gesture was considered as the base gesture and was used for comparison purpose. For the second gesture, called hand gesture, the pedestrian raised one hand without moving his/her body to seek yielding. The hand was raised straight at chest height and palm facing the driver (Fig. 3(b)). While in the third gesture, the pedestrian made an attempt to cross without raising the hand, this is called an attempt (Fig. 3(c)). The fourth gesture was similar to the third gesture, but the pedestrian was wearing a fluorescent vest, hence called v-attempt (Fig. 3(d)). This gesture was introduced to test low/no cost strategies to improve driver yielding behaviors without any infrastructure improvement.

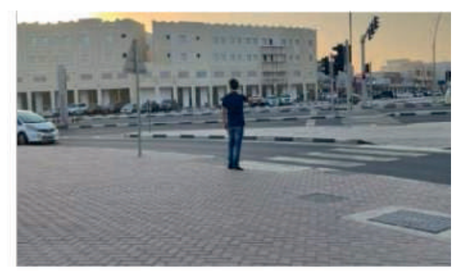
4.5. Procedure

The data were collected when the subject was crossing toward the intersection for all the cases. Further, the observers were at discreet locations to avoid any influence on driver behavior. If the driver yielded, the pedestrian crossed the street and returned back to the same place to repeat the procedure. If the driver didn't yield, then the pedestrian walked back for some distance and continued the experiment with another approaching driver. When the subject was performing the experiment, two other participants recorded the driver characteristics and behavior as well as vehicle characteristics from a discreet location. The observation sessions lasted until criteria for a minimum one hour period and 80 observations recorded per time period per gesture per site. This procedure was repeated for all four gestures, for two situations (male and female subjects), and for two conditions (daytime and nighttime) at each of the three crossings.

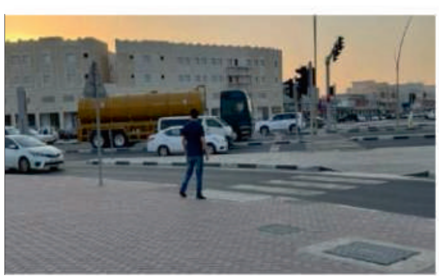
For each observation, the observers recorded the yielding behavior along with driver and vehicle characteristics. Yielding was considered when the driver reached a complete stop prior to the crosswalk allowing the pedestrian to cross. Meanwhile, when the driver passed the crosswalk without allowing the pedestrian to cross, it is considered as no yielding condition. Furthermore, the recorded data included the location of crossing, situation (daytime and nighttime), and the utilized gesture. The driver characteristics include estimated age group, gender, distractions (if using a handheld device), and whether the driver was accompanied by another person. On the other hand, the characteristics of the vehicle included type of vehicle and approach speed. The driver's age



a) Base gesture



b) Hand gesture



c) Attempt gesture

d) V-attempt gesture

Fig. 3. Gestures used in this study.

Table 1
Site characteristics.

Site	*Number of Lanes*	*Approach*	*Speed Limit (km/hr)*	*Type of Yield Indication*	*Availability of exit Acceleration Lane*	*Availability of entry Deceleration Lane*	*Crosswalk characteristics*	
							length (m)	*width (m)*
Tawar	Six lanes - Divided	North	80	Yield sign	No	Yes	3	5.50
The Mall	Six lanes - Divided	West	80	Yield sign	No	Yes	3	5.30
Al-khor	Four lanes - Divided	West	80	Yield sign	No	Yes	3	5.00

was categorized into three age groups; young (18–24 years), middle age (25–44 years), and older age (≥45 years). The observer subjectively assigned each driver to the suitable age group based on his/her visual assessment. Approach speed was noted as slow or other. Slow approach represented a speed of 20-25kmph. Observers were trained, through several trails, to recognize this speed and note down the same as slow. All other observed speeds were categorized as other. Fig. 4 summarizes the data framework used for this paper. All the experiments were conducted during weekends; between 3:00 PM to 6:00 PM for the daytime condition and 7:00 PM to 10:00 PM for the nighttime condition. The data collection was performed between November 2020 and February 2021. During this period of the year, the weather is usually clear with temperature between 18 and 25 degrees.

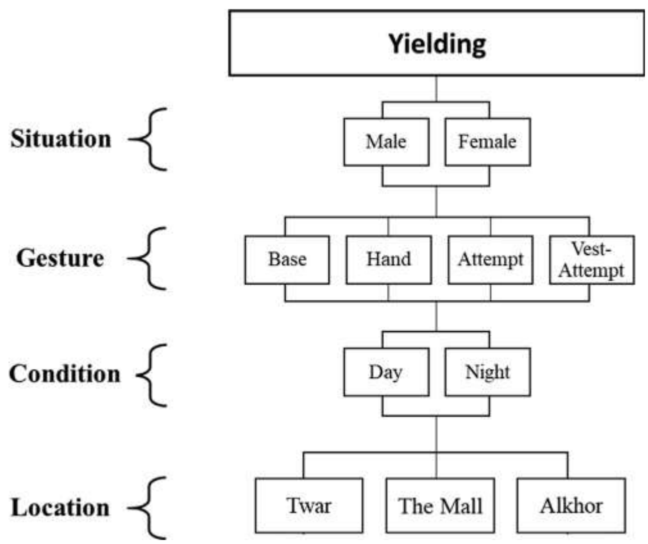


Fig. 4. Overview of the data collection framework.

5. Data overview

In total, 5,419 driver observations were collected while conducting this study, out of which 2,720 observations were for males and 2,699 were for females. The total number of observations for base, hand, attempt and vest-attempt gestures were 1597, 1366, 1220, and 1236, respectively. Alkhor intersection has the lowest number of observations with a total of 1,640, followed by The Mall intersection with a total number of 1,884 observations, and Tawar intersection with a total of 1,895 observations. The proportion of female drivers (24.5%) was found to be lower than male drivers (75.5%). Further, the percentage of SUVs were higher than sedan cars by 10.83%, as shown in Table 2. Table 3 summarizes the yielding rates for each experimental condition for male and female subjects. Overall, the female subjects received significantly higher yielding rate (Mean: 16.30%, Standard Deviation (SD): 36.95%) compared to male subjects (Mean: 11.40%, SD: 31.78%) (two-tailed/unpaired: $t_{(5287)} = 5.238$, $p < .001$). However, the increase in yielding rate for females in comparison to males was found to be insignificant when location, condition, and gesture were considered separately as shown in Table 3. Overall, 2.00%, 12.81%, 19.59%, and 24.60% yielding rates were observed for base, hand, attempt, and v-attempt gestures, respectively.

6. Model development and analysis

To test the impact of various driver and built environment characteristics, Binary Logistic Regression (BLR) models were developed using IBM SPSS 26.0 software. The model is aimed to predict the likelihood of driver's yielding for given conditions. Forward selection approach was used with a cut-off value of 0.05 significance level. The independent variable, driver yielding was coded as a binary variable, with a value of 1 when a driver yields to the pedestrian and 0 when the driver doesn't yield to the pedes-

Table 2
Number of observations for drivers' gender and vehicle types considering each location, condition and gesture.

Site	*Condition*	*Gesture*	*Drivers' gender*		*Vehicle types*	
			Male	*Female*	*Sedan*	*SUV*
Al khor	Day	Base	179	18	96	101
		Hand	163	23	77	109
		Attempt	164	22	82	104
		Vest Attempt	183	28	84	127
	Night	Base	187	38	99	126
		Hand	176	34	101	109
		Attempt	187	33	81	139
		Vest Attempt	170	35	90	115
Tawar	Day	Base	261	56	134	183
		Hand	261	44	137	168
		Attempt	172	33	85	120
		Vest Attempt	162	39	88	113
	Night	Base	181	64	104	141
		Hand	165	45	87	123
		Attempt	156	48	101	103
		Vest Attempt	159	49	83	125
The Mall	Day	Base	185	132	146	171
		Hand	152	94	108	138
		Attempt	122	79	108	93
		Vest Attempt	141	62	96	107
	Night	Base	181	115	138	158
		Hand	139	70	92	117
		Attempt	127	77	97	107
		Vest Attempt	109	99	102	106
Overall			4082	1337	2416	3003

Table 3
Driver yielding rates for all experimental conditions.

Site/intersection	Condition	Gesture	Male subject		Female subject		t-test	
			% yielding	Total observations	% yielding	Total Observations	t	Sig.
Al khor	Day	Base	0.00%	107	1.11%	90	1.000	0.32
		Hand	6.98%	86	11.00%	100	0.961	0.338
		Attempt	19.15%	94	27.17%	92	1.295	0.197
		Vest Attempt	21.90%	105	29.25%	106	1.221	0.224
	Night	Base	0.00%	124	0.99%	101	1.000	0.32
		Hand	3.00%	100	7.27%	110	1.414	0.159
		Attempt	10.28%	107	15.93%	113	1.243	0.215
		Vest Attempt	20.95%	105	30.00%	100	1.485	0.139
Tawar	Day	Base	1.95%	154	6.75%	163	2.119	**0.035**
		Hand	12.42%	161	12.50%	144	0.020	0.984
		Attempt	13.59%	103	14.71%	102	0.228	0.820
		Vest Attempt	13.86%	101	17.00%	100	0.613	0.540
	Night	Base	0.83%	120	3.20%	125	1.325	0.187
		Hand	6.54%	107	8.74%	103	0.596	0.552
		Attempt	8.82%	102	10.78%	102	0.469	0.640
		Vest Attempt	13.89%	108	15.00%	100	0.227	0.821
The Mall	Day	Base	0.71%	140	3.95%	177	1.984	**0.048**
		Hand	18.85%	122	27.42%	124	1.596	0.112
		Attempt	27.72%	101	42.00%	100	2.137	**0.034**
		Vest Attempt	28.16%	103	43.00%	100	2.223	**0.027**
	Night	Base	0.00%	153	2.10%	143	1.744	0.083
		Hand	12.26%	106	22.33%	103	1.928	0.055
		Attempt	20.19%	104	27.00%	100	1.142	0.255
		Vest Attempt	27.10%	107	35.64%	101	1.325	0.187
Overall			11.40%	2720	16.30%	2699	5.238	**<0.001**

trian. The coding of dependent variables is shown in Table 4. A total of five different BLR models were developed to explore the data. The first model determined the probability of a driver yielding considering all experimental conditions. Section 6.1 describes the details of the model. While the remaining four models analyzed the probability of driver yielding for each tested gesture independently; Section 6.2 explains details these BLR models.

6.1. Overall driver yielding BLR model

The first model was used to predict the overall driver's yielding behavior. Table 5 shows the details of the model. The Nagelkerke R Square was 20.1%, illustrating that the dependent variable yielding has variance that can be described by the independent variables satisfactorily. Initially, all variables were included as input variables for model development. Out of the 10 input variables, seven were significant in predicting the probability of driver yielding. Three variables (drivers' age group, accompanied, and using phone) were excluded as they did not have significant effect on the probability of driver's yielding. Gesture was found to be the most significant variable affecting the probability of yielding. In terms of gesture, hand, attempt and v-attempt are compared with the base gesture. The odds that the driver would yield when hand, attempt, and v-attempt gestures were used are 7.467, 13.190, and 17.596 times higher than the base gesture, respectively.

Further, as the approach speed reduced, the odds of the driver yielding increased by 2.8 compared to a faster approach speed. Female drivers had higher odds of yielding than their male counterparts by 1.795 and 1.563, respectively. Furthermore, the car type had a significant impact on the driver's yielding behavior, the yielding probability of drivers' with SUVs is 1.194 times higher than the yielding probability of sedan car drivers. Furthermore, the

Table 4
Independent variables coding used in the model.

Independent variables	Code	Frequency	Independent variables	Code	Frequency
Gesture			Drivers' gender		
Base *	0	1597	Female *	0	1337
Hand	1	1366	Male	1	4082
Attempt	2	1220			
V-Attempt	3	1236			
Drivers' age group			Car Type		
Young *	0	765	Sedan *	0	2416
Middle	1	3357	SUV	1	3003
Elder	2	1297			
Approach speed			Accompanied		
Slow *	0	1099	No *	0	2705
Other	1	4320	Yes	1	2714
Location			Using phone		
Rural *	0	1640	No *	0	4691
Urban	1	3779	Yes	1	728
Subject's gender			Time		
Female *	0	2699	Day *	0	2775
Male	1	2720	Night	1	2644

Note: * shows the reference category in the BLR models.

Table 5
BLR model for predicting drivers yielding behavior.

Independent variables	*β*	*SE*	*Sig.*	*Exp(β)*	**95% C.I. for** *Exp(β)*	
					Lower	Upper
Drivers' gender	−0.585	0.093	**<0.001**	0.557	0.465	0.668
Subject's gender	−0.446	0.085	**<0.001**	0.640	0.542	0.755
Car	0.177	0.085	**0.037**	1.194	1.011	1.411
Base Gesture			**<0.001**			
Hand Gesture	2.010	0.198	**<0.001**	7.467	5.065	11.007
Attempt Gesture	2.579	0.195	**<0.001**	13.190	9.000	19.331
V-Attempt Gesture	2.868	0.193	**<0.001**	17.596	12.056	25.684
Time	−0.418	0.085	**<0.001**	0.658	0.557	0.778
Location	0.356	0.097	**<0.001**	1.428	1.182	1.726
Approach speed	−1.029	0.093	**<0.001**	0.357	0.298	0.428
Constant	−2.734	0.221	**<0.001**	0.065		

Note: SE refers to Standard Error, Significance level α = 0.05.

Table 6
Direction and level of significance of each variable considering base and hand gestures in terms of the probability of yielding.

Independent variables	*Base case*						*Hand gesture*					
	β	*Sig.*	*SE*	*Exp(β)*	95% C.I. for *Exp(β)*		*β*	*Sig.*	*SE*	*Exp(β)*	95% C.I. for *Exp(β)*	
					Lower	Upper					Lower	Upper
Drivers' gender	-	0.110	-	-	-	-	−0.431	**0.019**	0.184	0.650	0.453	0.932
Pedestrians' gender	−1.603	**<0.001**	0.461	0.201	−2.645	−0.746	−0.423	**0.013**	0.170	0.655	0.469	0.914
Time	−0.759	**0.046**	0.384	0.467	−1.583	−0.013	−0.433	**0.013**	0.174	0.649	0.461	0.913
Location	−1.602	**0.003**	0.651	0.201	−3.209	−0.465	−1.024	**<0.001**	0.227	0.359	0.230	0.560
Approach speed	−1.748	**<0.001**	0.355	0.174	−2.470	−1.027	−1.230	**<0.001**	0.181	0.292	0.205	0.417
Accompanied	0.813	**0.029**	0.375	2.255	0.081	1.609	-	0.409	-	-	-	-
Constant	−2.043	**<0.001**	0.390	0.130	−2.878	−1.305	−0.104	0.640	0.222	0.901		
Nagelkerke R square	17.5%						11.1%					

Note: - indicates that variable not significant, values not available and Significance level α = 0.05, SE refers to Standard Error.

odds of yielding were significantly higher in the daytime (1.52 times) compared to the nighttime. Moreover, the drivers' yielding probability in urban areas is 1.428 times compared to the one in non-urban area.

6.2. Gesture based driver yielding BLR models

Four separate models were estimated to determine the probability of driver's yielding for each tested gesture during the experiments. These models were developed to assess the influence of independent variables for different gestures on driver yielding rates. It should be noted that the proportion of drivers' yielding for base gesture were low, hence Firth logistic regression model was developed. This model utilizes penalized maximum likelihood instead of standard maximum likelihood estimation to avoid biased estimation of results (Rahman & Sultana, 2017). Tables 6 and 7 show the details of BLR models. It should be noted that all the independent variables shown in Table 4 were included while developing these models. However, the results for only significant variables were reported in Tables 6 and 7. The Nagelkerke R Square for base, hand, attempt, and v-attempt gestures cases were 17.5%, 11.1%, 7.2%, and 7.8%, respectively.

Drivers' gender was found to be a significant factor affecting the yielding behavior for hand, attempt, and v-attempt gestures, while for the base gesture it was not significant. In all cases, subject's gender and approach speed had shown significant impact on the driver yielding behavior. Time variable was also significant in predicting the driver yielding behavior for all gestures, except v-attempt. The location variable was significant only for the base and hand gestures, while it was not significant for the attempt and v-attempt gestures. Although the car type was significant in the first/overall model, it was not significant for the individual gestures model. In general, drivers' age group, accompanied, and using

Table 7
Direction and level of significance of each variable considering attempt and V-attempt gestures in terms of the probability of yielding.

Independent variables	*Attempt gesture*						*V-attempt gesture*					
	β	*Sig.*	*SE*	*Exp(β)*	95% C.I. for *Exp(β)*		*β*	*Sig.*	*SE*	*Exp(β)*	95% C.I. for *Exp(β)*	
					Lower	Upper					Lower	Upper
Drivers' gender	−0.565	**<0.001**	0.163	0.568	0.413	0.782	−0.664	**<0.001**	0.148	0.515	0.386	0.688
Pedestrians' gender	−0.423	**0.005**	0.149	0.655	0.489	0.878	−0.367	**0.007**	0.137	0.693	0.530	0.906
Time	−0.595	**<0.001**	0.150	0.552	0.411	0.740	-	0.198	-	-	-	-
Location	-	0.146	-	-	-	-	-	0.763	-	-	-	-
Approach speed	−0.790	**<0.001**	0.163	0.454	0.330	0.625	−0.942	**<0.001**	0.149	0.390	0.291	0.523
Accompanied	-	0.976	-	-	-	-	-	0.503	-	-	-	-
Constant	0.071	0.733	0.209	1.074			0.229	0.193	0.176	1.258		
Nagelkerke R square	7.2%						7.8%					

Note: - indicates that variable not significant, values not available, Significance level α = 0.05, SE refers to Standard Error.

phone were not significant variables with regard to the driver's yielding probability.

7. Discussion

The results showed that the female subjects received higher yielding rates and female drivers showed higher probability of yielding in comparison with their male counterparts. This can be justified by the emotional and cultural factors that people tend to associate with females out of respect, especially in the Gulf Region, which is in line with previous studies (Guéguen et al., 2016; Schroeder & Rouphail et al., 2014). Another reason can be the safer driving behaviors of females compared to males, as it was found that female drivers have fewer lapses, errors, and violations (Soliman et al., 2018). Further, female pedestrian casualties are low over the past years (National Road Safety Strategy, 2018). Additionally, the proportion of female drivers is less than male drivers; as per a recent monthly statistic provided by the Ministry of Development Planning and Statistics (MDPS), the total number of issued drivers licenses for males was 6,363 compared to 1,237 for females in December 2021 (MDPS, 2021). Moreover, the type of gesture utilized by the pedestrian prior to crossing was determined to have the highest influence on the yielding behavior. The probability of driver yielding increased by 7.467, 13.190, and 17.596 in comparison with the base gesture for hand, attempt, and v-attempt gestures, respectively. This can be linked to the higher level of assertiveness exhibited by pedestrians in attempting to cross, compared to raising hand and base gesture. Similar findings were also reported by previous studies (Böckler et al., 2014; Crowley-Koch et al., 2011; Guéguen et al., 2015, 2016; Zhuang & Wu, 2014). Furthermore, a strong correlation between vehicle approach speed and the probability of yielding is found (a 2.8 time increase in driver's yielding when they are approaching at a lower speed). Similar findings were concluded in previous studies (Bertulis & Dulaski, 2014; Chen et al., 2016; Dutta & Ahmed, 2018; Fricker & Zhang, 2019; Lu et al., 2016; Schneider et al., 2018; and Wang et al., 2016). The aforementioned can be justified by the availability of a reasonable reaction time and the ability to stop before the crosswalk with a comfortable deceleration rate, which encourage drivers to yield to pedestrians. This highlights the importance of applying traffic calming measures that encourage drivers to reduce their speeds while approaching a pedestrian crossing.

The time of day was also found as an important parameter in influencing the yielding behavior. The drivers showed higher probability of yielding to subjects (1.52 times) during the daytime compared to nighttime. However, the time of day was found to have no impact for the v-attempt gesture. This might be due to the increased visibility and attraction during nighttime due to the fluorescent vest. Furthermore, attempt and v-attempt gestures were not significantly affected by the area type, unlike base and hand gestures, which had higher probability of yielding in non-urban areas than urban areas.

The type of vehicle was found to have no impact on the probability of yielding (i.e., sedan and SUV), indicating no clear power paradox existed. However, vehicle type showed adverse effects on the yielding behavior in a previous study (Sun et al., 2003). Similarly, the findings showed that the age group of drivers and distractions (especially using mobile phones while driving or being accompanied by other passengers) had negligible impact on the probability of yielding.

Overall, the approach speed was the most significant factor in improving the driver yielding probability. This indicates that the application of various traffic calming techniques and use of advanced markings will aid in increasing driver yielding rates. Further, use of gestures also increased probability of driver yielding. This shows yielding rates can be improved without substantial physical changes or investments. Campaigns should be conducted to encourage pedestrians to use gestures to promote yielding at crosswalks on channelized right turn lanes. The campaigns can be more targeted to male drivers and specific to male pedestrians. Further, pedestrians should be encouraged to use fluorescent vests while crossing, especially during the nighttime. It should be noted that a majority of pedestrian fatalities/injuries are for foreign nationalities or expats (i.e., 74.5%) (Planning and Statistics Authority, 2019a). So the awareness campaigns should be targeted to the most affected group of residents to maximize the benefits. Use of such simple technique can help to achieve significant benefits. Further, the campaigns can also include importance of the location and time of day while encouraging driver yielding at marked crosswalks. In summary, the authorities can work on engineering and educational measures to improve driver yielding and improve pedestrian risk assessment.

8. Conclusions

This study investigates the impact of four different pedestrian gestures on driver yielding behavior under different conditions using field experiments at three locations in the State of Qatar. This study utilized observations from 5,419 drivers at marked crosswalks located on the channelized right-turn lanes at signalized intersections. Overall 13.84% of drivers yielded to pedestrians with 2.00%, 12.81%, 19.59%, and 24.60% yielding rates for base, hand, attempt, and v-attempt gestures, respectively. BLR models were developed to assess the effect of driver characteristics, pedestrian characteristics, vehicle characteristics, and environment variables on the probability of driver yielding. The results showed that the drivers' yielding behavior was highly dependent on the gestures utilized by the subjects. The odds of driver yielding for hand, attempt, and v-attempt gestures were 7.467, 13.190, and 17.596 times than that of the base gesture, respectively. Furthermore, a strong correlation was found between vehicle approach speed and the driver yielding probability. The drivers approaching at lower speeds had 2.8 times greater probability of yielding in comparison to the drivers approaching with high speed. Furthermore, drivers had a significantly higher yielding probability when encountering female subjects compared to the male subjects. Also, the study showed that female drivers had higher probability of yielding to pedestrians in comparison with male drivers. Higher yielding rates were observed when base, hand, and attempt gestures were used during daytime compared to nighttime. Further, the yielding rates for V-attempt gesture were not affected by the time of day.

The outcomes of this study highlighted the importance of several variables on the driver's yielding behavior in the State of Qatar. These results can be applicable to the Arabian Gulf countries exhibiting similar characteristics. Accordingly, several engineering and education strategies can be employed to improve the yielding rates of drivers. Various techniques enforcing reduction in approach speed prior to pedestrian crossings by restricting the speed limit or providing speed calming measures such as road markings (i.e., zigzag lines, triangles, or horizontal lines) before the crossing to alert the drivers are highly recommended. Further, authorities can launch awareness campaigns for the pedestrians to use specific gesture showing assertiveness to seek yielding while crossing. On the other hand, public awareness campaigns targeting drivers to highlight the importance of yielding to pedestrians and giving them the right of way will also be helpful. At a later stage, authorities might consider law enforcement to warn/penalize drivers who do not give way to pedestrians. Also, proper lighting at crossings must be assured to enhance visibility during nighttime. These aforementioned measures will enhance pedestrians' safety and comfortability on the roads, which can help to promote walking and reduce fatalities. The results of this study can help transport planners and traffic engineers in improving driver behavior at unsignalized right turn lanes. Furthermore, the study was conducted in the State of Qatar, which has a heterogeneous driver population with different cultural backgrounds. It should be noted that the State of Qatar is home to expats from 94 different nationalities (Snoj, 2019), which have distinct backgrounds, exposure, and driving cultures. This makes the outcomes applicable to other GCC countries exhibiting similar diverse population characteristics.

Several recommendations for future research include other types of gestures and other pedestrian's characteristics such as age, type of clothing, and nationality. Further, the yielding behavior during weekdays in comparison with weekends can be assessed. One of the limitations in this study can be the low observations of female drivers in comparison to male drivers. Low female population in the State of Qatar, due to the presence of large proportion of male workers for the many mega infrastructural projects that are ongoing as part of the preparation for the FIFA World Cup 2022, can be a reason (i.e., female 26.3% and male 73.7% in 2019; Planning and Statistics Authority, 2019b). In addition, only SUVs and sedan cars were considered in this study since other vehicle types (i.e., heavy vehicles and motorcycles) were not observed in significant numbers at study sites.

Disclaimer

No funding was received for this study.

References

Alhajyaseen, W. K. M., Asano, M., & Nakamura, H. (2013). Left-turn gap acceptance models considering pedestrian movement characteristics. *Accident Analysis & Prevention, 50*, 175–185.

Alhajyaseen, W. K. M., & Iryo-Asano, M. (2017). Studying critical pedestrian behavioral changes for the safety assessment at signalized crosswalks. *Safety Science, 91*, 351–360.

Bella, F., Natale, V., & Silvestri, M. (2017). Driver-pedestrian interaction under different road environments. *Transportation Research Procedia, 27*, 148–155.

Bertulis, T., & Dulaski, D. M. (2014). Driver approach speed and its impact on driver yielding to pedestrian behavior at unsignalized crosswalks. *Transportation Research Record, 2464*(1), 46–51. https://doi.org/10.3141/2464-06.

Böckler, A., van der Wel, R. P. R. D., & Welsh, T. N. (2014). Catching Eyes: Effects of Social and Nonsocial Cues on Attention Capture. *Psychological Science, 25*(3), 720–727. https://doi.org/10.1177/0956797613516147.

Chen, P., Wu, C., & Zhu, S. (2016). Interaction between vehicles and pedestrians at uncontrolled mid-block crosswalks. *Safety Science, 82*, 68–76.

Clark, S., Coughenour, C., Bumgarner, K., de la Fuente-Mella, H., Reynolds, C., & Abelar, J. (2019). The Impact of Pedestrian Crossing Flags on Driver Yielding Behavior in Las Vegas, NV. *Sustainability, 11*(17), 4741.

Crowley-Koch, B. J., Van Houten, R., & Lim, E. (2011). Effects of pedestrian prompts on motorist yielding at crosswalks. *Journal of Applied Behavior Analysis, 44*(1), 121–126. https://doi.org/10.1901/jaba.2011.44-121.

de Albuquerque, F. D. B., & Awadalla, D. M. (2020). Characterization of road crashes in the emirate of Abu Dhabi. *Transportation Research Procedia, 48*, 1095–1110. https://doi.org/10.1016/j.trpro.2020.08.136.

Dutta, M., & Ahmed, M. A. (2018). Gap acceptance behavior of drivers at uncontrolled T-intersections under mixed traffic conditions. *Journal of Modern Transportation, 26*(2), 119–132.

Ferenchak, N. N. (2016). Pedestrian age and gender in relation to crossing behavior at midblock crossings in India. *Journal of Traffic and Transportation Engineering (English Edition), 3*(4), 345–351.

Fricker, J. D., & Zhang, Y. (2019). Modeling pedestrian and motorist interaction at semi-controlled crosswalks: The effects of a change from one-way to two-way street operation. *Transportation Research Record, 2673*(11), 433–446.

Fu, T., Miranda-Moreno, L., & Saunier, N. (2018). A novel framework to evaluate pedestrian safety at non-signalized locations. *Accident Analysis & Prevention, 111*, 23–33. https://doi.org/10.1016/j.aap.2017.11.015.

Guéguen, N., Eyssartier, C., & Meineri, S. (2016). A pedestrian's smile and drivers' behavior: When a smile increases careful driving. *Journal of Safety Research, 56*, 83–88. https://doi.org/10.1016/j.jsr.2015.12.005.

Guéguen, N., Meineri, S., & Eyssartier, C. (2015). A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing. *Safety Science, 75*, 87–89. https://doi.org/10.1016/j.ssci.2015.01.018.

Gulf Times. (2020). *Road accidents and fatalities drop in Qatar*. https://www.gulf-times.com/story/656066/Road-accidents-and-fatalities-drop-in-Qatar#:~:text=A total of 44 cases,down by 6.7%25 in 2019.

Hirun, W. (2016). Factors Affecting Driver Yielding Behavior. *International Journal of Engineering and Technology, 8*(2), 906–913.

Hu, W., & Cicchino, J. B. (2018). An examination of the increases in pedestrian motor-vehicle crash fatalities during 2009–2016. *Journal of Safety Research, 67*, 37–44.

Iryo-Asano, M., & Alhajyaseen, W. (2017a). Consideration of a pedestrian speed change model in the pedestrian–vehicle safety assessment of signalized crosswalks. *Transportation Research Procedia, 21*, 87–97.

Iryo-Asano, M., & Alhajyaseen, W. K. M. (2017b). Modeling pedestrian crossing speed profiles considering speed change behavior for the safety assessment of signalized intersections. *Accident Analysis & Prevention, 108*, 332–342. https://doi.org/10.1016/j.aap.2017.08.028.

Lantieri, C., Costa, M., Vignali, V., Acerra, E. M., Marchetti, P., & Simone, A. (2020). Flashing in-curb LEDs and beacons at unsignalized crosswalks and driver's visual attention to pedestrians during nighttime. *Ergonomics*, 1–12.

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. *Accident Analysis & Prevention, 95*, 425–437. https://doi.org/10.1016/j.aap.2016.04.014.

Malenje, J. O., Zhao, J., Li, P., & Han, Y. (2019). Vehicle yielding probability estimation model at unsignalized midblock crosswalks in Shanghai. *China. PLoS One, 14*(3), e0213876.

Ministry of Development Planning and Statistics [MDPS]. (2021). *Traffic first section 2021*. https://www.psa.gov.qa/en/statistics/Statistical Releases/General/QMS/QMS_PSA_96_Jan_2022.pdf.

MOI. (2007). *List of traffic violations that can be settled by paying fines*. Ministry of Interior, General Traffic Directorate. https://qatarofw.com/driving-violations-and-fines/.

Muley, D., Kharbeche, M., Alhajyaseen, W., & Al-Salem, M. (2017). Pedestrians' crossing behavior at marked crosswalks on channelized right-turn lanes at intersections. *Procedia Computer Science, 109*, 233–240.

Muley, D., Kharbeche, M., Downey, L., Saleh, W., & Al-Salem, M. (2019). Road Users' Behavior at Marked Crosswalks on Channelized Right-Turn Lanes at Intersections in the State of Qatar. *Sustainability, 11*(20), 5699.

Nasar, J. L. (2003). Prompting drivers to stop for crossing pedestrians. *Transportation Research Part F: Traffic Psychology and Behavior, 6*(3), 175–182. https://doi.org/10.1016/S1369-8478(03)00024-X.

National Center for Statistics and Analysis. (2019). *Traffic Safety Facts*. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/813079.

National Road Safety Strategy. (2018). Qatar Action Plan Sectors 2018–2022; National Traffic Safety Committee. In *State of Qatar, Doha, January 2018*.

Planning and Statistics Authority. (2019a). *Annual Statistical Abstract 2019*. https://www.psa.gov.qa/en/statistics1/pages/topicslisting.aspx?parent=General&child=StatisticalAbstract.

Planning and Statistics Authority. (2019b). *FIRST SECTION POPULATION AND SOCIAL STATISTICS*. https://www.psa.gov.qa/en/statistics/Statistical Releases/Population/Population/2019/Population_social_1_2019_AE.pdf.

QTL. (2007). *Qatar Traffic Law 19-2007*. State of Qatar. https://www.almeezan.qa/LawView.aspx?opt&LawID=3993&language=en.

Rahman, M. S., & Sultana, M. (2017). Performance of Firth-and logF-type penalized methods in risk prediction for small or sparse binary data. *BMC Medical Research Methodology, 17*, 1–15.

Ravishankar, K. V. R., & Nair, P. M. (2018). Pedestrian risk analysis at uncontrolled midblock and unsignalised intersections. *Journal of Traffic and Transportation Engineering (English Edition), 5*(2), 137–147.

Retting, R. A., Ferguson, S. A., & McCartt, A. T. (2003). A review of evidence-based traffic engineering measures designed to reduce pedestrian–motor vehicle crashes. *American Journal of Public Health, 93*(9), 1456–1463.

Samuel, S., Romoser, M. R. E., Gerardino, L. R., Hamid, M., Gómez, R. A., Knodler, M. A., ... Fisher, D. L. (2013). Effect of advance yield markings and symbolic signs on vehicle-pedestrian conflicts: Field evaluation. *Transportation Research Record, 2393*(1), 139–146. https://doi.org/10.3141/2393-16.

Schneider, R. J., Sanatizadeh, A., Shaon, M. R. R., He, Z., & Qin, X. (2018). Exploratory analysis of driver yielding at low-speed, uncontrolled crosswalks in Milwaukee, Wisconsin. *Transportation Research Record, 2672*(35), 21–32.

Schroeder, B. J., & Rouphail, N. M. (2011). Event-based modeling of driver yielding behavior at unsignalized crosswalks. *Journal of Transportation Engineering, 137*(7), 455–465.

Schroeder, B., Rouphail, N., Salamati, K., Hunter, E., Phillips, B., Elefteriadou, L., ... Mamidipalli, S. (2014). *Empirically-based performance assessment & simulation of pedestrian behavior at unsignalized crossings*. Southeastern Transportation Research, Innovation, Development and Education.

Snoj, J. 2019. Population of Qatar by nationality - 2019 report [Online]. Available: http://priyadsouza.com/population-of-qatar-by-nationality-in-2017/ [Accessed 10th March 2022].

Soliman, A., Alhajyaseen, W., Alfar, R., & Alkaabi, I. (2018). Changes in driving behavior across age cohorts in an Arab culture: The case of state of Qatar. *Procedia Computer Science, 130*, 652–659.

Stapleton, S., Kirsch, T., Gates, T. J., & Savolainen, P. T. (2017). Factors affecting driver yielding compliance at uncontrolled midblock crosswalks on low-speed roadways. *Transportation Research Record, 2661*(1), 95–102.

Sun, D., Ukkusuri, S., Benekohal, R. F., & Waller, S. T. (2003). Modeling of motorist-pedestrian interaction at uncontrolled mid-block crosswalks. *Transportation Research Record, TRB Annual Meeting CD-ROM, Washington, DC.*

Timmermans, C., Alhajyaseen, W., Reinolsmann, N., Nakamura, H., & Suzuki, K. (2019). Traffic safety culture of professional drivers in the State of Qatar. *IATSS research, 43*, 286–296.
Timmermans, C., Alhajyaseen, W., Ross, V., & Nakamura, H. (2020). Introducing a multi-variate classification method: Risky driving acceptance among different heterogeneous driver sub-cultures. *Journal of Safety Research, 73*, 81–91.
Turner, S., Fitzpatrick, K., Brewer, M., & Park, E. S. (2006). Motorist yielding to pedestrians at unsignalized intersections: Findings from a national study on improving pedestrian safety. *Transportation Research Record, 1982*(1), 1–12.
Van Houten, R., Malenfant, J. E. L., Blomberg, R., Huitema, B., & Hochmuth, J. (2017). High-visibility enforcement on driver compliance with pedestrian right-of-way laws: 4-year follow-up. *Transportation Research Record, 2660*(1), 58–65.
Van Houten, R., Malenfant, L., Huitema, B., & Blomberg, R. (2013). Effects of high-visibility enforcement on driver compliance with pedestrian yield right-of-way laws. *Transportation Research Record, 2393*(1), 41–49.
Wang, T., Jiang, X., Wang, W., & Cheng, Q. (2016). Analysis of vehicle speed change at non-signalized crosswalks based on driving behavior. *Procedia Engineering, 2016*(137), 547–553.
World Health Organization, (2020). *Road traffic injuries*. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.
Zhuang, X., & Wu, C. (2011). Pedestrians' crossing behaviors and safety at unmarked roadway in China. *Accident Analysis & Prevention, 43*(6), 1927–1936.
Zhuang, X., & Wu, C. (2014). Pedestrian gestures increase driver yielding at uncontrolled mid-block road crossings. *Accident Analysis & Prevention, 70*, 235–244. https://doi.org/10.1016/j.aap.2013.12.015.

Abdulkarim Almukdad Abdulkarim is a Research Assistant at Qatar Transportation and Traffic Safety Center, College of Engineering, Qatar University. His basic degree is in Civil Engineering field. His main field of research is traffic safety.

Deepti Muley Deepti is working as a Research Associate at Qatar Transportation and Traffic Safety Center, Department of Civil & Architectural Engineering, College of Engineering, Qatar University, Qatar. She is a Civil Engineer by profession and works in the area of pedestrian safety and traffic engineering along with transport planning. She has experience of handling research as well as in consulting/industry projects.

Radwan Alfahel Radwan is a graduate students at the Department of Civil & Architectural Engineering, College of Engineering, Qatar University. His basic degree is in Civil Engineering.

Firas Alkadour Radwan is a graduate students at the Department of Civil & Architectural Engineering, College of Engineering, Qatar University. His basic degree is in Civil Engineering.

Reem Ismail Reem is a graduate student and research assistant at the Department of Civil & Architectural Engineering, College of Engineering, Qatar University, Qatar. Her basic degree is in Civil Engineering.

Wael K. M. Alhajyaseen Wael is an Associate Professor at Qatar Transportation and Traffic Safety Center, Department of Civil & Architectural Engineering, College of Engineering, Qatar University, Qatar. His basic degree is in Civil Engineering and his primary research area is Traffic Safety using field studies and driving simulator. He has published several journal and conference papers in highly reputed journals. He is handling several large scale research projects in the area of traffic safety.

Journal of Safety Research 84 (2023) 86–98

Contents lists available at ScienceDirect

Journal of Safety Research

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

Cars dent, horse riders break: Analysis of police-recorded injury incidents involving ridden horses on public roads in Great Britain

Danica Pollard [a,*], John Duncan Grewar [b]

[a] The British Horse Society, Abbey Park, Stareton, Kenilworth, Warwickshire CV8 2XZ, United Kingdom
[b] jDATA Pty (Ltd), Tambali Village, Sandbaai 7200, South Africa

ARTICLE INFO

Article history:
Received 13 December 2021
Received in revised form 21 June 2022
Accepted 17 October 2022
Available online 28 October 2022

Keywords:
Equestrian road safety
Casualties
Injury risk
Risk reduction
Near-miss

ABSTRACT

Introduction: Police-recorded road injury data are frequently used to approximate injury risk for different road user groups but a detailed analysis of incidents involving ridden horses has not previously been conducted. This study aims to describe human injuries resulting from interactions between ridden horses and other road users on public roads in Great Britain and identify factors associated with severe to fatal injuries. *Method:* Police-recorded road incident data involving ridden horses (2010–2019) were extracted from the Department for Transport (DfT) database and described. Multivariable mixed-effects logistic regression modeling was used to identify factors associated with severe/fatal injury outcomes. *Results:* A total of 1,031 injury incidents involving ridden horses were reported by police forces, involving 2,243 road users. Out of 1,187 road users injured, 81.4% were female, 84.1% were horse riders, and 25.2% (n = 293/1,161) were in the 0–20 year age category. Horse riders represented 238/267 serious injuries and 17/18 fatalities. Vehicle types involved in incidents where horse riders were seriously/fatally injured were mostly cars (53.4%, n = 141/264) and vans/light goods vehicles (9.8%, n = 26). Horse riders, cyclists, and motorcyclists had higher odds of severe/fatal injury compared to car occupants ($p < 0.001$). Severe/fatal injuries were more likely on roads with 60–70 mph speed limits versus 20–30 mph roads, while odds of severe/fatal injury increased with increasing road user age ($p < 0.001$). *Conclusions:* Improved equestrian road safety will largely impact females and young people as well as reducing risk of severe/fatal injuries in older road users and those using modes of transport such as pedal-cycles and motorcycles. Our findings support existing evidence that reductions in speed limits on rural roads would help reduce the risk of serious/fatal injuries. *Practical applications:* More robust equestrian incident data would better inform evidence-based initiatives to improve road safety for all road users. We suggest how this can be done.

1. Introduction

Equestrianism is a common sport and leisure activity in Great Britain (GB [comprised of England, Scotland and Wales]). In 2019, as part of the British Equestrian Trade Association's National Equestrian Survey, it was estimated that 3 million people in GB have ridden a horse at least once in the past year, while 1.8 million are regular riders that ride at least once a month (British Equestrian Trade Association (BETA), 2019). Exercising horses by using a combination of off-road routes and public roads (a term referred to as hacking) remains a common equestrian activity in GB with more than half of 797 horse owners surveyed between 2009 and 2011 reporting they had "hacked" their horses in the previous week (Wylie, Ireland, Collins, Verheyen, & Newton, 2013). Use of public off-road routes by equestrians is restricted in England and Wales, but not Scotland, to only those designated for horse use, meaning only a small proportion of the existing public off-road network is available to equestrians in contrast to pedestrians and cyclists (Rights of way and accessing land, 2020). Even where access to equestrian off-road routes exists, road use is often required to reach those routes or travel between them. A more recent study in 2020 found that road use by equestrians in GB and Northern Ireland is common, with 84% (4481/ 5335) of equestrians surveyed reporting they use roads at least once per week and not just for riding their horse but also for leading the horse while walking next to them, carriage driving, and riding while leading another horse (Pollard & Furtado, 2021).

* Corresponding author.
E-mail addresses: dee.pollard@bhs.org.uk (D. Pollard), info@jdata.co.za (J.D. Grewar).

https://doi.org/10.1016/j.jsr.2022.10.010
0022-4375/

Road incidents involving equestrians are also reported to be common. A survey of 426 horse riders in GB and Northern Ireland found that 60.3% had reported having a near-miss while using roads in the previous year (Scofield, Savin, & Randle, 2013). A more recent and extensive survey of equestrians in GB and Northern Ireland found that out of 6,390 equestrians participating, 67.7% reported having a near-miss and 6.1% an injury incident in the previous year (injury sustained by them and/or their horse) (Pollard & Furtado, 2021). At a more regional level, out of 1,976 equestrians surveyed in Devon, a South West county of England, 79.1% reported having a near-miss, 15.6% experienced a collision, and 7.7% sustained personal injury (Trump & Parkin, 2020). Real or perceived risk presented by high volumes of traffic and/or fast-moving vehicles has been identified as a barrier to walking and cycling (Anciaes, Stockton, Ortegon, & Scholes, 2019; Jacobsen, Racioppi, & Rutter, 2009; Sanders, 2015). Similar perceptions have been identified in equestrians with road use centering around an individualized assessment of risk comprised of the actions of other road users, the non-inclusive characteristics of the road network, the relationship with the individual horse, and the equestrian's own emotional management (Pollard & Furtado, 2021; Simsekoglu, Dalland, & Robertsen, 2020). The incident-causing actions of vehicle drivers around horses have been associated with differences in hazard perception, leading to an underestimation of risk when approaching a horse on a road, or frustration at encountering a slow-moving road user (Chapman & Musselwhite, 2011).

Similar equestrian road safety concerns have been reported by equestrians in Australia and Norway (Simsekoglu et al., 2020; Thompson & Matthews, 2015). However, published data regarding equestrian road safety from other countries are currently lacking, particularly regarding road incidents in countries where equids (horses, donkeys, and mules) are used extensively as working animals for the transportation of people and goods. Despite the negative experiences of a considerable proportion of equestrians when using the road network, equestrian road safety figures are generally under-represented by road safety stakeholder reports and government communications in GB. For example, The Department for Transport (Dft) Reported Road Casualties Great Britain Annual Report 2019 (Reported road casualties Great Britain, 2019) does not represent equestrians alongside pedestrians, cyclists, and motorcyclists. This may be because equestrians are considered to represent only a minimal proportion of road users in GB. While this may be true in relation to motorized vehicles, it is at odds with the existing evidence as to the number of people in GB that take part in equestrian activities, the regularity with which the equestrian population use roads with their horses, and the frequency of incidents they experience. Additionally, rate-based equestrian casualty data are not currently available, making it challenging to determine how the incidence rates of equestrian casualties are changing over time.

Equestrian road incident data in GB are collated independently by two organizations, the DfT and the British Horse Society (BHS). The DfT, together with the Scottish and Welsh Governments, has been collating road safety data for road collisions resulting in personal injury in Great Britain (GB) since 1979. These data represent injury collisions reported to British police forces. The records are made freely available and are often used by local authorities, policy-makers, and road safety stakeholders. It is possible to extract road incidents involving ridden horses from these data. The British Horse Society (BHS) has been collating horse-related road incidents via the Horse Incidents website since 2010, including both injury and non-injury incidents (Report Your Horse Incident, 2020). Incidents are reported by the public and submitted by anyone directly involved in, or having witnessed, an incident involving a ridden or non-ridden horse on a public road. A description and analysis of BHS road incident data has been published elsewhere (Pollard & Grewar, 2020), including an analysis of factors associated with collision risk and fatality outcome for the main horse involved in the incident. Close passing by vehicle drivers was one of the most significant contributors to collision risk between a vehicle and a horse and/or their handler, while collisions and speeding were significantly more likely to result in a horse fatality. Horse and human injuries were related; equestrians were 12 times as likely to be severely or fatally injured in incidents that resulted in horse fatality.

It is well recognized that road incidents are often under-reported, particularly if they do not result in serious or fatal human injury (James, 1991; Murphy et al., 2020). For example; National Health Service [NHS] Hospital Episode Statistics document that between April 2019 and April 2020 in England alone, 3,298 people were admitted to hospital due to an animal-rider or animal-drawn vehicle transport incident (Hospital Admitted Patient Care Activity 2019-20, 2020). In contrast; DfT road safety data for 2019 record only 124 people to have been involved in a road incident involving a ridden horse across England, Scotland, and Wales (Road Safety Data - data.gov.uk, 2020). Additionally; under-reporting may be more prevalent in certain road user groups due to, for example, differing perceptions of injury severity. British horse racing staff were less likely to report or take time-off for "invisible" injures, such as concussions and musculoskeletal injuries in comparison to fractures (Davies, McConn-Palfreyman, Parker, Cameron, & Williams, 2022); while anecdotal experiences from a major trauma surgeon in GB suggest that equestrians do not tend to seek medical treatment unless "something is hanging off" or that treatment of relatively serious injuries is often delayed (Research shines linght on equestrian-related injuries, 2022). In a large survey of equestrians in GB and Northern Ireland about their road use habits and experiences; participants were asked to comment on what would make them more or less likely to report a horse-related road incident – feelings that their reporting would not make a difference to their individual case or to equestrians in their area often dissuaded them from reporting an incident (Pollard & Furtado, 2021). If this under-reporting is not recognized it can lead to underestimation of road safety problems and lack of prioritization for policy changes and funding to improve road safety. It is also important to recognize who is most likely to be seriously injured or killed in interactions between different road user groups (Webster & Davies, 2020).

Investigating the data that are currently available regarding police-recorded road incidents involving ridden horses will provide a better understanding of equestrian incidents on the road network and help focus future equestrian road safety policy and research. Additionally, it will help to highlight the gaps and limitations of the DfT road safety database in light of more recent publications in the field of equestrian road safety. The aims of this study are to describe and investigate data on human injuries resulting from interactions between ridden horses and other road users on public roads in GB. The specific objectives are to:

(i) describe road incidents involving ridden horses and resulting in human injury as reported by police to the DfT database between 2010 and 2019
(ii) use the DfT data to identify factors associated with higher odds of serious to fatal injury sustained by road users involved in incidents with ridden horses
(iii) provide recommendations on how equestrian road incident data could be made more robust in order to better inform evidence-based road safety initiatives

2. Materials and methods

2.1. The DfT road safety data

The DfT road safety data for road incidents resulting in personal injury include the location, circumstances, types of vehicles involved, and resultant casualties. Casualties in this instance are defined as any person sustaining personal injury as a direct result of the road incident. These data represent injury-causing incidents reported to police forces across GB and recorded by police to the DfT road safety database using the STATS 19 reporting form.[1] Instructions on how the STATS 19 forms should be completed are also available, which provide definitions of the data to be collected.[2] These data are made publicly available via three separate but related coded datasets representing individual incidents, the vehicles involved and the resultant casualties, including a variable lookup data guide[3] for the codes. The incident index field provides a unique identifier for each incident and links vehicles and casualties to each incident, while the vehicle reference field links casualties to each vehicle.

All incident, vehicle, and casualty datasets between 2010 and 2019 were screened to identify incidents that involved ridden horses. All data including ridden horses or horse riders as casualties were extracted alongside data of other road users and casualties involved in the incidents. The unique identifier fields were used to merge all data into a single dataset for statistical analysis. A limitation of these data are that, for equestrians, only data on ridden horses are available; data including horse-drawn vehicles and horses being handled in another way on public roads, although collected, cannot be extracted as it is part of the 'other vehicles' category.

Severity of injury to road users involved in horse-related road incidents, as defined in the instructions on how the STATS 19 forms should be completed,[4] was categorized as:

1. *Fatal* – death occurring within 30 days of the incident and as a direct result of the incident.
2. *Serious* – examples include broken neck or back, severe head injury, severe chest injury, any difficulty breathing, internal injuries, multiple severe injuries, loss of arm or leg (or part), deep penetrating wound, fracture, deep cuts/lacerations, head injury, crushing, burns (excluding friction burns), concussion, loss of consciousness, severe general shock requiring hospital treatment, detention in hospital as an in-patient, either immediately or later, and death occurring more than 30 days after the incident but as a direct result of the incident.
3. *Slight* – examples include whiplash or neck pain, shallow cuts/lacerations/abrasions, sprains and strains (not necessarily requiring medical treatment), bruising and slight shock requiring roadside attention.
4. *None* – shaken but no other injury or medical treatment received.

In order to investigate factors associated with severe to fatal injury outcomes among the reported horse-related road incidents, a binary severity of injury variable was generated with a no injury to slight injury (0) category and a severe to fatal injury (1) category.

2.2. Data analysis

The DfT road-related incident data were stored in Microsoft Excel (Office 365) spreadsheets and imported into Stata (IC v. 13.0) statistical software for statistical analysis. Ordinal variables were summarized as medians with an interquartile range (IQR) and range (minimum to maximum), and categorical variables were described as proportions (%). The spatial distribution of road casualties with serious or fatal injuries that occurred between 2010 and 2019 were mapped in QGIS 3.10 (http://qgis.org) using reported latitude and longitude coordinates to visually describe the regional location of the most serious incidents.

Initial relationships between variables of interest (e.g., injury severity and whether police attended the incident or not) were assessed using the Chi-square (χ^2) test or Fisher's exact test for categorical data and the Mann-Whitney *U* test for continuous/ordinal data. The significance level was set as $p < 0.05$.

The data were analyzed at road user level with each observation representing a road user and their injury outcome. Univariable mixed-effects logistic regression modeling was used to identify factors associated with higher risk of severe to fatal injury outcome as a direct result of the incident. Risk was presented by calculating odds ratios (OR) and corresponding 95% confidence intervals (CI). As there was a lack of independence between observations for road users involved in the same incident, the statistical model had to be adjusted to take into account similarities within these incidents (e.g., occurring at the same time, on the same day, and on the same road). The statistical model was additionally adjusted for the year the incident took place to adjust for any similarities between incidents occurring in the same year. This was done by including year as a random effect in the model and the incident index (the unique identifier for each incident) as a random effect nested within year. Following univariable analyses, variables where the likelihood ratio statistic (LRS) $p < 0.25$ were selected for multivariable modeling (Dohoo, Martin, & Stryhn, 2009). The final multivariable mixed-effects logistic regression model was built using manual, stepwise, forward selection, with variables individually added into the model from most to least significant based on their Wald p-values and retained in the final model if model fit was significantly improved (Wald $p < 0.05$). Missing data remained missing and responses with missing data were automatically excluded from the analyses during model building, with the exception of some variables where a large number of records were missing (e.g., driver or rider age band). In these instances an "unknown" category was created. All variables that were excluded during the model building process were forced back into the final model individually at the end to assess any potential interaction or confounder effects.

3. Results

3.1. Description of police-recorded injury incidents involving ridden horses on public roads

Screening of the DfT road safety data between 2010 and 2019 revealed a total of 1,031 injury incidents involving ridden horses on the public road network were reported by police forces in GB. Tables 1 and 2 present a summary of the descriptive data obtained from the DfT road safety dataset at both incident (Table 1) and road user (Table 2) level.

The South East (19.4%), South West (16.5%), Yorkshire and The Humber (12.6%) and West Midlands (10.8%) regions of England had the highest frequency of reported incidents while Scotland (3.7%) and Greater London (2.1%) had the lowest frequency. These incidents involved a total of 2,243 road users. Of these 1,187 (52.9%) were injured; 1.5% (n = 18) fatally, 22.5% (n = 267) seriously and 76.0% (n = 902) slightly. The frequency of serious and fatal injuries, although distributed across GB, had a localised spatial distribution with higher frequencies around Greater London

1 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/230590/stats19.pdf.
2 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/230596/stats20-2011.pdf.
3 https://data.gov.uk/dataset/cb7ae6f0-4be6-4935–9277-47e5ce24a11f/road-safety-data.
4 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/230596/stats20-2011.pdf.

Table 1
A summary of injury road incidents (n = 1,031) involving ridden horses as recorded by police forces in Great Britain between 2010 and 2019.

Variable	Number of incidents	Percentage (%)
Incident year		
2010	126	12.2
2011	135	13.1
2012	127	12.3
2013	108	10.5
2014	115	11.2
2015	101	9.8
2016	103	10.0
2017	84	8.2
2018	73	7.1
2019	59	5.7
Incident region		
East of England	98	9.5
East Midlands	81	7.9
Greater London	22	2.1
North East	49	4.8
North West	83	8.1
Scotland	38	3.7
South East	200	19.4
South West	170	16.5
Wales	49	4.8
West Midlands	111	10.8
Yorkshire and The Humber	130	12.6
Incident month		
January	81	7.9
February	88	8.5
March	74	7.2
April	103	10.0
May	86	8.3
June	78	7.6
July	97	9.4
August	85	8.2
September	100	9.7
October	77	7.5
November	84	8.2
December	78	7.6
Incident day		
Sunday	178	17.3
Monday	118	11.5
Tuesday	156	15.1
Wednesday	142	13.8
Thursday	133	12.9
Friday	124	12.0
Saturday	180	17.5
Incident time of day		
09:00–12:00	318	30.8
12:00–14:00	202	19.6
14:00–17:00	295	28.6
17:00–20:00	156	15.1
20:00–00:00	15	1.5
00:00–09:00	45	4.4
Incident area		
Urban	154	14.9
Rural	877	85.1
Incident road type		
A	118	11.5
B	140	13.6
C	194	18.8
Unclassified	579	56.2
Speed limit of road (mph)		
20	8	0.8
30	435	42.2
40	84	8.2
50	27	2.6
60	474	46.0
70	3	0.3
Weather conditions		
Fine no high winds	928	90.0
Rain no high winds	41	4.0
Snow no high winds	1	0.1
Fine with high winds	19	1.8
Rain with high winds	5	0.5
Fog/mist	1	0.1
Other	10	1.0
Unknown	26	2.52
Road surface condition		
Dry	855	83.4
Wet/damp	167	16.3
Frost/ice	2	0.2
Flood > 3 cm	1	0.1
Attended by police		
Yes	554	53.7
No	477	46.3

and the South East, the South West, West Midlands and Yorkshire regions (Fig. 1).

Incidents were generally least frequent during the winter months (December to February; 19.6%, n = 790) and most frequent during the summer months (June to August; 29.2%, n = 1,177). Incidents were more frequent on Saturdays (17.5%) and Sundays (17.3%) and least frequent on Mondays (11.5%). Most incidents (30.8%) occurred in the morning between the hours of 09:00 and 12:00 and in the afternoon (28.6%) between the hours of 14:00 and 17:00 hours. More than half of incidents occurred on unclassified roads (56.2%) and in rural areas (85.1%) with either 30 mph (42.2%) or 60 mph (46.0%) speed limits. Most of the incidents occurred during fine weather conditions with no high winds (89.9%) and dry road surface conditions (83.4%). Although all incidents were recorded by the police, the police attended only 53.7% of the incidents.

A total of 1,129 horse riders were involved in the incidents; 83.5% (n = 943) were female, 13.7% (n = 155) were male, and 31 did not have gender recorded. Age category data were available for 1,050 horse riders; the largest proportion belonged to the 46–55 year (18.7%, n = 196) and 36–45 year (17.5%, n = 184) age categories while 26.3% (n = 276) were in the 0–20 year age category. A considerable proportion of interactions did not involve impact between the ridden horse and the vehicle (46.9%) but where impact did occur, most first points of impact were either on the offside of the horse (40.2%) or from the rear (28.6%).

A median of two vehicles and/or ridden horses were involved in each incident (IQR 2 to 2; range 1 to 18). Of the other road users (n = 1,114) involved in the incidents with ridden horses, most were car occupants (62.7%), van/light goods vehicle occupants (8.8%), and motorcycle riders/passengers (6.8%). Of the other road users, 56.4% were male, 20.3% were female, and 23.3% did not have a gender recorded. Age category data were available for 771 vehicle drivers/riders; the largest proportion belonged to the 46–55 year age category (18.6%).

Out of the 1,187 road users injured a median of 1 were injured per incident (range 1 to 5). The majority of road users injured were female (81.5%) and belonged to the 36–55 year (35.6%, n = 413/1,161) year age category while 25.2% (n = 293) were in the 0–20 year age category. Horse riders were the main road user group injured in these incidents (84.1%) and represented 238 out of 267 serious injuries and 17 out of the 18 fatalities (Fig. 2). Out

Table 2
A summary of data regarding road users (n = 2,243) involved in injury road incidents with ridden horses as recorded by police forces in Great Britain between 2010 and 2019.

Variable	Number of incidents	Percentage (%)
Vehicle type involved in incident with ridden horse (n = 1,114)		
Car	698	62.7
Van	98	8.8
Motorcycle	76	6.8
Heavy goods vehicle	74	6.6
Agricultural vehicle	63	5.7
Minibus/bus	33	3.0
Pedal cycle	28	2.5
Other/unknown	44	4.0
Vehicle occupant gender (n = 1,114)		
Female	226	20.3
Male	628	56.4
Not recorded	260	23.3
Vehicle driver/rider age category (n = 771)		
11–15	2	0.3
16–20	46	6.0
21–25	59	7.7
26–35	125	16.2
36–45	126	16.3
46–55	143	18.6
56–65	112	14.5
66–75	82	10.6
>75	76	9.9
Horse rider gender (n = 1,129)		
Female	943	83.5
Male	155	13.7
Not recorded	31	2.8
Horse rider age category in years (n = 1,050)		
0–5	2	0.2
6–10	22	2.1
11–15	93	8.9
16–20	159	15.1
21–25	125	11.9
26–35	153	14.6
36–45	184	17.5
46–55	196	18.7
56–65	86	8.2
66–75	26	2.5
>75	4	0.4
Impact between ridden horse and vehicle (n = 1,127)		
Yes	599	53.1
No	528	46.9
Location of first point of impact between vehicle and ridden horse (n = 599)		
Front	130	21.7
Rear	171	28.6
Offside	241	40.2
Nearside	57	9.5
Injuries in incidents involving ridden horses (n = 2,243)		
Yes	1,187	52.9
No	1,056	47.1
Injured road users in incidents involving ridden horses (n = 1,187)		
Horse rider	998	84.1
Car occupant	105	8.9
Pedestrian	37	3.1
Cyclist	24	2.0
Motorcyclist	15	1.3
Other vehicle	8	0.7
Gender of injured road users (n = 1,187)		
Female	967	81.5
Male	220	18.5
Age category of injured road users (n = 1,161)		
0–5	7	0.6
6–10	19	1.6
11–15	94	8.1
16–20	173	14.9
21–25	128	11.0
26–35	168	14.5
36–45	199	17.1
46–55	214	18.4
56–65	98	8.4
66–75	39	3.4
>75	22	1.9
Injury severity (n = 1,187)		
Slight	902	76.0
Serious	267	22.5
Fatal	18	1.5

of the 120 road users aged 15 years and younger that were injured, 109 (90.8%) were horse riders. Police were more likely to attend incidents that resulted in severe to fatal injury ($p < 0.001$, χ^2 22.2).

Twenty-seven incidents where the horse rider was seriously or fatally injured included multiple ridden horses; 26 involved two ridden horses and one incident involved four ridden horses. Vehicle types involved in incidents where horse riders were either seriously or fatally injured were mostly cars (53.4%, n = 141) and vans/light goods vehicles (9.8%, n = 26) (Fig. 3). A considerable proportion (12.5%, n = 33) of horse rider serious injuries and fatalities were reported to not have included another vehicle. As specific details of incidents were not available, it is not possible to speculate as to how the injuries to horse riders occurred where no other vehicles were involved.

3.2. Factors associated with severe to fatal injury outcomes

A total of 2,243 observations were available for mixed-effects logistic regression modeling. There were 10 year categories (2010 to 2019) with a minimum of 120 and a maximum of 293 observations per year within which were nested 1,031 incidents with a minimum of 1 and a maximum of 19 observations per incident – equating to the number of road users involved in each incident.

Univariable mixed-effects logistic regression results are presented in Table S1. Multivariable modeling identified six variables significantly associated with severe to fatal injury, after adjustment for injury year and incident (Table 3). There was evidence of a lack of independence of observations across the calendar years and at incident level (i.e., across observations of road users involved in the same incident and in incidents occurring in the same year; LRS = 0.0284). The severity of injury was associated with mode of transport, first point of impact, the speed limit of the road, the month of the year, the region, and the age of the driver or rider involved in the incident. Compared to people traveling in cars, the odds of severe to fatal injury in incidents involving ridden horses were higher for cyclists (OR 108.5, 95% CI 24.1, 487.2), horse riders (OR 73.4, 95% CI 26.1, 206.2), and motorcyclists (OR 18.2, 95% CI 4.4, 75.6). Incidents where the first point of impact was from the front (OR 2.5, 95% CI 1.3, 4.7) or the rear (OR 2.0, 95% CI 1.1, 3.6) were more likely to result in severe or fatal injury compared to when the first point of impact was from the offside. Non-impact incidents were similarly associated with higher odds of severe or fatal injury (OR 1.7, 95% CI 1.1, 2.9). The odds of serious or fatal injury almost doubled (OR 1.6, 95% CI 1.1, 2.4) for roads with speed limits of between 60 and 70 mph compared to roads with speed limits of between 20 and 30 mph. Injury severity odds were higher between January to February (OR2.3, 95% CI 1.2, 4.4) and May to June (OR 2.8, 95% CI 1.5, 5.3) compared to between March to April. Regionally, serious and fatal injuries were more

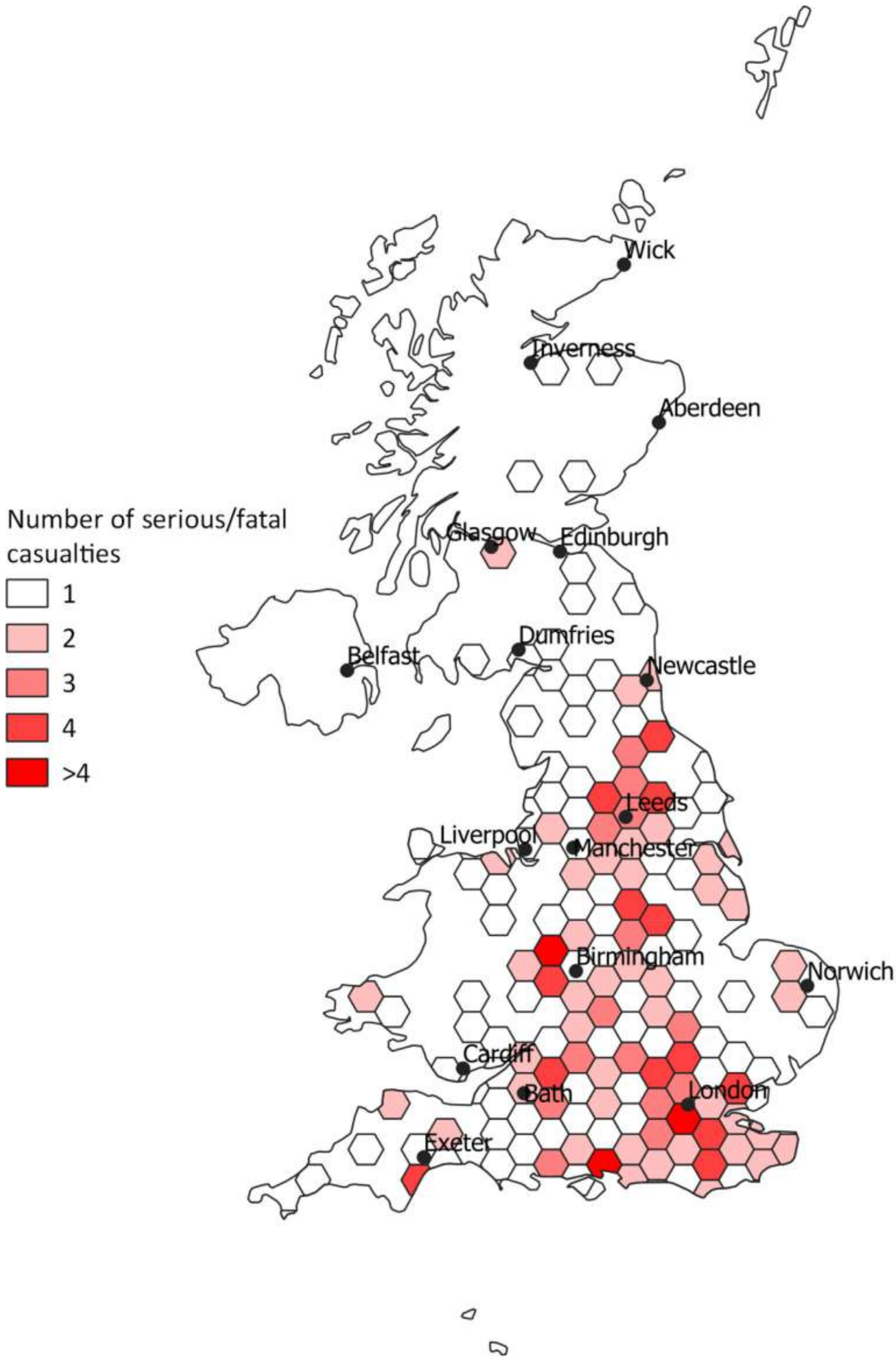


Fig. 1. The frequency of serious and fatal casualties resulting from road incidents involving ridden horses (n = 1,031) on public roads in Great Britain between 2010 and 2019 as recorded by police forces.

likely in the South East (OR 1.6, 95% CI 1.5, 7.5), West Midlands (OR 2.6, 95% CI 1.1, 6.1) and Yorkshire (OR 2.5, 95% CI 1.1, 5.9) compared to the North West. Lastly, the odds of serious or fatal injury increased with increasing age band of the driver or rider involved in the incident with the odds of road users over 66 years of age sustaining serious or fatal injuries nearly 9-fold higher ($p < 0.001$) than that of the youngest age group (0–15 years).

4. Discussion

The analysis of road incidents and their contributory factors is vital to help prevent future injuries and fatalities on the road network. We have previously described in detail and analyzed horse-related road incidents reported to the BHS and occurring between 2010 and 2020 in England, Scotland, Wales, and Northern Ireland,

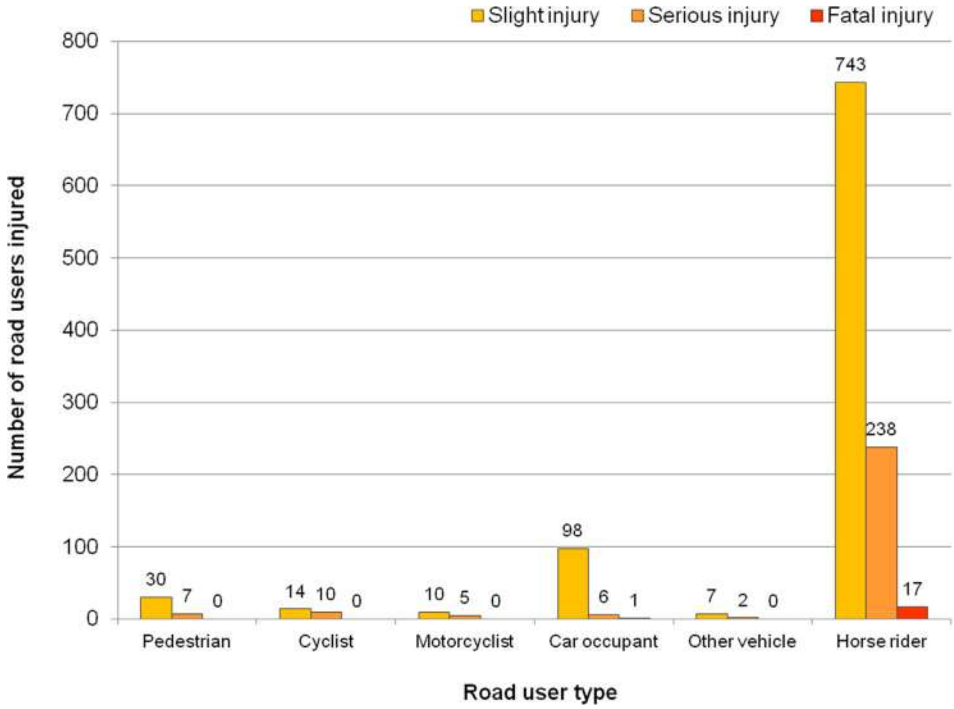


Fig. 2. The number and type of road users injured and the severity of injury sustained in police-reported incidents involving ridden horses (n = 1,031) on public roads in Great Britain between 2010 and 2019.

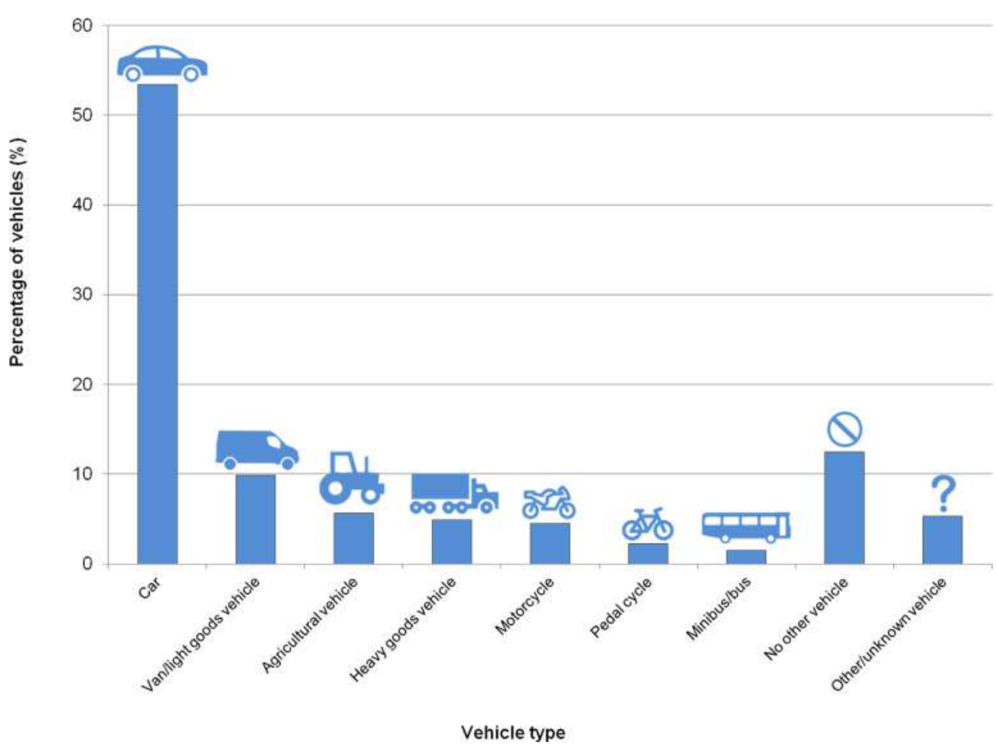


Fig. 3. The types of vehicles (n = 264) involved in police-reported incidents resulting in severe or fatal injury to a horse rider on public roads in Great Britain between 2010 and 2019.

which provided important insight into collision and horse fatality risk (Pollard & Grewar, 2020). As evidence-based data regarding injuries sustained by road users involved in incidents with ridden horses were lacking, we set out to describe the type of road incidents recorded by police forces and the impact on the road users involved. We also present recommendations on how the robustness of equestrian road incident data can be improved to better feed into the current knowledge on the frequency, location, and circumstances surrounding road incidents involving horses, and how best to prevent them.

Table 3
Multivariable mixed-effects logistic regression modeling, with incident year and id as random effects, of factors associated with higher odds of severe to fatal injury to road users (n = 2,239) involved in incidents including ridden horses (n = 1,031) and reported by police to the Department for Transport road safety database between 2010 and 2019.

Variable	Coefficient	Standard error	Odds ratio (OR)	95% confidence interval (OR)	Wald P-value
Mode of transport					
Car	Reference		1.0		
Pedal cycle	4.7	0.8	113.5	24.8, 518.6	<0.001
Motorcycle	2.9	0.7	18.5	4.4, 77.6	<0.001
Ridden horse	4.3	0.5	75.7	26.8, 214.1	<0.001
Other/unknown*	−0.7	1.1	0.5	0.1, 4.4	0.535
First point of impact					
Offside	Reference		1.0		
Nearside	0.2	0.4	1.2	0.5, 2.8	0.721
Back	0.7	0.3	2.0	1.1, 3.7	0.030
Front	0.9	0.3	2.6	1.3, 4.8	0.004
No impact	0.6	0.4	1.8	1.1, 2.9	0.025
Speed limit of road (miles per hour)					
20–30	Reference		1.0		
40–50	0.5	0.3	1.7	0.9, 3.1	0.084
60–70	0.5	0.2	1.7	1.1, 2.5	0.014
Month					
January-February	0.9	0.3	2.4	1.2, 4.6	0.010
March-April	Reference		1.0		
May-June	1.1	0.3	2.9	1.5, 5.6	0.002
July-August	0.4	0.3	1.4	0.8, 2.7	0.273
September-October	0.4	0.3	1.5	0.8, 2.8	0.259
November-December	0.4	0.3	1.5	0.7, 2.8	0.266
Region					
East of England	0.6	0.5	1.8	0.7, 4.5	0.214
East Midlands	0.8	0.5	2.1	0.8, 5.6	0.124
Greater London	0.7	0.8	2.0	0.4, 9.6	0.362
North East	0.5	0.6	1.6	0.5, 4.9	0.398
North West	Reference		1.0		
Scotland	0.9	0.6	2.6	0.8, 8.3	0.110
South East	1.2	0.4	3.4	1.5, 7.8	0.004
South West	0.5	0.4	1.6	0.7, 3.8	0.252
Wales	0.6	0.5	1.8	0.6, 5.3	0.268
West Midlands	1.0	0.5	2.6	1.1, 6.3	0.036
Yorkshire and the Humber	0.9	0.4	2.5	1.1, 6.0	0.036
Driver/rider age band (years)					
0–15	Reference		1.0		
16–20	0.4	0.4	1.4	0.6, 3.2	0.392
21–25	0.1	0.4	1.1	0.4, 2.6	0.879
26–35	0.8	0.4	2.3	1.1, 5.1	0.037
36–45	1.0	0.4	2.7	1.3, 5.9	0.011
46–55	1.0	0.4	2.7	1.2, 5.6	0.012
56–65	1.3	0.4	3.8	1.6, 9.0	0.003
>66	2.2	0.5	8.8	3.2, 24.4	<0.001
Unknown	−1.2	0.6	0.3	0.1, 1.0	0.046

* Other modes of transport include minibuses/buses, agricultural vehicles, vans, light and heavy goods vehicles.

4.1. Police-recorded injury incidents involving ridden horses on public roads

The spatial distribution of DfT road incidents corresponds with highest frequencies of serious or fatal outcomes with the South East, South West, West Midlands, and Yorkshire areas of England having the highest proportions of incidents as well as those with the most serious outcomes. This was partially reflected in the results of the multivariable regression modeling where the South East, West Midlands, and Yorkshire had approximately-three times the odds of severe to fatal injury outcomes for road users (most of whom were horse riders) involved in incidents including ridden horses in comparison to the North West. Similarly, the South East was found to have significantly higher odds of collision incidents between vehicles and equestrians compared to the North West when the BHS horse incidents data were analyzed (Pollard & Grewar, 2020). Equestrians in the West Midlands and Yorkshire regions of England were found to be more likely to use roads and had higher odds of a road-related near-miss in the previous year compared to equestrians in Scotland (Pollard & Furtado, 2021). Corroboration of data for certain regions being more risky for equestrians than others can prompt further investigation into the location of regional incidents, such as the characteristics of the road and road user behavior, and lead to interventions to improve equestrian road safety. Additionally, this information could help secure funding for more sophisticated rate-based regional estimates of risk, based on the distance and/or time equestrians spent on roads and the density of equestrians in the region.

The frequency of incidents reported coincided with the days (weekend days), times (predominantly mornings and afternoons), seasons (summer months) and weather conditions (fine, dry conditions) when equestrians report they are more likely to be riding their horses and, therefore, accessing roads (Trump & Parkin, 2020). However, incident month was the only time variable retained in the final multivariable model with the odds of having a severe or fatal injury outcome higher in late winter and early summer compared to the spring. This finding suggests that winter months have a disproportionally higher frequency of incidents with a serious outcome compared to spring months when the weather starts improving. This is similar to trends seen in cycling road incidents and injuries in the UK; cycling is more common during the spring and summer months but the casualty rate per mile traveled is higher in autumn and winter months (Reported road accidents, 2021). This is likely due to a complex combination of factors, such as adverse road surface conditions due to rain and ice/snow, poorer visibility (shorter daylight hours, lower light or low sun in the mornings or evenings) and strong winds (Pazdan, 2020). Location wise, most incidents occurred in rural areas and on minor unclassified roads. By definition, unclassified roads are local roads intended for local traffic that should be used by smaller amounts of traffic traveling at lower speeds over shorter distances (Guidance on road classification and the primary route network, 2021). The 2019 annual report for road casualties in Great Britain showed that the majority of road fatalities (57%) occurred on rural roads (Reported road casualties Great Britain, 2019). These include narrow, single-track roads with areas of poor visibility, due to overhanging vegetation or high hedges, and often speed limits of 60 mph; speed limits often not suitable for the road conditions, types of road users using the roads, and surroundings.

In our multivariable model, however, the speed limit of the road was a much more important determinant of injury severity than a rural, urban, or suburban location. Odds of severe or fatal injury almost doubled on roads with speed limits of 60–70 mph compared to roads with speed limits of 20–30 mph. It is well-established that speed is one of the main determinants of collision risk and collision severity (Aarts & Van Schagen, 2006; Richter, Berman, Friedman, & Ben-David, 2006). While the speed limit of a road may not necessarily equal higher speeds, it has been shown to be an important proxy measurement; raising speed limits contributes to increased road fatalities while lowering speed limits has the opposite effect. As an example, a Swedish study evaluating insurance data between 2005 and 2017 found that a reduction in speed limits (from 50-60 km/h to 30–40 km/h) was associated with lower risk of moderate to fatal injury for cyclists involved in collisions with cars (Isaksson-Hellman & Töreki, 2019), while research in the United States found an increase in road fatalities between 1995 and 2005, which could be attributed to raised speed limits on all road types in the United States, with the highest fatality increase on rural interstates (Friedman, Hedeker, & Richter, 2009). Therefore, there is substantial evidence that supports our conclusions that reduction of speed limits alongside improved enforcement, particularly on rural roads, would lead to fewer and less severe road incidents.

eAlmost half of all injury incidents involving ridden horses did not result in physical contact or impact between the ridden horse and the vehicle, which is in keeping with previously published research using both BHS and survey data from GB and Northern Ireland (Pollard & Grewar, 2020; Trump & Parkin, 2020). Where impact did occur, it most commonly occurred on the offside or rear of the horse, however, rear and front (head-on) impacts were more likely to result in severe or fatal injury compared to offside impacts. Interestingly, the odds of severe or fatal injury were also almost doubled for non impact incidents compared to offside impacts. This is important to understand; lack of a collision in equestrian incidents does not equate to lack of injury even when another road user is involved. This invites the concept of three minds at work when, for example, a vehicle driver encounters a ridden horse on the road; that of the vehicle driver, the rider, and the horse. Horses are prey animals and their reactions to a perceived hazard (which could be a loud, noisy trailer, a fast-moving vehicle, or a silent bicycle) most commonly set off the 'flight' response where the horse will attempt to escape from the danger but sometimes the 'fight' response will be activated, where a horse may kick out at the hazard (Keaveney, 2008; Norwood et al., 2000; Thompson, McGreevy, & McManus, 2015). This may not only be dangerous for the horse-rider combination, but also for other road users. Helping non-equestrians understand simple horse behavior would help put the risk faced by equestrians on roads into perspective.

Similar to cyclists (Aldred & Crosweller, 2015), fear of injury or witnessing or being involved in a road incident are important contributors to equestrians' avoidance of roads (Pollard & Furtado, 2021) and, as such, may be barriers to the uptake of equestrian activities. This also presents a problem in terms of the 'Safety in numbers' effect, which has been demonstrated to exist for pedestrians and cyclists; it is used to explain the inverse statistical relationship between the number of pedestrians and cyclists in a population and the number of injuries they sustain due to road incidents (Elvik & Bjørnskau, 2017; Fyhri, Sundfør, Bjørnskau, & Laureshyn, 2017; Jacobsen, 2003). Although the exact mechanisms behind this effect have not yet been elucidated, the most common theories include the concept of motorists becoming more attentive when exposed to higher numbers of pedestrians and cyclists (Jacobsen, 2003), that road users that gain more experience and become more familiar with other road users develop better expectations of behaviors (Phillips, Bjørnskau, Hagman, & Sagberg, 2011), or that the demands of larger populations of cyclists and pedestrians drive safer transport infrastructure, norms, and behaviors (Bhatia & Wier, 2011). Whether one or several of these proposed mechanisms play a part in increasing safety of pedestrians and cyclists, it is likely a similar Safety in Numbers effect exists for equestrians and avoidance of roads by equestrians could lead to a downward spiral in terms of equestrian road safety. Equestrians have previously described feeling increasingly less safe on roads (Pollard & Furtado, 2021), leading to fewer equestrians using roads and in turn contributing to higher incident risk for those equestrians still using roads.

This leads onto who is most likely to be involved in road incidents with ridden horses and who is most likely to be injured. Over 80% of road users injured in these incidents were female, 36% were aged between 36 and 55 years of age, while 25% were aged between 0 and 20 years of age, and 84% were horse riders. Perhaps of particular concern is that out of 120 injured road users aged 15 years and younger, over 90% were horse riders. Our previous research also identified that younger equestrians were more likely to use roads but also to report experiencing a near-miss incident on roads in the previous year compared to older equestrians, while equestrians riding while leading another ridden horse (often a child on a pony) were more likely to have experienced an injury incident on the road in the previous year (Pollard & Furtado, 2021). These findings are at odds with the vision of Sport England's "Uniting the Movement" strategy in GB, which aims to encourage sport and physical activity uptake in communities, particularly by involving more women and young people (Uniting the Movement, 2021). If these groups are the ones having more negative experiences and being injured while doing equestrian activities on roads, it is likely to dissuade them from participating in activities related to equestrianism, which for some may be the only physical activity they are involved in (Church, Taylor, Maxwell, Gibson, & Twomey, 2020). Conversely, physical activity and mental stimulation are of equal importance to older age categories. Our multivariable injury outcome model identified that odds of severe or fatal injury increased with increasing age of the road user involved with odds of severe or fatal injury almost nine times as likely in the over 66 year old category compared to the 15 years and younger category. Improving equestrian road safety will therefore help to safeguard our younger and older generations.

When further taking into account severity of injury, it is almost always the horse rider, a cyclist, or motorcyclists, rather than a person enclosed in a vehicle, that are severely or fatally injured in road incidents involving ridden horses. This also highlights that other road users, such as cyclists and motorcyclists, should be made aware to take particular care when interacting with horses on the road to additionally reduce any injury risk to themselves. For leisure and sport cyclists this could be done via cycling clubs and groups, while for commuter cyclists this could be done via cycle to work schemes or other environmental and sustainability frameworks adopted by employers. The vehicle types involved in the highest proportion of severe or fatal injuries to horse riders were cars and vans/light goods vehicles. While these are the road user groups that are also likely to frequently encounter ridden horses on roads, this finding is something that should be investigated further.

4.2. Recommendations on improving the robustness of equestrian road incident data

The DfT data have several limitations when representing equestrian road incidents, which are discussed below and also comprise the limitations of this study. We provide some recommendations on how the robustness of equestrian road incident data could be improved.

4.2.1. The type of data collected

The DfT data represent only a proportion of actual road incidents as recorded by police forces. Recording, therefore, relies on the incident being reported to the police and is subject to considerable under-reporting for nonfatal injury incidents. The only type of equestrian activity recorded is ridden activity and while equestrians often ride their horses on roads, the current data do not acknowledge other types of equestrian activities that also regularly occur on roads, such as leading a horse on foot and horse-drawn vehicles (Pollard & Furtado, 2021). We recommend that all equestrian activities are included as identifiable categories in the dataset to better understand overall equestrian injury risk. Another consideration is inclusion of incidents where a loose horse (absence of rider/handler) is involved in a road incident with another road user. This can happen in several ways, including horses not being secured properly or escaping onto the road network due to damaged fencing or gates being left open, a horse that is running loose on the road because the rider has become unseated, or vehicles encountering semi-feral free-roaming ponies in large conservation areas such as the New Forest and Dartmoor. In fact, collisions with a loose horse are one of the most common causes of road-related horse fatality in Britain (Pollard & Grewar, 2020) and additionally pose a high risk of injury for the vehicle occupant. Enabling the extraction of these data would provide important information on human injury risk as a consequence of being involved in a collision with a large animal.

The DfT data contain only incidents resulting in personal injury to one or more road users. While this is vital information, collation of data regarding near-misses (involving close passes or speeding) or incidents causing distress (aggressive or intimidating behavior aimed at the rider/handler or the horse) to the equestrian contain vital early warning indicators of specific road environments or road use behaviors that could escalate to injury (Aldred, 2016). According to Dee et al. (Dee, Cox, & Ogle, 2013)there is often a misinterpretation of a near-miss as being a high-probability, low-consequence severity occurrence rather than a narrowly averted low-probability, high-consequence severity occurrence. Proper investigation, evaluation, and intervention following a near-miss often prevents the occurrence of a more severe incident. Not only that, but experience of near-misses by cyclists can have a considerable impact on risk perception and future participation in cycling (Aldred & Crosweller, 2015; Sanders, 2015). Self-reported near-misses by cyclists were shown to reasonably accurately represent actual events with many driver behaviors reportedly similar between near-misses and police-recorded slight injury collisions involving cyclists (Aldred, 2016). However, these near-miss data need to be accepted and utilized by road safety policy makers and authorities in order to be adopted into intervention-based strategies. Being aware of the frequency and location of these non-injury incidents could help police forces and local authorities work with their local equestrian communities to improve road safety. We suggest ways in which this could be done in the following subsection.

The current data analyzed were retrospectively collected and although recorded in a standardized way, are still liable to subjective interpretation of events. The incidents also lacked context whether provided by eye-witness accounts or video footage, which would have been useful to explain how some of the injury incidents occurred. For example, incidents with ridden horses where no other road users were involved – is this because the other road user could not be identified or because it truly was an incident not involving another road user? While this may not be available in public records of police-recorded data, it is an important consideration for any future equestrian road safety research.

Equestrians are currently not represented in any rate-based casualty estimates (based on casualties per mile traveled), which indicate that pedestrians, cyclists, and motorcyclists, although having lower numbers of casualties compared to vehicle drivers, have high rates of casualties when the miles traveled are taken into account. It is likely that equestrians have similarly elevated casualty rates if the distance or time spent using roads is taken into consideration. Therefore, studies designed to determine rate-based incident or casualty estimates would better represent the actual risk experienced by equestrians when using the public road network.

4.2.2. Combining multiple sources of data

In an ideal world, a central database would exist linking multiple sources of data (e.g., incident circumstances, road characteristics, road user demographics, healthcare data regarding injury outcomes and objective video footage if available). This would ensure that all data and outcomes of interest were reported and coded in the same way.

Information on equestrian road incidents is currently fragmented. There are two databases which collate, store, and make available equestrian road incident data to the public: the Dft data, as described here, and the BHS Horse Incidents database described in a previous study (Pollard & Grewar, 2020). The BHS data contain information on incidents across all equestrian activities on roads in GB and Northern Ireland, injury outcomes for both the main rider/handler and horse involved in the incident, as well as data on near-miss incidents (including road rage and aggressive behavior directed at the rider/handler and or horse). These databases are independent although there is likely considerable overlap between injury incidents reported to both the police and the BHS as the BHS actively advocates the reporting of both injury and non-injury incidents to the police. However, this does mean that equestrians have to go through two separate reporting processes for a single incident, potentially contributing to lower reporting of frequent road incidents. As discussed above, the importance of presenting near-miss data, where available, should not be underestimated. We suggest that road safety stakeholders consider both police-recorded injury data alongside BHS near-miss data when assessing equestrian road safety in their local areas.

Some police forces have created self-reporting online portals that have streamlined the collection of video and photographic evidence related to driving offenses from members of the public using dashboard, body, or helmet camera footage (e.g., Operation SNAP (GoSafe - Op snap, 2022; Devon and Cornwall Police Operation Snap digital submissions, 2022; Operation Snap | Warwickshire Police, 2022; Operation Snap | West Mercia Police, 2002). Equestrians can submit evidence regarding injury and near-miss incidents to this portal, however, these data are not publicly available nor is it currently possible to obtain equestrian-specific non-injury data from police forces. Submitting of this evidence also relies on equestrians wearing suitable video recording equipment and being able to capture the moment in time from the right perspective.

Finally, linking road incident data to both medical and veterinary healthcare data could provide a better understanding of injuries, hospitalization period, and the financial implication of equestrian road incidents. Although summary-level data for NHS Hospital Episode Statistics are publicly available (Hospital Episode Statistics (HES), 2022), special permissions have to be obtained to access the full NHS records to provide meaningful insight into the healthcare impact of equestrian road incidents. Similarly, screening of large-scale veterinary hospital records from practices that use the same practice management software could provide important horse injury data (Welsh, Duz, Parkin, & Marshall, 2016).

4.2.3. The importance of language

Encouraging the reporting of road-related equestrian incidents will help create a more complete picture of what is happening on the road network. However, the language we use around road safety can be problematic when it comes to describing equestrian incidents in relation to other road incidents. While horses are used as modes of transport, they are not vehicles. Classing horses as vehicles fails to acknowledge their role as autonomous road users in their own right as well as failing to recognize the complex bond that most equestrians have with their horses; often perceiving them as valued friends, companions or family members (Dashper, 2017; Lee Davis, Maurstad, & Dean, 2015; McGowan, Phillips, Hodgson, Perkins, & McGowan, 2012). We recommend that reference to vehicles in the DfT database is changed to "modes of transport," which would be a better representation of how roads are used by all road users. Additionally, collating data on whether the horses involved were injured in the incidents would elevate their role on the road from being a vehicle to a special type of road user.

Avoidance of the word "accident" is well established in current road safety culture, with the word accident implying a lack of attributable blame or a sense of inevitability; but, it is well-established that most injury events on roads are largely predictable and preventable (Davis & Pless, 2001; Stewart & Lord, 2002). Terminology such as "crash" or "collision" have, therefore, been widely favored by road safety professionals and academics. However, this can be problematic for describing horse incidents because, as shown by the current data and elsewhere, a considerable proportion of horse incidents (even those resulting in injury and involving other road users) do not involve physical contact between another road user and the horse or their rider/handler (Pollard & Grewar, 2020; Trump & Parkin, 2020). In cycling, the terms 'non-collision incident' and 'single cyclist collision' are used to describe incidents where no other road users were involved and the cyclist injured themselves by, for example, slipping on an icy road (Gildea, Hall, & Simms, 2021). We propose that use of the word road incidents should be considered by road safety stakeholders to encompass the experiences of all road users and include the analysis of near-miss incidents. These can subsequently be broken down into collision or non-collision incidents and injury or non-injury incidents as required.

5. Conclusions

Despite the considerable number of people involved in equestrian pursuits and the frequency with which they use roads with their horses, horse riders and handlers in GB often feel like low-priority road users and equestrian activities are seldom promoted as part of government active travel or green exercise initiatives. In road incidents involving ridden horses, it is almost always the horse rider that is injured. The horse riders injured in these incidents are largely women and just over a quarter are young adults or children. These findings are at odds with the aims of several national campaigns that seek to increase participation in sport, physical activity, and access to green spaces, particularly for women and young people.

When taking into consideration injury severity in incidents involving ridden horses, horse riders, cyclists and motorcyclists were more likely than car occupants to be severely or fatally injured, while odds of severe or fatal injury increased with increasing road user age. The vehicles associated with the most severe injury to horse riders were cars and light goods vehicles or vans. These findings warrant further investigation in order to assess opportunities to create a safer, more inclusive road network and promote positive interactions between road users. Our findings also support the evidence that reductions in speed limits on roads frequented by horse riders, which would be the majority of the rural and some non-rural road networks, would help reduce the risk of serious or fatal injuries.

6. Practical applications

Although the frequency of police-recorded road incidents involving ridden horses is low in relation to incidents involving other road users, the limitation of the DfT database should be kept in mind. These data present only a proportion of true injury incidents for one subset of equestrians. Additionally, rate-based casualty estimates (based on time spent on the road or distance traveled) for equestrians are currently lacking, making it difficult to fully understand the risk faced by equestrians while using roads and time changing trends. However, the BHS database presents a complementary and readily-available source of data that can be used to analyze patterns of near-miss incidents, which are extremely frequent, and pinpoint problem areas where action can be taken before serious or fatal incidents occur. Improving the robustness of equestrian road incident data would ensure that road safety stakeholders and police forces are made aware of the frequency and location of near-misses and injury incidents that occur between equestrians and other road users that would help them work within their local communities to improve road safety.

Funding

DP's position at the British Horse Society was funded by the Department for Transport.

Acknowledgments

To colleagues at the British Horse Society Safety Department (Emily Yates, Des Payne and Alan Hiscox). To Heather Ward from the University College London Centre for Transport Studies for advice and help with understanding the STATS 19 data. The Department for Transport for investing in equestrian road safety research and for making police-recorded road incident data freely available. To all the people that have submitted an incident report to the BHS and shared their experiences.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Supplementary material

S1: Univariable mixed-effects logistic regression modeling, with incident year and id as random effects, of factors associated with higher odds of severe to fatal injury to road users (n=2,239) involved in incidents including ridden horses (n=1,031) and reported by police to the Department for Transport road safety database between 2010 and 2019. Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsr.2022.10.010.

References

Aarts, L., & Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis and Prevention, 38*, 215–224. https://doi.org/10.1016/J.AAP.2005.07.004.

Aldred, R. (2016). Cycling near misses: Their frequency, impact, and prevention. *Transportation Research Part A: Policy and Practice, 90*, 69–83. https://doi.org/10.1016/J.TRA.2016.04.016.

Aldred, R., & Crosweller, S. (2015). Investigating the rates and impacts of near misses and related incidents among UK cyclists. *Journal of Transport & Health, 2*, 379–393. https://doi.org/10.1016/j.jth.2015.05.006.

Anciaes, P. R., Stockton, J., Ortegon, A., & Scholes, S. (2019). Perceptions of road traffic conditions along with their reported impacts on walking are associated with wellbeing. *Travel Behaviour and Society, 15*, 88–101. https://doi.org/10.1016/j.tbs.2019.01.006.

Bhatia, R., & Wier, M. (2011). "Safety in Numbers" re-examined: Can we make valid or practical inferences from available evidence? *Accident Analysis and Prevention, 43*, 235–240. https://doi.org/10.1016/J.AAP.2010.08.015.

British Equestrian Trade Association (BETA). (2019). *The National Equestrian Survey 2019*; Whetherby, West Yorkshire.

Chapman, C., & Musselwhite, C. B. A. (2011). Equine road user safety: Public attitudes, understandings and beliefs from a qualitative study in the United Kingdom. *Accident Analysis and Prevention, 43*, 2173–2181. https://doi.org/10.1016/j.aap.2011.06.009.

Church, A., Taylor, B., Maxwell, N., Gibson, O., & Twomey, R. (2020). The health benefits of horse riding in the UK Available online: https://research.brighton.ac.uk/en/publications/the-health-benefits-of-horse-riding-in-the-uk (accessed on Oct 26, 2020).

Dashper, K. (2017). Listening to Horses Developing Attentive Interspecies Relationships through Sport and Leisure. *Society & Animals, 25*, 207–224. https://doi.org/10.1163/15685306-12341426.

Davies, E., McConn-Palfreyman, W., Parker, J. K., Cameron, L. J., & Williams, J. M. (2022). Is Injury an Occupational Hazard for Horseracing Staff? *International Journal of Environmental Research and Public Health, 19*, 2054. https://doi.org/10.3390/IJERPH19042054/S1.

Davis, R. M., & Pless, B. (2001). BMJ bans "accidents". *BMJ, 322*, 1320–1321. https://doi.org/10.1136/BMJ.322.7298.1320.

Dee, S. J., Cox, B. L., & Ogle, R. A. (2013). Using near misses to improve risk management decisions. *Process Safety Progress, 32*, 322–327. https://doi.org/10.1002/PRS.11632.

Devon and Cornwall Police Operation Snap digital submissions Available online: https://operationsnap.devon-cornwall.police.uk/ (accessed on Apr 8, 2022).

Dohoo, I., Martin, W., & Stryhn, H. (2009). *Veterinary Epidemiologic Research* (2nd ed.). Canada: VER Inc..

Elvik, R., & Bjørnskau, T. (2017). Safety-in-numbers: A systematic review and meta-analysis of evidence. *Safety Science, 92*, 274–282. https://doi.org/10.1016/J.SSCI.2015.07.017.

Friedman, L. S., Hedeker, D., & Richter, E. D. (2009). Long-term effects of repealing the national maximum speed limit in the United States. *American Journal of Public Health, 99*, 1626–1631. https://doi.org/10.2105/AJPH.2008.153726.

Fyhri, A., Sundfør, H. B., Bjørnskau, T., & Laureshyn, A. (2017). Safety in numbers for cyclists—conclusions from a multidisciplinary study of seasonal change in interplay and conflicts. *Accident Analysis and Prevention, 105*, 124–133. https://doi.org/10.1016/J.AAP.2016.04.039.

Gildea, K., Hall, D., & Simms, C. (2021). Configurations of underreported cyclist-motorised vehicle and single cyclist collisions: Analysis of a self-reported survey. *Accident Analysis & Prevention, 159*. https://doi.org/10.1016/J.AAP.2021.106264.

GoSafe - Op snap Available online: https://gosafe.org/faq/operation-snap/ (accessed on Apr 8, 2022).

Guidance on road classification and the primary route network - GOV.UK Available online: https://www.gov.uk/government/publications/guidance-on-road-classification-and-the-primary-route-network/guidance-on-road-classification-and-the-primary-route-network (accessed on Nov 4, 2021).

Hospital Admitted Patient Care Activity 2019-20 - NHS Digital Available online: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2019-20 (accessed on Oct 26, 2020).

Hospital Episode Statistics (HES) - NHS Digital Available online: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics (accessed on Apr 8, 2022).

Isaksson-Hellman, I., & Töreki, J. (2019). The effect of speed limit reductions in urban areas on cyclists' injuries in collisions with cars. 20, 39–44, https://doi.org/10.1080/15389588.2019.1680836.

Jacobsen, P. L. (2003). Safety in numbers: More walkers and bicyclists, safer walking and bicycling. *Injury Prevention, 9*, 205–209. https://doi.org/10.1136/IP.9.3.205.

Jacobsen, P. L., Racioppi, F., & Rutter, H. (2009). Who owns the roads? How motorised traffic discourages walking and bicycling. *Injury Prevention, 15*, 369–373. https://doi.org/10.1136/ip.2009.022566.

James, H. F. (1991). Under-reporting of road traffic accidents. *Traffic Engineering and Control, 32*, 574–583.

Keaveney, S. M. (2008). Equines and their human companions. *Journal of Business Research, 61*, 444–454. https://doi.org/10.1016/j.jbusres.2007.07.017.

Lee Davis, D., Maurstad, A., & Dean, S. (2015). My Horse Is My Therapist: The Medicalization of Pleasure among Women Equestrians. *Medical Anthropology Quarterly, 29*, 298–315. https://doi.org/10.1111/maq.12162.

McGowan, T. W., Phillips, C. J. C., Hodgson, D. R., Perkins, N., & McGowan, C. M. (2012). Euthanasia in Aged Horses: Relationship between the Owner's Personality and Their Opinions on, and Experience of, Euthanasia of Horses. *Anthrozoos, 25*, 261–275. https://doi.org/10.2752/175303712X13403555186091.

Murphy, A., Dark, M., Bougdah, H., Baden, P., Bhagat, A., Ma, A., ... Benabbas, E. (2020). *Reported road casualties Great Britain: 2019 annual report*. Londong.

Norwood, S., McAuley, C., Vallina, V. L., Fernandez, L. G., McLarty, J. W., & Goodfried, G. (2000). Mechanisms and patterns of injuries related to large animals. In *Proceedings of the Journal of Trauma - Injury, Infection and Critical Care; Lippincott Williams and Wilkins* (Vol. 48, pp. 740–744).

Operation Snap | Warwickshire Police Available online: https://www.warwickshire.police.uk/police-forces/warwickshire-police/areas/warwickshire-police/campaigns/campaigns/2019/operation-snap/ (accessed on Apr 8, 2022).

Operation Snap | West Mercia Police Available online: https://www.westmercia.police.uk//police-forces/west-mercia-police/areas/west-mercia/campaigns/campaigns/2019/operation-snap/ (accessed on Apr 8, 2022).

Pazdan, S. (2020). The impact of weather on bicycle risk exposure. *Archives of Transport, 56*, 89–105. https://doi.org/10.5604/01.3001.0014.5629.

Phillips, R. O., Bjørnskau, T., Hagman, R., & Sagberg, F. (2011). Reduction in car-bicycle conflict at a road-cycle path intersection: Evidence of road user adaptation? *Undefined, 14*, 87–95. https://doi.org/10.1016/J.TRF.2010.11.003.

Pollard, D., & Furtado, T. (2021). Public roads as places of interspecies conflict: A study of horse-human interactions on UK roads and impacts on equine exercise. *Animals, 11*, 1072. https://doi.org/10.3390/ani11041072.

Pollard, D., & Grewar, J. D. (2020). Equestrian road safety in the United Kingdom: Factors associated with collisions and horse fatalities. *Animals, 10*, 2403. https://doi.org/10.3390/ani10122403.

Report Your Horse Incident | Riding Accidents | The BHS Available online: https://www.bhs.org.uk/our-work/safety/report-an-incident (accessed on Oct 26, 2020).

Reported road accidents, vehicles and casualties tables for Great Britain - GOV.UK Available online: https://www.gov.uk/government/statistical-data-sets/reported-road-accidents-vehicles-and-casualties-tables-for-great-britain (accessed on Nov 16, 2021).

Reported road casualties Great Britain, annual report: 2019 - GOV.UK Available online: https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2019 (accessed on Nov 3, 2020).

Research shines linght on equestrian-related injuries - Horse & Hound Available online: https://www.horseandhound.co.uk/news/they-dont-go-to-hospital-unless-somethings-hanging-off-research-shines-light-on-equestrian-related-injuries-767344 (accessed on Apr 7, 2022).

Richter, E. D., Berman, T., Friedman, L., & Ben-David, G. (2006). Speed, road injury, and public health. *Annual Review of Public Health, 27*, 125–152. https://doi.org/10.1146/ANNUREV.PUBLHEALTH.27.021405.102225.

Rights of way and accessing land: Use public rights of way - GOV.UK Available online: https://www.gov.uk/right-of-way-open-access-land/use-public-rights-of-way (accessed on Oct 26, 2020).

Road Safety Data - data.gov.uk Available online: https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data (accessed on Oct 26, 2020).

Sanders, R. L. (2015). Perceived traffic risk for cyclists: The impact of near miss and collision experiences. *Accident Analysis and Prevention, 75*, 26–34. https://doi.org/10.1016/j.aap.2014.11.004.

Scofield, R. M., Savin, H., & Randle, H. (2013). Road safety: Is there a relationship between "near misses" and the use of rider and horse reflective/flourescent equipment. *Proceedings of the Proceedings of the 9th International Equitation Science Conference, University of Delaware, 17th-20th July 2013; University of Delaware Printing Services: Newark, Delaware*.

Simsekoglu, Ö., Dalland, E. B., & Robertsen, K. (2020). Predictors of perceived road collision/incident risk among horse users: A survey study from Norway. *Transportation Research Part F: Traffic Psychology, 74*, 248–258. https://doi.org/10.1016/j.trf.2020.08.022.

Stewart, A. E., & Lord, J. H. (2002). Motor Vehicle Crash versus Accident: A Change in Terminology Is Necessary. *Journal of Traumatic Stress, 154*(15), 333–335. https://doi.org/10.1023/A:1016260130224.

Thompson, K., & Matthews, C. (2015). Inroads into Equestrian Safety: Rider-Reported Factors Contributing to Horse-Related Accidents and Near Misses on Australian Roads. *Animals, 5*, 592–609. https://doi.org/10.3390/ani5030374.

Thompson, K., McGreevy, P., & McManus, P. (2015). A critical review of horse-related risk: A research agenda for safer mounts, riders and equestrian cultures. *Animals, 5*, 561–575. https://doi.org/10.3390/ani5030372.

Trump, D., & Parkin, J. (2020). Equestrian road collisions: An analysis of the collision record, riding behaviours and experiences. *Safety Science, 123*, 104523. https://doi.org/10.1016/J.SSCI.2019.104523.

Uniting the Movement | Sport England Available online: https://www.sportengland.org/why-were-here/uniting-the-movement#thefivebigissues-12074 (accessed on Nov 17, 2021).

Webster, E., & Davies, D. (2020). What kills most on the roads? New analysis for the new transport agenda; London.

Welsh, C. E., Duz, M., Parkin, T. D. H., & Marshall, J. F. (2016). Prevalence, survival analysis and multimorbidity of chronic diseases in the general veterinarian-attended horse population of the UK. *Preventive Veterinary Medicine, 131*, 137–145. https://doi.org/10.1016/j.prevetmed.2016.07.011.

Wylie, C. E., Ireland, J. L., Collins, S. N., Verheyen, K. L. P., & Newton, J. R. (2013). Demographics and management practices of horses and ponies in Great Britain: A cross-sectional study. *Research in Veterinary Science, 95*, 410–417. https://doi.org/10.1016/j.rvsc.2013.05.004.

Dr Dee Pollard - Dee finished her undergraduate BSc(Hons) degree at Rhodes University in South Africa, followed by a MSc in Animal Breeding and Genetics from Wageningen University (The Netherlands) and the Swedish University of Agricultural Sciences, investigating inherited disorders in horses. In 2013 Dee started a World Horse Welfare-funded laminitis PhD offered by the Royal Veterinary College. This saw her make the move to the Animal Health Trust, where she spent 7 years, taking on a post-doc position after the completion of her PhD. Dee's PhD and post-doc work looked at the epidemiology of laminitis in Great Britain and what we can do to help reduce the risk. Dee currently works at the British Horse Society and completed a Department for Transport-funded project investigating the perceptions and attitudes to road safety by equestrians in the UK. Dee also works with other researchers and veterinary surgeons to consult on data analysis and statistical methods.

Dr John Grewar - John completed his Bachelor of Veterinary Science and MSc (Veterinary Science) degree in 2006 and 2009 respectively. He worked as an epidemiologist for a Provincial Government in South Africa from 2009 until 2016 and then began work as a senior researcher for the Equine Health Fund with Wits Health Consortium. This work has evolved into his current position as Research and Innovation Manager at the South African Equine Health and Protocols NPC. His work includes the development of systems for the processing and reporting of equine health in South Africa with the goal of developing sustainable systems within South Africa that will enable the country to develop protocols with target trade partners. He has a special interest in surveillance and outbreak investigation, and has recently completed a PhD focussing on the evaluation of African horse sickness surveillance programs.

Journal of Safety Research 84 (2023) 364–370

Contents lists available at ScienceDirect

Journal of Safety Research

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

Comparison of management and workers' perception, attitudes and beliefs toward health and safety in the Ontario manufacturing sector

Chun-Yip Hon [a,*], Jaskaren Randhawa [b], Norman Lun [a], Craig Fairclough [a,b], Linda Rothman [a]

[a] School of Occupational and Public Health, Toronto Metropolitan University, Canada
[b] Workplace Safety & Prevention Services, Canada

ARTICLE INFO

Article history:
Received 9 May 2022
Received in revised form 17 September 2022
Accepted 30 November 2022
Available online 12 December 2022

Keywords:
Perceptions, attitudes, and beliefs
Occupational health and safety
Workers
Managers
Manufacturing

ABSTRACT

Introduction: The Ontario manufacturing sector is over-represented when it comes to workers' compensation claims in the province. A previous study suggested that this may be the result of compliance gaps with respect to the province's occupational health and safety (OHS) legislation. These gaps may be, in part, due to differences in perceptions, attitudes, and beliefs toward OHS between workers and management. This is noteworthy as these two cohorts, when working well together, can foster a healthy and safe work environment. Therefore, this study sought to ascertain the perceptions, attitudes, and beliefs of workers and management with respect to OHS in the Ontario manufacturing sector and to identify differences between the groups, if any. *Methods*: A survey was created and disseminated online to get the widest reach across the province as possible. Descriptive statistics were used to present the data and chi-square analyses were performed to determine if there were any statistically significant differences in responses between workers and managers. *Results*: In total, 3,963 surveys were included in the analysis, which consisted of 2,401 (60.6%) workers and 1,562 (39.4%) managers. Overall, workers were more likely to state that their workplace was 'a bit unsafe' relative to managers and this difference was statistically significant. There were also statistically significant differences between the two cohorts with respect to health and safety communication matters, the perception of safety as a high priority, whether people work safely when unsupervised, and whether control measures are adequate. *Conclusions*: In summary, there were differences in perception, attitudes, and beliefs toward OHS between workers and managers in Ontario manufacturing and these differences must be addressed in order to improve the sector's health and safety performance. *Practical Applications*: Manufacturing workplaces can improve their health and safety performance by strengthening labor-management relationships, including having routine health and safety communication.

1. Introduction

Ensuring a healthy and safe work environment is not only the 'right thing to do,' but there is also a rather significant financial burden associated with work-related incidents. According to the Workplace Safety and Insurance Board (WSIB) in Ontario, Canada, nearly $2.5 billion in benefit payments were issued for work-related injury and illness claims in 2020 (Workplace Safety and Insurance Board, 2022a). In particular, Ontario's manufacturing sector reported 7,205 lost-time injury claims, which represents 15 % of all claims in the province (Workplace Safety and Insurance Board, 2022b). This is noteworthy as this sector only employs about 12 % of the total workforce in Ontario (Government of Ontario, 2011). Given the above WSIB claim statistics, a reasonable person would argue that the current situation is unacceptable and that measures need to be taken to rectify it. One means of addressing this issue is to examine compliance gaps within the manufacturing sector with respect to the province's occupational health and safety legislation. A study by Hon and Fairclough (2017) found that many Ontario manufacturing workplaces were not meeting the minimum requirements mandated in the province's Occupational Health and Safety Act and its Regulations, such as education and training as well as health and safety policies. To understand why these gaps exist, the authors recommended that future research should include an assessment of the attitudes, beliefs, and perceptions of those in the province's manufacturing sector regarding occupational health and safety (Hon &

* Corresponding author at: School of Occupational and Public Health, Toronto Metropolitan University, 350 Victoria Street, DCC304, Toronto ON M5B 2K3, Canada.

E-mail address: cyhon@ryerson.ca (C.-Y. Hon).

https://doi.org/10.1016/j.jsr.2022.11.011
0022-4375/

Fairclough, 2017). According to a review of 30 years of safety climate research, past studies have primarily focused on workers' perceptions toward safety in the workplace (Zohar, 2010). Given that occupational health and safety is rooted by the internal responsibility system which states that every-one in the workplace (regardless of position) plays a role in health and safety (Government of Ontario, 2022), it is important to understand the perspectives toward health and safety from the two key workplace parties – workers and managers. Survey-based studies have previously identified workplace health and safety perception differences between management and workers in steel mills in southeastern United States (Prussia et al., 2003), the nuclear sector in the United States (Findley et al., 2007), the trucking industry in the United States (Huang et al., 2014), as well as in construction in Columbia (Marín et al., 2019). These previous studies found that managers tended to have a higher perception of workplace health and safety than workers, which can lead to organizational conflict (Findley et al., 2007). To the best of our knowledge, this is the first study of its kind to explore this issue in Ontario – the province with the largest workforce in Canada (Statistics Canada, 2022). Therefore, the objective of this study was to compare management and workers' perceptions, attitudes, and beliefs toward occupational health and safety in the Ontario manufacturing sector, since an understanding of these viewpoints is critical toward addressing compliance gaps.

2. Methods

This was a survey-based cross-sectional study in which institutional ethics was approved prior to the collection of any data (Ryerson REB 2020–385).

2.1. Survey design

As no similar study has been conducted previously, a *de novo* survey was developed with questions related to attitudes, beliefs, and perceptions of occupational health and safety. Most of the questions were extracted and modified from existing surveys – many of which were previously pre-tested and/or validated (Adebola, 2014; Health & Safety Executive, 2004; Prairie Research Associates, 2015). The survey was arranged into various sections with the first section containing demographic questions such as job role and employment length. The second section, with multiple Likert-type responses, asked about an individual's perception toward health and safety (sourced primarily from the Health & Safety Executive questionnaire; Health & Safety Executive, 2004). The third section sought the respondent's attitudes and beliefs regarding health and safety initiatives in the workplace, such as communication and hazard control measures (sourced primarily from the Prairie Research Associates survey; Prairie Research Associates, 2015). In this section, respondents gave their level of agreement toward various statements using a 5-point Likert-type response ranging from strongly agree (1) to strongly disagree (5). Lastly, in the fourth section, respondents were asked to identify their top three workplace health and safety concerns.

The questionnaire has been included as a Supplemental File. It is evident in the survey that questions 1 to 10 were related to demographics, and the question on page 5 asked about specific hazards found in the workplace. Cronbach's alpha was calculated for the remaining questions related to perceptions, attitudes, and beliefs (after recoding to ensure the answers were going in the same direction) to ensure internal consistency. The result was 0.89, which suggests a good consistency (Bland et al., 1997). In addition, the survey was pre-tested by several volunteers to ensure that the survey was functional on multiple browser types and that the questions were unambiguous. Where relevant, questions were revised based on feedback from pilot respondents.

2.2. Participant recruitment

Workplace Safety & Prevention Services (WSPS), a not-for-profit health and safety association in Ontario that serves the manufacturing sector, led the recruitment of prospective participants. WSPS marketed the survey on their social media outlets (Facebook, Twitter, LinkedIn) and in electronic newsletters that encouraged both managers and workers from the Ontario manufacturing sector to complete the survey. Upon clicking the link to the survey, a respondent was taken to the consent preamble that preceded the survey questions. To encourage participation, prospective respondents were given the opportunity to enter a draw for a $50 e-gift card after completing the questionnaire.

The inclusion criteria for workers were that the individual must have been employed in their current workplace for at least three months so that they had an opportunity to gain an understanding of the occupational health and safety practices/protocols in their workplace. Meanwhile, the manager's inclusion criteria were that they had to be in their current role for at least three months to have had some time to understand their role and responsibilities from an occupational health and safety perspective.

The survey was hosted on a web-based platform, Opinio, and was open from February 15, 2021, to April 30, 2021.

2.3. Data analysis

Descriptive frequency statistics were used to report the participants' responses stratified by worker and manager. Chi-square analysis was conducted to determine whether there was a statistically significant difference (p-value < 0.05) in the responses regarding the perceptions, attitudes, and beliefs of occupational health and safety between the two cohorts. For those variables found to be significant, post-hoc testing was performed to identify response categories that were statistically significant from one another. The top five occupational health and safety concerns were ranked by worker and manager/supervisor. The statistical analyses were performed using SAS version 9.4 (Toronto, ON) and figures were created using R version 4.0.4.

3. Results

3.1. Demographics

A total of 5,245 responses were received, of which 1,282 were excluded from analysis because the respondent's job role was not specified, they did not reside in Ontario, they were not employed in the Ontario manufacturing industry, or less than 90 % of their survey was completed. Overall, 3,963 (75.6 %) of all responses were included in the analysis, consisting of 2,401 (60.6 %) workers and 1,562 (39.4 %) managers/supervisors (however, the actual sample size for each question differed from these values as respondents might have opted to skip some of the questions). A response rate could not be determined accurately because the number of people who were invited to participate was unknown. Relative to managers, a greater proportion of the workers were younger, female, had part-time job status, were employed for a shorter length of time, were employed at facilities with a smaller workforce (<50 employees), and more often worked evenings and nights (Table 1).

Table 1
Characteristics of respondents stratified by Worker and Manager/Supervisor.

Variable	Subcategory	Worker (%)	Manager/Supervisor (%)	Chi-square p-value
Age	25 and under	290 (12.1)	82 (5.3)	<0.0001
	25 to 35	1183 (49.4)	689 (44.2)	
	36 to 45	790 (33.0)	620 (39.8)	
	46 to 55	122 (5.1)	149 (9.6)	
	55 and above	12 (0.5)	19 (1.2)	
Sex	Female	1150 (48.3)	641 (41.4)	<0.0001
	Male	1121 (51.3)	898 (58.0)	
	Do not identify as either male or female	10 (0.4)	10 (0.7)	
Job Status	Full-time	1701 (71.9)	1258 (81.7)	<0.0001
	Part-time	631 (26.7)	245 (15.9)	
	Casual	35 (1.5)	36 (2.3)	
Postal Code	K	303 (12.6)	145 (9.3)	0.007
	L	507 (21.1)	331 (21.2)	
	M	805 (33.5)	519 (33.2)	
	N	407 (17.0)	276 (17.7)	
	P	379 (15.8)	291 (18.6)	
Employment Length	Less than 1 year	314 (13.2)	52 (3.4)	<0.0001
	At least 1 year, but less than 5 years	1179 (49.4)	614 (39.6)	
	At least 5 years, but less than 10 years	764 (32.0)	696 (44.9)	
	At least 10 years, but less than 20 years	115 (4.8)	164 (10.6)	
	20 years or more	13 (0.6)	26 (1.7)	
Industry	Computer and electronic manufacturing	515 (21.6)	305 (19.7)	0.002
	Food, textiles and related manufacturing	691 (29.0)	380 (24.5)	
	Machinery, electrical equipment and miscellaneous manufacturing	647 (27.1)	451 (29.1)	
	Metal, transportation equipment and furniture manufacturing	363 (15.2)	267 (17.2)	
	Non-metallic and mineral manufacturing	94 (3.9)	85 (5.5)	
	Printing, petroleum and chemical manufacturing	77 (3.2)	64 (4.1)	
Unionized workplace?	Yes	1964 (85.6)	1253 (85.4)	0.83
	No	330 (14.4)	215 (14.7)	
Facility Size	Small (less than 50 employees)	729 (30.6)	250 (16.1)	<0.0001
	Medium (50 to 250 employees)	1408 (59.1)	947 (60.9)	
	Large (251+ employees)	247 (10.4)	358 (23.0)	
Normal work shift	Day shift	1264 (52.8)	896 (57.5)	<0.0001
	Evening shift	522 (21.8)	239 (15.3)	
	Night shift	132 (5.5)	60 (3.9)	
	A mix of shifts	474 (19.8)	364 (23.4)	

3.2. Perceptions of health and safety

In terms of perception of overall safety, there was a statistically significant difference between workers and management whereby a larger proportion of workers (14.97 %) felt the workplace was 'a bit unsafe' compared to management (11.45 %) (Fig. 1). There were also statistically significant differences between the two cohorts regarding the workplace party they believed was primarily responsible for controlling health and safety risks in the workplace, the amount of health and safety training that had been provided, and whether they perceived the amount of health and safety training as being adequate (Table 2). A larger proportion of workers believed that "employer" and "employee" were primarily responsible for controlling the health and safety risks in the workplace, whereas managers generally felt that this was within the purview of the "occupational health and safety coordinator/department" or the "Ontario Ministry of Labour." The response of having had 'not very much' health and safety training was significantly higher in workers (21.97 %) compared to management (11.72 %), and significantly more workers felt that the training that they have received was "not enough" (14.81 %) compared with managers/supervisors (11.72 %).

3.3. Attitude and beliefs toward occupational health and safety issues

There was a statistically significant difference in the beliefs of workers and managers regarding various aspects of communication of occupational health and safety matters (Fig. 2). Specifically, managers were more likely to agree that workers are involved in safety decisions (69 % vs 64 %), that there are frequent communications about safety in the workplace (71 % vs 66 %), and workers are regularly asked about their safety concerns (70 % vs 66 %).

Approximately 70 % of both cohorts agreed or strongly agreed with the statement, "workplace injuries and accidents are an inevitable part of life" (Table 3). There were also statistically significant differences between the two cohorts in their level of agreement with respect to the following statements (i.e., agree or strongly agree):

- The safety of workers is a high priority for my workplace (77 % managers vs 73 % workers)
- Workplace health and safety requirements negatively impacts business operations (56 % workers vs 53 % managers)
- At my workplace, people here always work safely even when they are not being supervised (67 % managers vs 66 % workers)

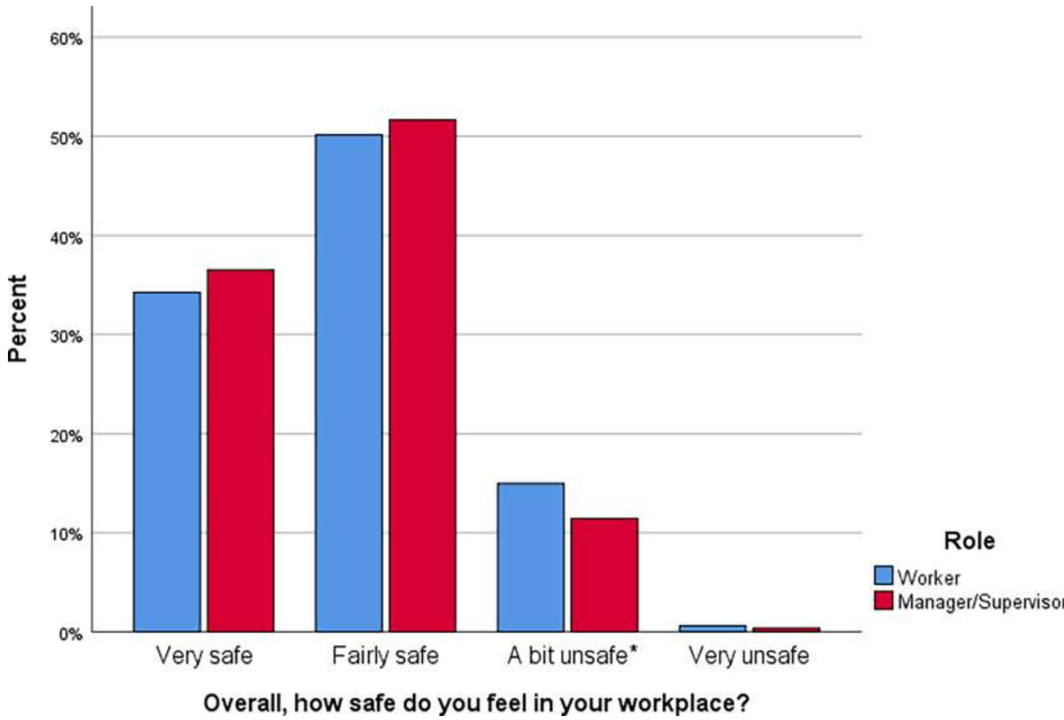


Fig. 1. Responses by Worker (n = 2391) and Manager/Supervisor (n = 1555) to the question "Overall, how safe do you feel in your workplace?" (p < 0.001); **statistically significant in post-hoc testing (p < 0.05).*

Table 2
Perceptions of occupational health and safety stratified by Worker and Manager/Supervisor.

Perception	Subcategory	Worker (%)	Manager/Supervisor (%)	Chi-square p-value
Which party do you think is primarily responsible for controlling health and safety risks in the workplace?	Employer	737 (30.8%)**	370 (23.8%)**	<0.001
	Employee	561 (23.4%)**	260 (16.7%)**	
	Joint Health and Safety Committee	465 (19.4%)	362 (23.3%)	
	OHS Coordinator/Department	437 (18.3%)**	397 (25.5%)**	
	Ontario Ministry of Labour	163 (6.8%)*	150 (9.6%)*	
	Other (please specify)	0 (0.0%)	2 (0.1%)	
	Don't know	32 (1.3%)	15 (1.0%)	
How much health and safety training have you received from your current employer?	A great deal	444 (18.6)	248 (16.0)	<0.001
	A fair amount	1347 (56.6)**	985 (63.4)**	
	Not very much	523 (22.0)**	182 (11.7)**	
	None at all	67 (2.8)**	12 (0.8)**	
The amount of health and safety training that you have received from your current employer is...	Too much	599 (25.2)	374 (24.1)	0.01
	About right	1396 (58.8)*	985 (63.4)*	
	Not enough	352 (14.8)*	182 (11.7)*	
	Don't know	29 (1.2)	12 (0.8)	

Note: statistically significant in post-hoc testing: * p < 0.05; ** p < 0.01.

- Not all the health and safety procedures/rules are strictly followed at my workplace (63 % managers vs 61 % workers)
- At my workplace, hazard control measures are adequate (65 % workers vs 62 % managers)
- My workplace has a joint health and safety committee that is effective at improving safety (72 % managers vs 66 % workers)

3.4. Top occupational health and safety concerns

While machine safety was the top work-related health and safety concern for both cohorts, it was significantly higher in management (32.65 %) compared to workers (26.16 %) (Table 4). The other top health and safety concerns were similar between the two cohorts and included chemical hazards, radiation, fire hazards, biological hazards, as well as electrical hazards.

4. Discussion

This study examined the perceptions, attitudes, and beliefs regarding occupational health and safety between workers and managers in the Ontario manufacturing sector. The results indicate that there are indeed differences between the two cohorts in some key areas. Overall, workers feel less safe than managers in the workplace. This finding is not surprising as similar results have been reported previously, albeit in other industrial sectors (Hallowell, 2010; Marín et al., 2019). Also, this result may be attributed to the fact that workers are more regularly exposed to haz-

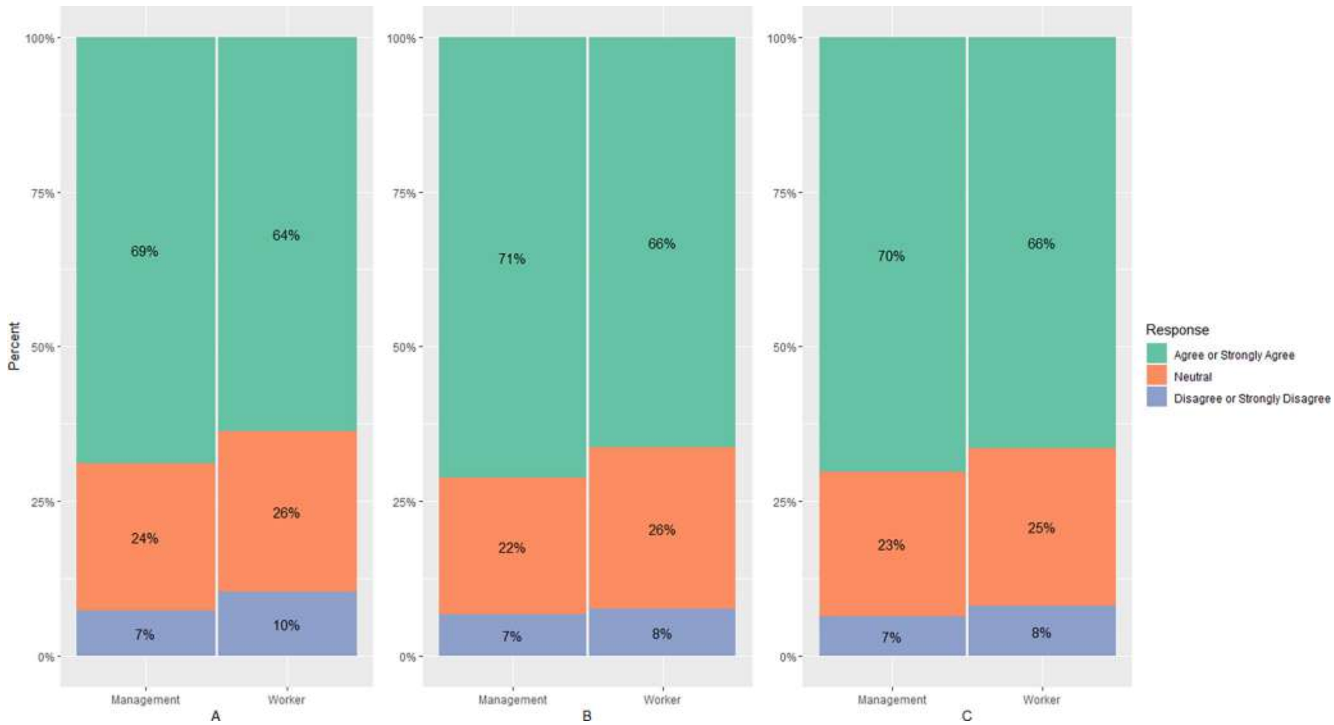


Fig. 2. Perception of workers and managers with respect to health and safety communication. Specific questions are: (A) "Workers here are involved in decisions affecting their safety" ($p < 0.01$); (B) "There are frequent communications about safety in my workplace" ($p < 0.01$); and (C) "Workers are regularly asked about their safety concerns" ($p < 0.05$).

ards, whereas managers typically perform less hazardous administrative tasks (Marín et al., 2019; Nordlöf et al., 2015).

There were also statistically significant differences between workers and managers with respect to their perception of health and safety training – both in terms of amount and whether it was believed to be enough. A systematic review found that health and safety training leads to improvements with respect to knowledge, safe behaviors, as well as health outcomes (Robson et al., 2012). As such, it would be prudent for the participating organizations to continue offering and possibly expand the amount of health and safety training that is available. In addition, it is suggested that this training review the roles and responsibilities of the various workplace parties, in particular, that the employer has ultimate responsibility for health and safety in Ontario (Occupational Health and Safety Act, 2016), as less than 30 % of respondents correctly identified the employer as the primary party responsible for managing health and safety risks in the workplace. Prussia et al. (2003) also found differences in the way managers and workers attribute responsibility for safety. This bears mentioning as employers that invest in workplace health and safety and are proactive are more likely to experience fewer injuries and illnesses (Battaglia et al., 2015; Geldart et al., 2010).

Statistically significant differences in perceptions regarding health and safety communication between managers and workers were found, including frequency of workplace safety information. This disconnect is noteworthy as open and timely communication is correlated with improved health and safety performance (Cigularov et al., 2010; Gittleman et al., 2010).

A notable finding is that both cohorts feel that workplace injuries and accidents are an inevitable part of life. This fatalistic attitude has been attributed to poor safety culture (Henning et al., 2009; Nordlöf et al., 2015), which bears mentioning as studies have shown that a more positive safety culture in a workplace leads to better health and safety performance (Wu et al., 2008; Noweir et al., 2013). It would therefore be important to address these negative attitudes so that both workers and management recognize that many work-related incidents are preventable and, in turn, improve their organization's health and safety performance.

When comparing the beliefs and attitudes of the two cohorts toward health and safety, there were several statistically significant differences. This included the finding that workers were less likely to agree that: (a) their safety is a high priority, (b) people work safely even when not being supervised, (c) hazard control measures are adequate, and (d) the Joint Health and Safety Committee is effective at improving health and safety. Meanwhile, managers were less likely to agree that (a) health and safety rules/procedures are being followed and (b) workplace health and safety requirements negatively impacts operations. These findings are consistent with previous studies that found differences in safety beliefs and attitudes between these two cohorts in other industries (Findley et al., 2007; Gittleman et al., 2010; Huang et al., 2012; Huang et al., 2014).

Such differences between managers and workers can result in misunderstandings, conflict, increased risk, and poorer safety performance (Findley et al., 2007; Human Factors Group, 2002). Studies have also shown that workers' safety performance is greatly influenced by their relationship with their managers, including a reduction in injury rates when their managers showed greater concern for them and supported the workers' positive safety behaviors (Geldart et al., 2010; Hofmann & Morgeson, 2009). Therefore, the next logical step would be to identify those factors that lead to these differences in beliefs and attitudes between workers and managers and, subsequently, address these gaps in order to improve health and safety performance (Marín et al., 2019).

Despite differences in perceptions and attitudes, both workers and managers were in agreement with respect to the top health and safety issue in the Ontario manufacturing sector – machine safety. This is not surprising as this hazard was highlighted as a common concern in the manufacturing sector by the Ontario Ministry of Labour (Government of Ontario, 2021). However, the

Table 3
Attitudes and beliefs toward occupational health and safety stratified by Worker and Manager/Supervisor.

Attitude or belief	Cohort	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Chi-square p-value
Workplace injuries and accidents are an inevitable part of life	Worker (%)	659 (27.5)	1017 (42.5)	569 (23.8)	133 (5.6)	17 (0.7)	0.68
	Manager (%)	421 (27.1)	667 (42.9)	363 (23.3)	86 (5.5)	18 (1.2)	
At my workplace, safety is as important as quality of the work and getting the work done on time	Worker (%)	660 (27.7)	1020 (42.8)	543 (22.8)	141 (5.9)	20 (0.8)	0.186
	Manager (%)	469 (30.2)	614 (39.6)	367 (23.7)	84 (5.4)	18 (1.2)	
The safety of workers is a high priority for my workplace	Worker (%)	667 (28.3)	1065 (44.5)	526 (22.0)	108 (4.5)	16 (0.7)	0.048
	Manager (%)	489 (31.4)	702 (45.1)	298 (19.2)	62 (4.0)	5 (0.3)	
Workplace health and safety requirements negatively impacts business operations	Worker (%)	398 (16.6)	946 (39.5)	591 (24.7)	291 (12.2)	169 (7.1)	0.043
	Manager (%)	257 (16.5)	573 (36.8)	407 (26.2)	173 (11.1)	146 (9.4)	
Formal safety inspections are regularly conducted in my workplace	Worker (%)	632 (26.5)	1023 (42.9)	564 (23.6)	152 (6.4)	16 (0.7)	0.193
	Manager (%)	431 (27.8)	648 (41.8)	387 (25.0)	73 (4.7)	10 (0.7)	
New employees at my workplace learn quickly that they are expected to follow safety rules	Worker (%)	574 (24.1)	1037 (43.6)	580 (24.4)	169 (7.1)	21 (0.9)	0.44
	Manager (%)	373 (24.1)	700 (45.2)	376 (24.3)	87 (5.6)	13 (0.8)	
At my workplace, safety is given a high priority in training programs	Worker (%)	548 (23.0)	1045 (43.8)	592 (24.8)	181 (7.6)	22 (0.9)	0.071
	Manager (%)	365 (23.6)	695 (44.9)	388 (25.1)	82 (5.3)	19 (1.2)	
At my workplace, there are rules and procedures about how to work safely	Worker (%)	535 (22.4)	1063 (44.5)	605 (23.4)	168 (7.0)	16 (0.7)	0.113
	Manager (%)	366 (23.6)	731 (47.2)	346 (22.3)	93 (6.0)	13 (0.8)	
At my workplace everyone has the tools and equipment they need to do their job safely	Worker (%)	549 (23.1)	1050 (44.2)	599 (25.2)	162 (6.8)	16 (0.7)	0.156
	Manager (%)	378 (24.5)	705 (45.6)	368 (23.8)	80 (5.2)	14 (0.9)	
At my workplace, people here always work safely even when they are not being supervised	Worker (%)	519 (21.8)	1045 (43.9)	598 (25.1)	184 (7.7)	33 (1.4)	0.04
	Manager (%)	317 (20.6)	719 (46.7)	402 (26.1)	84 (5.5)	18 (1.2)	
Not all the health and safety procedures/ rules are strictly followed at my workplace	Worker (%)	467 (19.5)	992 (41.5)	625 (26.2)	246 (10.3)	60 (2.5)	0.018
	Manager (%)	293 (18.9)	685 (44.2)	393 (25.3)	123 (7.9)	57 (3.7)	
At my workplace, disciplinary action is taken against people who break health and safety procedures/ instructions/rules	Worker (%)	515 (21.6)	1073 (45.0)	587 (24.6)	184 (7.7)	23 (1.0)	0.267
	Manager (%)	308 (19.9)	717 (46.3)	401 (25.9)	102 (6.6)	21 (1.4)	
At my workplace, hazard control measures are adequate	Worker (%)	488 (20.5)	1053 (44.3)	626 (26.3)	185 (7.8)	32 (1.3)	0.017
	Manager (%)	272 (17.6)	696 (44.9)	457 (29.5)	97 (6.3)	28 (1.8)	
Workplace health and safety requirements benefits my workplace	Worker (%)	568 (23.8)	1074 (45.0)	549 (23.0)	164 (6.9)	33 (1.4)	0.376
	Manager (%)	351 (22.6)	735 (47.2)	332 (21.3)	109 (7.0)	29 (1.9)	
My workplace has a joint health and safety committee that is effective at improving safety	Worker (%)	541 (22.7)	1039 (43.7)	609 (25.5)	174 (7.3)	22 (0.9)	0.003
	Manager (%)	398 (25.8)	717 (46.4)	328 (21.2)	89 (5.8)	12 (0.8)	

Table 4
Top five occupational health and safety concerns stratified by Worker and Manager/Supervisor.

Worker Concerns (%)	Rank	Manager/Supervisor Concerns (%)
Machine safety (26.2)	1	Machine safety (32.7)
Chemical hazards (25.7)	2	Fire hazards (27.2)
Radiation (24.1)	3	Radiation (26.0)
Fire hazards (23.5)	4	Chemical hazards (24.7)
Biological hazards (22.6)	5	Electrical hazards (22.5)

results found that a statistically significant larger percentage of managers ranked machine safety as the top hazard, suggesting that this is another gap that needs to be addressed in order to improve health and safety performance.

There are limitations associated with this study that need to be discussed. As this was a cross-sectional study, the results are only representative of the time of sample collection and perceptions may change over time. Therefore, a longitudinal examination of perceptions is suggested as a future study to determine any trends. As the response rate of the survey is not known, the study could have experienced response and/or non-response bias. Social desirability bias could not be ruled out as managers tends to answer more positively so as not to tarnish the reputation of their respective organizations (Marín et al., 2019). Also, this study did not identify how the respondents' perceptions, attitudes, and beliefs were formulated; examining the reasons for these perceptions in a future study would aid in developing appropriate interventions. Lastly, the purpose of this study was solely descriptive and meant to compare the responses between management and workers. Seeing that we did identify differences between the two cohorts, future work should include performing multivariate analyses to examine associations, if any, and to identify potential confounders.

In summary, this study found that the perceptions, attitudes, and beliefs regarding occupational health and safety of workers compared with managers in Ontario's manufacturing sector differ – in some cases, the differences were statistically significant. These differences must be narrowed through improvements in labor-management relations. In turn, this will lead to improved health and safety awareness and performance, including compliance with legislation, in the manufacturing sector in Ontario.

5. Practical Applications

This study concluded that changes are required to labor-management relationships in the Ontario manufacturing sector to improve health and safety performance. Based on our findings, some examples of changes include, but are not limited to: increased education and training; managers/supervisors to lead by example on safety issues; organizations being proactive with respect to health and safety concerns; and improving worker-management communication, especially with respect to health and safety issues.

Competing interest statement

The authors have no conflicts of interest to declare.

Acknowledgements

The authors wish to thank the Centre for Labour Management Relations at Toronto Metropolitan University (formerly Ryerson University) for funding this study. The authors are also appreciative of all the participants who took time to complete the survey.

References

Adebola, J. O. (2014). Knowledge, Attitude and Compliance with Occupational Health and Safety Practices among Pipeline Products and Marketing Company (PPMC) Staff in Lagos. *Merit Research Journal of Medicine and Medical Sciences, 2* (8), 2354–3238 http://www.meritresearchjournals.org/mms/index.htm.

Battaglia, M., Passetti, E., & Frey, M. (2015). Occupational health and safety management in municipal waste companies: A note on the Italian sector. *Safety Science, 72*, 55–65. https://doi.org/10.1016/j.ssci.2014.08.002.

Bland, J. M., Altman, D. G., & Bland, J. M. (1997). Statistics Notes : The Odds Ratio Linked references are available on JSTOR for this article : Statistics Notes The odds ratio. *British Medical Journal, 314*(7080), 572.

Cigularov, K. P., Chen, P. Y., & Rosecrance, J. (2010). The effects of error management climate and safety communication on safety: A multi-level study. *Accident Analysis and Prevention, 42*(5), 1498–1506. https://doi.org/10.1016/j.aap.2010.01.003.

Findley, M., Smith, S., Gorski, J., & O'neil, M. (2007). Safety climate differences among job positions in a nuclear decommissioning and demolition industry: Employees' self-reported safety attitudes and perceptions. *Safety Science, 45*, 875–889. https://doi.org/10.1016/j.ssci.2006.08.027.

Geldart, S., Smith, C. A., Shannon, H. S., & Lohfeld, L. (2010). Organizational practices and workplace health and safety: A cross-sectional study in manufacturing companies. *Safety Science, 48*(5), 562–569. https://doi.org/10.1016/J.SSCI.2010.01.004.

Gittleman, J. L., Gardner, P. C., Haile, E., Sampson, J. M., Cigularov, K. P., Ermann, E. D., Stafford, P., & Chen, P. Y. (2010). [Case Study] CityCenter and Cosmopolitan Construction Projects, Las Vegas, Nevada: Lessons learned from the use of multiple sources and mixed methods in a safety needs assessment. *Journal of Safety Research, 41*, 263–281. https://doi.org/10.1016/j.jsr.2010.04.004.

Government of Ontario (2011). *Manufacturing sector: Meeting report*. https://www.ontario.ca/page/manufacturing-sector-meeting-report.

Government of Ontario (2021). *Hazards in the Industrial Sector*. https://www.ontario.ca/page/hazards-industrial-sector.

Government of Ontario (2022). *The Internal Responsibility System*. https://www.ontario.ca/document/guide-occupational-health-and-safety-act/internal-responsibility-system.

Hallowell, M. (2010). Safety risk perception in construction companies in the Pacific Northwest of the USA. *Construction Management and Economics, 28*(4), 403–413. https://doi.org/10.1080/01446191003587752.

Health & Safety Executive. (2004). *Attitudes toward health and safety: A quantitative survey of stakeholder opinion*. https://www.ipsos.com/sites/default/files/migrations/en-uk/files/Assets/Docs/Archive/Polls/hse.pdf.

Henning, J. B., Stufft, C. J., Payne, S. C., Bergman, M. E., Mannan, M. S., & Keren, N. (2009). The influence of individual differences on organizational safety attitudes. *Safety Science, 47*(3), 337–345. https://doi.org/10.1016/j.ssci.2008.05.003.

Hofmann, D. A., & Morgeson, F. P. (2009). The role of leadership in safety. *The psychology of workplace safety*. https://doi.org/10.1037/10662-008.

Hon, C.-Y., & Fairclough, C. (2017). Occupational Health & Safety Legislative Gap Analysis of the Food Processing Sector in Ontario, Canada. *Journal of Safety, Health & Environmental Research, 13*(1), 338–344.

Huang, Y.-H., Verma, S. K., Chang, W.-R., Courtney, T. K., Lombardi, D. A., Brennan, M. J., & Perry, M. J. (2012). Supervisor vs. employee safety perceptions and association with future injury in US limited-service restaurant workers. *Accident Analysis and Prevention, 47*, 45–51. https://doi.org/10.1016/j.aap.2011.11.023.

Huang, Y. H., Robertson, M. M., Lee, J., Rineer, J., Murphy, L. A., Garabet, A., & Dainoff, M. J. (2014). Supervisory interpretation of safety climate versus employee safety climate perception: Association with safety behavior and outcomes for lone workers. *Transportation Research Part F: Traffic Psychology and Behaviour, 26*(PB), 348–360. https://doi.org/10.1016/J.TRF.2014.04.006.

Human Factors Group. (2002). *Safety Culture: A review of the literature*.

Marín, L. S., Lipscomb, H., Cifuentes, M., & Punnett, L. (2019). Perceptions of safety climate across construction personnel: Associations with injury rates. *Safety Science, 118*, 487–496. https://doi.org/10.1016/j.ssci.2019.05.056.

Nordlöf, H., Wiitavaara, B., Winblad, U., Wijk, K., & Westerling, R. (2015). Safety culture and reasons for risk-taking at a large steel-manufacturing company: Investigating the worker perspective. *Safety Science, 73*, 126–135. https://doi.org/10.1016/j.ssci.2014.11.020.

Noweir, M. H., Alidrisi, M. M., Al-Darrab, I. A., & Zytoon, M. A. (2013). Occupational safety and health performance of the manufacturing sector in Jeddah Industrial Estate, Saudi Arabia: A 20-years follow-up study. *Safety Science, 53*, 11–24. https://doi.org/10.1016/j.ssci.2012.09.005.

Occupational Health and Safety Act, (2016). https://www.ontario.ca/laws/statute/90o01.

Prarie Research Associates (2015). *Workplace Safety Climate Survey*. Safety Survey Tool.pdf.

Prussia, G. E., Brown, K. A., & Willis, P. G. (2003). Mental models of safety: Do managers and employees see eye to eye? *Journal of Safety Research, 34*(2), 143–156. https://doi.org/10.1016/S0022-4375(03)00011-2.

Robson, by L., Lynda Robson, by S., Stephenson, C. M., Schulte, P. A., Amick III, B. C., Irvin, E. L., Eggerth, D. E., Chan, S., Bielecky, A. R., Wang, A. M., Heidotting, T. L., Peters, R. H., Clarke, J. A., Cullen, K., Rotunda, C. J., & Grubb, P. L. (2012). A systematic review of the effectiveness of occupational health and safety training. *Scandinavian Journal of Work, 38*(3), *193-208*. https://doi.org/10.5271/sjweh.3259

Statistics Canada. (2022). *Labour Force Characteristics by Province*. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410028703.

Workplace Safety and Insuance Board. (2022a). *WSIB By the Numbers*. https://www.wsib.ca/en/bythenumbers/schedule-1-2020-highlights.

Workplace Safety and Insurance Board. (2022b). *Injury Claims by Sector*. https://www.wsib.ca/en/bythenumbers/schedule-1-industry-sector-claims-and-lti-rate.

Wu, T. C., Chen, C. H., & Li, C. C. (2008). A correlation among safety leadership, safety climate and safety performance. *Journal of Loss Prevention in the Process Industries, 21*(3), 307–318. https://doi.org/10.1016/j.jlp.2007.11.001.

Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis & Prevention, 42*(5), 1517–1522. https://doi.org/10.1016/J.AAP.2009.12.019.

Chun-Yip Hon: Dr. Hon is an Associate Professor with the School of Occupational and Public Health at Toronto Metropolitan University (formerly Ryerson University) in Toronto, Canada. Dr. Hon received his PhD in 2012 in occupational and environmental hygiene from the University of British Columbia. Dr. Hon is also a certified industrial hygienist (CIH) as well as a Canadian registered safety professional (CRSP) with nearly 25 years of experience as an occupational hygienist. In general, his research interests are occupational exposure assessments, risk assessment and evaluation of intervention measures to prevent occupational injury and/or illness.

Jaskaren Randhawa: Ms. Randhawa has a Master's of Public Health from the University of Sydney. She was the Research Program Specialist at the Workplace Safety and Prevention Services which provides health and safety consultation services for various sectors in the province of Ontario.

Norman Lun: Mr. Lun is a recent graduate of the Occupational Health and Safety Program at the School of Occupational and Public Health at Toronto Metropolitan University (formerly Ryerson University) in Toronto, Canada.

Craig Fairclough: Mr. Fairclough an occupational hygiene consultant with nearly 25 years of experience in occupational health and safety. He has a Master's in Occupational Health and possesses the Registered Occupational Hygienist (ROH), Certified Industrial Hygienist (CIH) and Canadian registered Safety Professional (CRSP) designations. Mr. Fairclough is also an Adjunct Professor with the School of Occupational and Public Health at Toronto Metropolitan University (formerly Ryerson University) in Toronto, Canada.

Linda Rothman: Dr. Rothman is an Assistant Professor with the School of Occupational and Public Health at Toronto Metropolitan University (formerly Ryerson University) in Toronto, Canada. She has a PhD from the Institute for Medical Sciences (IMS) with a focus on epidemiology at the University of Toronto and completed postdoctoral training in Kinesiology and Health Sciences at York University. Her area of expertise is in unintentional injury prevention and is also involved in research related to sport, playground, road traffic and car occupant injuries.

Journal of Safety Research 84 (2023) 61–73

Contents lists available at ScienceDirect

Journal of Safety Research

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

Comparison of motor-vehicle involved e-scooter fatalities with other traffic fatalities

Elizabeth Karpinski [a,*], Ellie Bayles [a], Lisa Daigle [b], Dan Mantine [c]

[a] The MITRE Corporation, 202 Burlington Road, Bedford, MA 01730, United States
[b] The MITRE Corporation, 2275 Rolling Run Drive, Windsor Mill, MD 21244, United States
[c] The MITRE Corporation, 7525 Colshire Drive, McLean, VA 22102, United States

ARTICLE INFO

Article history:
Received 29 October 2021
Received in revised form 22 June 2022
Accepted 17 October 2022
Available online 28 October 2022

Keywords:
E-scooters
Micromobility
Transportation fatalities
Transportation safety

ABSTRACT

Introduction: Shared e-scooters are an emerging mode of transportation with many features that make their physical properties, behavior, and travel patterns unique. Safety concerns have been raised concerning their usage, but it is difficult to understand effective interventions with so little data available. *Methods:* Using media and police reports, a crash dataset was developed of rented dockless e-scooter fatalities in crashes involving motor vehicles that occurred in the United States in 2018–2019 (n = 17) and the corresponding records from the National Highway Traffic Safety Administration data were identified. The dataset was used to perform a comparative analysis with other traffic fatalities during the same time period. *Results:* Compared to fatalities from other modes of transportation, e-scooter fatality victims are younger and more likely male. More e-scooter fatalities occur at night than any other mode, except pedestrians. E-scooter users are comparatively as likely as other unmotorized vulnerable road users to be killed in a hit-and-run crash. While e-scooter fatalities had the highest proportion of alcohol involvement of any mode, this was not significantly higher than the rate seen in pedestrian and motorcyclist fatalities. E-scooter fatalities were more likely than pedestrian fatalities to be intersection-related, and to involve crosswalks or traffic signals. *Conclusions:* E-scooter users share a mix of the same vulnerabilities as both pedestrians and cyclists. Although e-scooter fatalities are demographically most similar to motorcycle fatalities, crash circumstances share more similarities with pedestrian or cyclist fatalities. Other characteristics of e-scooter fatalities are notably distinct from other modes. *Practical Applications:* E-scooter use must be understood by users and policymakers to be a distinct mode of transportation. This research highlights the similarities and differences between similar modes, like walking and cycling. By using this information on comparative risk, e-scooter riders and policymakers can take strategic action to minimize the number of fatal crashes.

1. Introduction

The shared e-scooter is a powered two-wheel (PTW) vehicle with many features that make it a unique transportation mode. E-scooters offer many advantages as an alternative mode of transportation in dense urban areas, particularly the potential of reducing pollution and motor-vehicle traffic (Shaheen & Cohen, 2019). However, an increasing number of crashes and fatalities publicized in the news media have raised public concern about the safety of these devices. Families of victims have called for the removal of shared e-scooters in cities like Atlanta, Georgia (Hansen, 2019) and Ft. Lauderdale, Florida (Wallman & Maines, 2019). In August 2019, the Nashville City Council rejected a proposed ban on e-scooters following a fatal accident (Morgan, Lewis, & Bockius LLP, 2021). These concerns about the safety of these vehicles exist despite the small absolute number of fatalities compared to the number of other surface transportation fatalities (NHTSA, 2020b).

The different vehicle characteristics of and user behavior with e-scooters have raised questions about the similarities and differences between e-scooter crashes and other similar crashes involving vulnerable road users, such as cyclists and pedestrians (Kleinertz, Ntalos, Hennes, et al., 2021). Previous research on the nature of rented e-scooter fatalities has largely not been possible, due to the newness of the mode. While many cities have enforced different policies and restrictions on e-scooter use (Unagi, 2021), the lack of data makes it difficult to find significant evidence on crash risk. However, by developing a better understanding of

* Corresponding author.
E-mail address: Ekarpinski@mitre.org (E. Karpinski).

https://doi.org/10.1016/j.jsr.2022.10.008
0022-4375/

how shared e-scooter users face similar and different risks to people using better-understood modes of transportation (i.e., pedestrians, pedalcyclists, and motorcyclists) it is possible to form a clearer picture of e-scooter fatality risk.

2. Background

2.1. E-Scooters

Starting in 2017, e-scooter companies began rolling out shared e-scooter services in cities across the United States. Since over a third of residential trips made by vehicles in the United States are under two miles, the potential market for short-range transportation is large (Federal Highway Administration, 2017). Early pro e-scooter studies examined possible economic and environmental benefits of decreased automobile use (Shaheen & Cohen, 2019) and increased mobility (Smith & Schwieterman, 2018). However, safety concerns regarding e-scooter implementation have been prevalent since early adoption (Abcarian, 2018).

Shared e-scooters are extremely lightweight and small, even compared to motorcycles or mopeds. The Electric Scooter Guide (2021) database lists specifications for nearly 200 different e-scooter models. E-scooters are between 38 inches and 55 inches high, with a median weight of 47 pounds. Some e-scooters listed in the guide weigh as little as 23 pounds, but most of their analyzed models are between 32 pounds (first quartile) and 72 pounds (third quartile). Standing e-scooters are distinct for their 'standing design' where the rider remains upright while riding, standing on a narrow foot platform (Shaheen & Cohen, 2019). E-scooters are also designed for single-person transportation (SAE International, 2019). E-scooter companies typically forbid taking on a second rider because the second rider is in a hazardous position (Bird, 2020; Lime, 2021). In two of the fatal crashes identified that involved "double riding" with a second person on the e-scooter, it was the second rider or "passenger" who was fatally injured while the primary rider survived (Griggs & Lou, 2019; Carmel, 2019). Much like pedestrians and cyclists, the e-scooter rider is particularly vulnerable in comparison to the automobile occupant, with no barrier and no safety features like seatbelts or airbags to protect them in the event of an accident.

Even before the coronavirus pandemic started in 2020, over a hundred million trips had been taken in the shared e-scooter market (National Association of City Transportation Officials, 2019). Many companies operating in the area had achieved billion-dollar valuations in record time (Yakowicz, 2018). In 2019, 109 cities around the United States were host to these dockless systems, with mixed reception (NACTO, 2019). Although travel has decreased substantially across all modes due to the coronavirus pandemic, micromobility has emerged as an increasingly viable alternative (Perry, 2020); new research has targeted micromobility as a potential option to reduce the spread of contagious disease, since unlike public transit, it does not require many individuals to share the same air (National Science Foundation, 2020).

Early studies have largely focused on injuries that presented to emergency rooms (such as Austin Public Health and City of Austin, 2019; Badeau, Carman, et al., 2019) and other impacts to the healthcare system (such as Bekhit, Le Fevre, et al.; Mayhew & Bergin, 2019; Mitchell et al., 2019). But without significant fatality data, it has not been possible to extend this analysis to crashes that resulted in a fatality. Interventions to improve safety have also generally relied on analogies to cyclists, motorcyclists, or pedestrians. Interventions such as mandatory helmets (U.S. Consumer Product Safety Commission, 2014) or bike lanes (Bird, 2019) depend on justifying the similarity to cyclists. Other ongoing policy debates, such as whether e-scooters should ride on the sidewalk or the street (Unagi, 2021), also depend on an understanding of e-scooter risk, and how their features compare to other modes.

2.2. E-Scooter users

The unique features of e-scooters have also prompted unique behavior from users. Naturalistic studies have observed a "hybrid" phenomenon, where e-scooter riders may switch patterns of behavior rapidly and repeatedly; for example, acting as a pedestrian, then as a cyclist, then back to acting as a pedestrian (Todd, Krauss, Zimmermann, & Dunning, 2019). In addition, the shared aspect of the e-scooter allows users to rent the scooter for a short time, typically under 30 minutes (NACTO, 2019), and then leave the e-scooter at their destination for another user to pick up. This distributed model theoretically allows for high vehicle utilization (Tilleman & Feasley, 2018), but it has raised concerns over whether users can be responsible for handling the devices safely, bringing their own safety equipment (Penney and Associates, 2021), and whether they have the proper training to do so (Fielding, 2019). Early studies have suggested that e-scooter riders often have poor awareness of the e-scooter laws in the city, for example, whether or not riding on sidewalks is legal (James et al., 2019).

Previous literature suggests that e-scooter use, like bikeshare and privately owned bicycle ridership, tends to lean male. Naturalistic observations of e-scooter riders in Brisbane reported that 76% were "apparently" male, which was less than the sex-ratio they observed for private bicyclists (84%) but more than for shared bikes (72%) (Haworth, Schramm, & Twisk, 2021). Similarly, a survey of shared e-scooter riders in Vienna found that 74% identified as male (Laa & Leth, 2020), while 62% of users in a survey of 75,000 in Portland, Oregon self-identified as male (Portland Bureau of Transportation, 2018).

Males have made up a more slender majority in some of the studies conducted on e-scooter injuries. An early study by Austin Public Health (2019) found that injured e-scooter riders were 55% male, based on emergency room and medical services data. A similar rate (52%) was found in a Washington, D.C. study (Cicchino, Kulie, & McCarthy, 2021) that also looked at emergency department presentations; they also found a higher ratio of males in riders injured on the road (88%), but found that gender did not play a significant role in injury severity, as tested through ratings on the Abbreviated Injury Scale. A Singapore review of emergency department records using 2015–2016 data found that 66% of injured riders were male (Liew, Wee, & Pek, 2020). Nellamattathil and Amber (2020) used radiology data to study e-scooter injuries from a sample that was 76% male. Similarly, an analysis of media-reported crashes in the United States by Yang et al. (2020) found that 72% of the e-scooter riders involved were male.

Research has also suggested gendered differences in e-scooter perception and use. A survey of university staff in Arizona found that males were significantly more likely than females to perceive e-scooters as "very safe" (Sanders et al., 2020) and Dill (2019) indicated that males are more likely than females to use e-scooters for commuting and to say that their preference for e-scooters is due to speed, which may be related to their increased risk. In addition, previous research has suggested that e-scooter riders tend to go faster when commuting than when riding recreationally (Almannaa et al., 2021). Injury studies similarly suggest that commuters are more likely to be involved in on-road crashes, sustaining more severe injuries (Cicchino, Kulie, & McCarthy, 2021).

Although studies have found people of all ages injured in e-scooter crashes, measures of centrality have consistently suggested that the typical e-scooter rider is in their 20 s or 30 s. Cicchino et al. (2021) found that the mean age of the injured rider was 39.5 (SD: 15.2). Their results suggested that younger riders were significantly more likely to be involved in motor-vehicle crashes, and that their crashes were more likely to have happened at intersections. Badeau et al. (2019) found the average age of 34 years for an injured e-scooter rider who presented to an emergency department in Salt Lake City, Utah. Laa and Leth (2020) describe the e-scooter users in Vienna, Austria identified by their surveys and field observations as generally young or middle aged, with 24% of their sample under the age of 25, and 46% between the ages of 25 and 34.

2.3. Transportation safety

Our analysis also focused on some of the major aspects of transportation safety that have been studied with regards to motor vehicles, in order to study whether these factors may play a similar role in e-scooter fatalities.

2.4. Demographics

The role of age has long been researched as an important moderating factor in transportation fatalities. NHTSA's 1993 Report to Congress on safety issues related to younger and older drivers noted that crash involvement rates by age are highest for drivers aged 15–19 (at over 150 crashes per 1,000 licensed drivers, according to NHTSA data), and that this rate decreases substantially as drivers age (to roughly 50 crashes per 1,000 licensed drivers for the 40–44 age group) without rising again. However, in fatal crashes, the *age* of at-fault drivers follows a "U" shaped-distribution: high for younger and older drivers, and low for the ages in between (Eustace & Wei, 2010). The influence of age is marked in several ways: effect on behavior (i.e., risk-taking in younger populations; NHTSA, 1993); effect on susceptibility (e.g., the physical fragility in older populations; as in Kim et al., 2008); and the underlying distribution of usage rates, which also vary by mode of transportation with age. The final count of fatalities is a product of both the risk and the usage rates for each age bracket.

2.5. Hit-and-Runs

Hit-and-run refers to crashes where the driver illegally leaves the scene without stopping: hit-and-run crashes are considered a criminal offense, and they can increase crash severity by causing a delay in medical attention for the victims. In recent years, both the prevalence and proportion of fatal hit-and-run crashes have been increasing (Benson, Arnold, Tefft, & Horrey, 2018). This makes hit-and-run crash risk a question of increasing relevance and importance in traffic safety.

2.6. Involvement of alcohol

Although all 50 states and the District of Columbia have passed legislation against operating a motor vehicle with a blood alcohol concentration (BAC) of 0.08 or above, alcohol impairment remains a pressing traffic safety concern (NHTSA, 2021b). Overall, 28% of traffic fatalities occur in alcohol-impaired driving crashes (NHTSA, 2021b). The rate of alcohol impairment is also known to vary by state. In some states, like Rhode Island, over one-third of drivers involved in fatal crashes are impaired, while in Utah the rate is 11% (NHTSA, 2021b). However, it can be difficult to get reliable statistics on the rate of alcohol impairment in crashes because of the incentive for drivers to flee the scene and differing standards for testing between medical establishments and jurisdictions (NHTSA, 2021b).

A lesser-publicized traffic safety issue is the high rate of alcohol involvement in non-motorist fatalities. Many of the same risks present for drivers, including impaired psychomotor skills, visual-perceptual difficulties, and slowed information processing (Moskowitz & Burns, 1990) also endanger pedestrians, pedalcyclists, and other VRUs when they are impaired. For example, Oxley et al. (2006) suggest that road crossing behavior may be less safe in pedestrians with higher BAC levels. They attribute this to impaired judgment in gauging risk based on the speed and distance of the approaching vehicle. Research has also shown higher BACs to be associated with delayed response times and worse performance for motorcyclists (Creaser, Ward, et al., 2009).

Previous studies on e-scooter injuries have drawn attention to the rate of alcohol involvement among injured riders. Estimates have ranged widely. Studies have reported 16% (patient-reported) in a Salt Lake City, Utah emergency room (Badeau et al., 2019); 13% in Dunedin, New Zealand based on routine alcohol screening in emergency room presentations (Beck, Barker, et al., 2020); 38% tested above the legal limit at admission in three Level 1 trauma centers (Kobayashi et al., 2019): 5% at emergency departments in Southern California (Trivedi et al., 2019), 12% in Washington D.C. (Cicchino et al., 2021), and 29% self-reported as drinking alcohol in the 12 hours preceding their injury in Austin, TX (Austin Public Health, 2019). An earlier comparison of e-scooter and cyclist injuries in Hamburg, Germany by Kleinertz et al. (2021) suggested that alcohol played a greater role in e-scooter injuries (28%) than cycling injuries (6%); however, another Canadian study found that alcohol involvement in nonfatal crashes by bicyclists was as high as 14.5% (Asbridge et al., 2014).

3. Materials

Without a national system to record and track e-scooter deaths, it is difficult to collect data on shared e-scooter fatalities. NHTSA's Fatality Analysis Reporting System (FARS) collects data on motor-vehicle traffic fatalities, but all these fatalities must include a motor vehicle, which not all e-scooter fatalities do. In addition, NHTSA's FARS criteria limit its police-reported fatality data to deaths occurring on public highways (i.e., on a publicly owned road, which excludes driveways or private parking lots) and deaths occurring within 30 days of the initial crash (NHTSA, 2021a). E-scooter users are accounted for under the FARS schema as "users of personal conveyance," which also includes any non-motorized, non-pedaling modes like skateboards and roller blades (NHTSA, 2020a). However, e-scooters are not identified any more specifically in the raw data available.

In order to create an e-scooter fatality data set, we started with the lists of e-scooter fatalities compiled by other researchers, such as the Collaborative Sciences Center for Road Safety at the University of North Carolina Chapel Hill (2021), a National Broadcasting Company (NBC) list of fatal micromobility crashes (Fleischer, Yarborough & Jones, 2019), and a list of international e-scooter incidents by Quartz (Griswold, 2020). We limited our analysis specifically to fatalities involving a shared, standing e-scooter in 2018 and 2019. Many incidents listed by other sources included privately owned or seated devices, which we removed from our dataset.

We used the identifying information of location, time, and victim age/gender to match each e-scooter fatality to its corresponding record in FARS. We flagged these records as e-scooter fatalities in our dataset. This allowed us to compare e-scooter fatalities alongside the other vehicle modes in FARS. All but three of 20 fatalities meeting our criteria had a corresponding record in FARS. Those three fatalities were not recorded in FARS because no motor vehicle was deemed to be involved. This final list of fatalities and the news source identifying them as e-scooter related, are listed

Table 1
E-Scooter fatalities involving motor vehicles in 2018–2019.

Victim Age (Sex)	Date of Crash	Source
20 (M)	09/21/18	Cho, DiMargo, and Swalec, 2018
26 (M)	12/22/18	Cervantes and Stickney, 2018
21 (M)	02/01/19	CBS Austin, 2019
27 (M)	04/11/19	Lohrmann, 2019
31 (M)	04/13/19	Cosgrove, 2019
5 (M)	04/23/19	Griggs & Lou, 2019
20 (M)	04/23/19	Prince, 2019a
26 (M)	05/16/19	Alund, 2019
33 (M)	06/20/19	Marrero, 2019
34 (F)	07/27/19	Pozen, 2019
37 (M)	07/17/19	Prince, 2019b
45 (M)	08/06/19	Jones, 2019
26 (M)	08/04/19	Sukut, 2019
28 (M)	10/09/19	KHQ Q6, 2019
16 (M)	10/27/19	Carmel, 2019
16 (M)	11/20/19	News 12, 2019
35 (M)	11/04/19	Stunson, 2019

Note: In this table, male is signified by (M) and female is signified by (F).

in Table 1 E-Scooter Fatalities Involving Motor Vehicles in 2018–2019.

In our analysis, we considered fatalities associated with each mode to be fatalities in which a person using that mode was a fatality: when a crash appears in the "Passenger Car" row, the crash involved a "Passenger Car" occupant (driver or passenger) who died as a result of the crash. If the Passenger Car instead hit and killed a pedalcyclist (a cycle powered by pedaling), the fatality would be in the Pedalcyclist row. If a Passenger Car hit a cyclist and both the cyclist and an occupant of the car died, it counted as a fatality for both modes.

As points of comparison, in addition to conventional motor vehicles like passenger cars and motorcycles, and conventional VRUs like pedestrians and pedalcyclists, this analysis includes modes of motorist and non-motorist transport that may share similarities with e-scooters: "moped/motorized bicycles" is a vehicle code that includes e-bikes and mopeds; "motor scooter" refers to Vespa-style seated scooters; and, as previously mentioned, "personal conveyance" is a catch-all for any wheeled non-pedaling transportation, such as skateboards, roller blades, and wheelchairs (NHTSA, 2020a). The category for Passenger Cars includes conventional motor vehicles, but not pickup trucks, vans, or anything bigger (NHTSA, 2020a). We used these categories to further group vehicles in a variety of ways. The National Safety Council (2018) defines vulnerable road users (VRUs) as anybody in or near a trafficway who is not inside an enclosed vehicle. In our analysis, this includes pedestrians, pedalcyclists, and users of e-scooters, personal conveyances, motor scooters, mopeds, and motorcycles (i.e., every category except for passenger cars). Another category is vehicles that NHTSA defines as "Motor Vehicles," which includes passenger cars, motor scooters, mopeds, and motorcycles. NHTSA does not consider e-scooters or personal conveyances to be motor vehicles, even though they often have motors, so in this analysis, "unmotorized vehicles" refers to pedestrians, pedalcyclists, e-scooter users, and people on personal conveyances, even though the latter category may have motors.

4. Methodology

Pearson's Chi-Squared Test of Independence was the primary statistical test used in this analysis. However, the accuracy of the chi-squared test relies on the expectation of at least five data points per category, a criterion that was not always met due to the low number of data points for e-scooters. In these situations, Fisher's Exact Test of Independence was used in place of chi-square for its ability to handle small sample sizes, similar to Shah, Aryal, Wen, and Cherry (2021). These tests were conducted using chisq.test() and fisher.test() from the R 'stats' library (R Core Team, 2013). The smallest p-value that can be calculated using these functions is $p < 2.2 \times 10^{-16}$. Additionally, because this analysis involved performing many different statistical tests on the same dataset, Bonferroni's Correction was applied to the significance level to reduce the likelihood of producing an erroneously significant result through random chance (Weisstein, n.d.). Since 15 statistical tests were performed in this analysis, a significance level of 0.05/15, or 0.0033, was used for all tests.

5. Results

5.1. Victim demographics

5.1.1. Gender

In the NHTSA 2018–2019 data, the lowest proportion of male fatalities was in passenger cars (61% male), followed by pedestrians (70% male). In contrast, all but one e-scooter fatalities in the data were male (94%, see Table 2). This same gender ratio is only approached among motorcyclist fatalities (91% male) and pedalcyclists (86%), as illustrated in Fig. 1 Gender of Victim by Mode of Transportation (see Table 3).

A Pearson's Chi-Squared Test of Independence was conducted comparing e-scooters to other VRUs (all categories except Passenger Cars), which yielded a $p = 0.24$, suggesting that the percentage of male victims using e-scooters is not significantly different from other VRUs. However, pedestrian victims are significantly less likely to be male (Chi-Squared Test $p < 2.2 \times 10^{-16}$) compared to other VRUs.

5.1.2. Age

The age distribution of e-scooter fatalities is particularly notable for the young age of the victims: all but one e-scooter fatality in the 2018–2019 data was under the age of 40. The cumulative distribution is strikingly convex compared to the distribution of other modes (Fig. 2 Cumulative Distribution of Traffic Fatalities by Age). While motorcycle deaths are sometimes associated with young men, only 46% of motorcyclist fatalities are between the ages of 20–40. E-scooter victims are significantly more likely (Fisher's Exact Test $p = 0.0006$) to be between those ages even when compared to motorcyclists, who have the next greatest share of fatalities aged 20–40.

The young slant in casualties is illustrated as a cumulative distribution in Fig. 2 Cumulative Distribution of Traffic Fatalities by Age. Passenger car and pedestrian deaths are spread uniformly

Table 2
Traffic fatalities by gender.

Victim's Mode	Female	Male
E-scooter	5.9% (n = 1)	94% (n = 16)
Moped/Motorized Bicycle	11% (n = 16)	89% (n = 128)
Motor Scooter	11% (n = 42)	89% (n = 339)
Passenger Car	39% (n = 10099)	61% (n = 15650)
Pedalcyclist	14% (n = 235)	86% (n = 1480)
Pedestrian	30% (n = 3818)	70% (n = 8809)
Personal Conveyance	20% (n = 65)	80% (n = 268)
Regular Motorcycle	8.6% (n = 796)	91% (n = 8422)

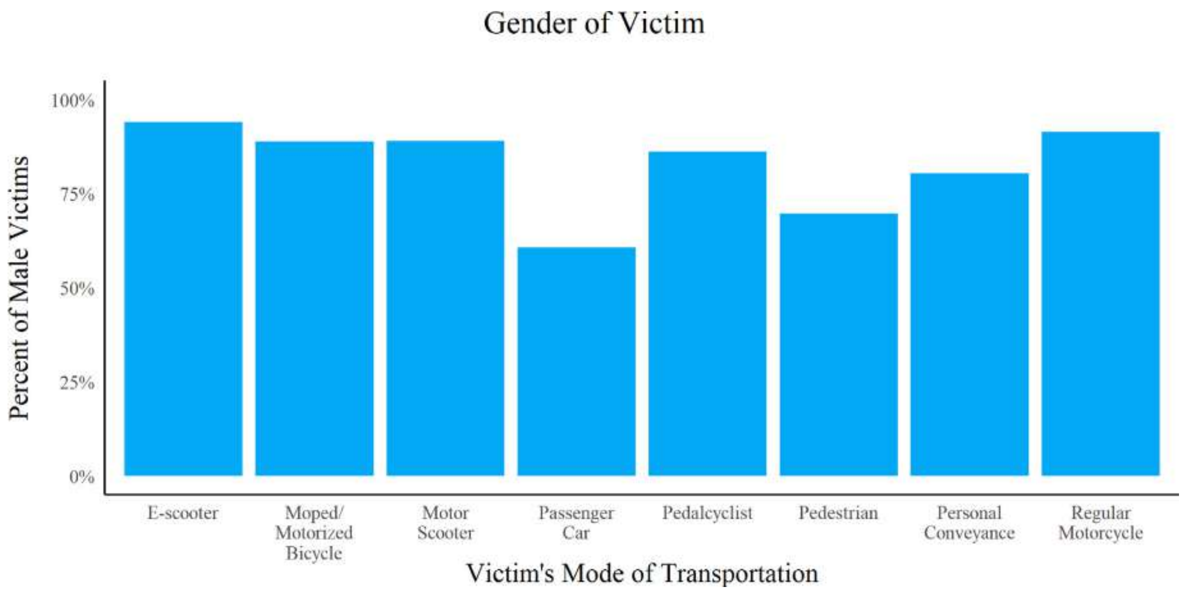


Fig. 1. Gender of victim by mode of transportation.

Table 3
Traffic fatalities by age group.

Victim's Mode	Age 0–19	Age 20–40	Age 41+
E-scooter	18% (n = 3)	76% (n = 13)	5.9% (n = 1)
Moped/Motorized Bicycle	5.6% (n = 8)	33% (n = 47)	62% (n = 89)
Motor Scooter	4.2% (n = 16)	29% (n = 112)	66% (n = 253)
Passenger Car	12% (n = 3186)	42% (n = 10752)	46% (n = 11811)
Pedalcyclist	10% (n = 171)	26% (n = 442)	64% (n = 1092)
Pedestrian	6.2% (n = 783)	32% (n = 4036)	62% (n = 7753)
Personal Conveyance	16% (n = 53)	14% (n = 48)	70% (n = 231)
Regular Motorcycle	2.8% (n = 262)	46% (n = 4214)	51% (n = 4742)

across age groups. Other modes, like golf carts and three-wheeled motorcycles, lean heavily older in comparison, with a minority of fatalities under the age of 40.

5.2. Crash circumstances

5.2.1. Alcohol involvement

Rates of alcohol involvement among fatally injured e-scooters users was high (41%) even compared to the next highest, pedestrians (30%) and motorcyclists (29%). These rates are illustrated in Fig. 3 Alcohol Involvement by Mode and Role in Fatal Crash. However, Fisher's Exact Test of Independence comparing the rate of impairment in e-scooter users to other VRUs yields $p = 0.818$, suggesting that the difference is not statistically significant. However, this may be attributable to the small sample size. The involvement of another party (a non-fatality driving a motor vehicle) under the influence was low in comparison to the rate of alcohol involvement among victims (see Table 4 and Table 5).

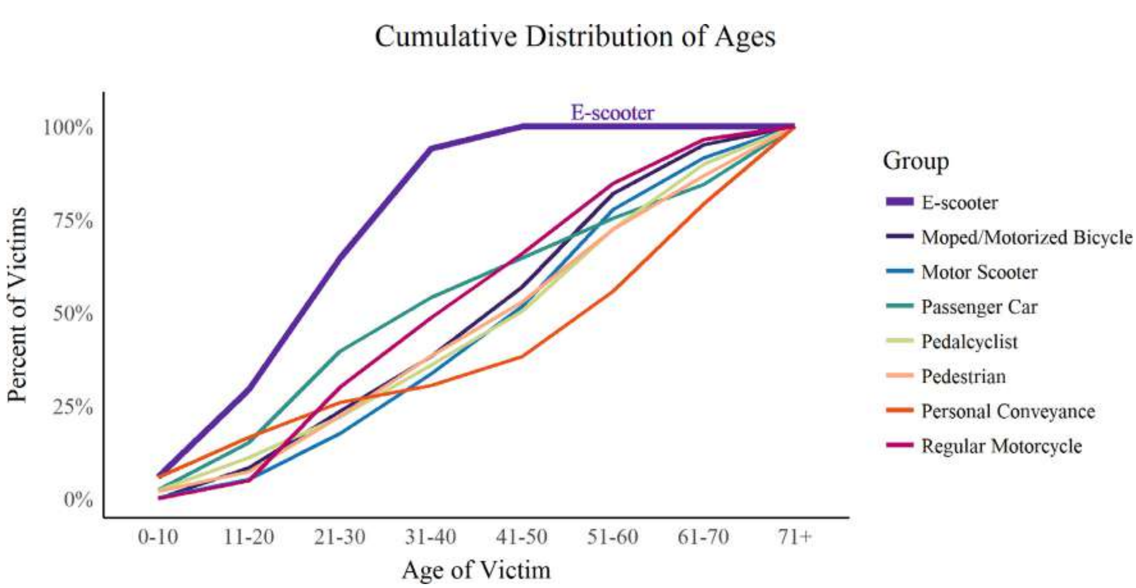


Fig. 2. Cumulative distribution of traffic fatalities by age.

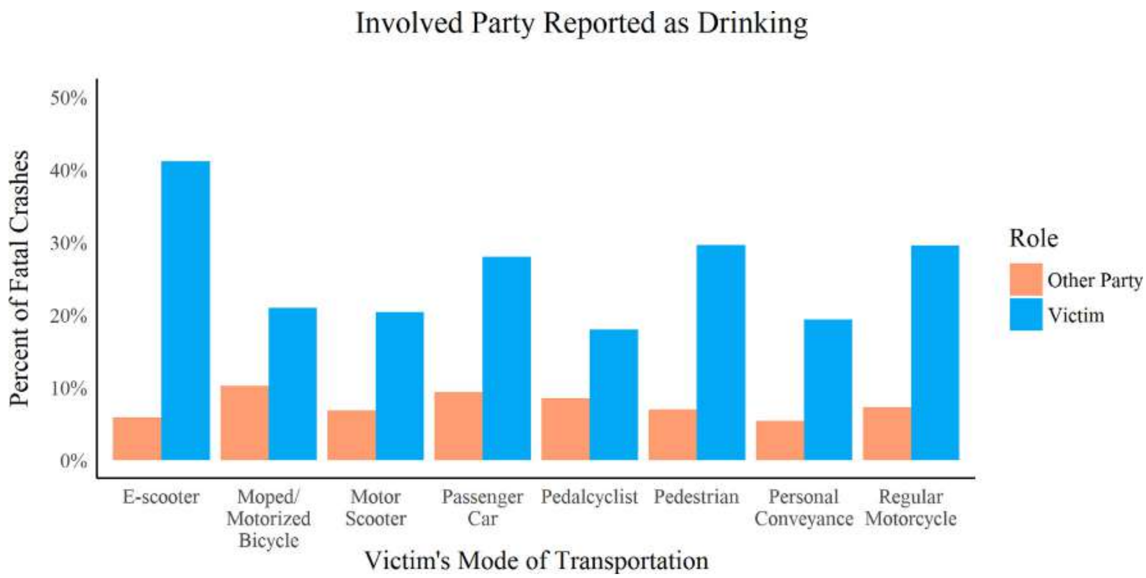


Fig. 3. Alcohol involvement by mode and role in fatal crash.

Table 4
Fatalities where another party, other than the victim/victim's driver, had been drinking.

Victim's Mode	Alcohol Involvement in Other Party	No Alcohol Involvement	Unknown
E-scooter	5.9% (n = 1)	29% (n = 5)	65% (n = 11)
Moped/Motorized Bicycle	10% (n = 11)	14% (n = 15)	76% (n = 81)
Motor Scooter	6.9% (n = 19)	16% (n = 45)	77% (n = 213)
Passenger Car	9.4% (n = 1221)	18% (n = 2315)	73% (n = 9445)
Pedalcyclist	8.5% (n = 145)	19% (n = 328)	72% (n = 1237)
Pedestrian	7% (n = 863)	18% (n = 2260)	75% (n = 9206)
Personal Conveyance	5.5% (n = 18)	18% (n = 61)	76% (n = 251)
Regular Motorcycle	7.3% (n = 387)	20% (n = 1035)	73% (n = 3878)

5.2.2. Daylight conditions

E-scooter fatalities occur at night in the dark more than any other mode of transportation (see Table 6 and Fig. 4 Percent of Fatal Crashes Occurring at Night by Mode). However, when compared to each other using Fisher's Exact Test of Independence, the rate of nighttime fatalities between pedestrians and e-scooter riders are not significantly different ($p = 0.78$). Pearson's Chi-Squared Test of Independence was conducted, comparing e-scooters and pedestrians combined to all other VRUs (mopeds, motor scooters, pedalcyclists, and motorcycles), which yielded $p < 2.2 \times 10^{-16}$, suggesting that pedestrians and people on e-scooters are both more likely to be killed at night compared to other VRUs.

5.2.3. Hit-and-run crashes

Although the e-scooter fatalities in the sample had a higher absolute percentage of hit-and-runs (24%) than any other mode (Table 7), they were not significantly different from the hit-and-run rates for pedalcyclists (18%) and pedestrians (20%) (Fisher's Exact Test $p = 0.031$) at the significance level of 0.0033 required by Bonferroni's Correction. However, the rates of hit-and-runs *are* significantly higher for unmotorized VRUs (pedestrians, bicycles, e-scooters, and personal conveyance) than for motorized VRUs (mopeds, motor scooters, and motorcycles) (Fisher's Exact Test $p < 2.2 \times 10^{-16}$) (see Fig. 5).

Table 5
Fatalities where victim (or victim's driver) had been drinking.

Victim's Mode	Alcohol Involvement in Victim	No Alcohol Involvement	Unknown
E-scooter	41% (n = 7)	29% (n = 5)	29% (n = 5)
Moped/Motorized Bicycle	20% (n = 29)	34% (n = 49)	45% (n = 65)
Motor Scooter	20% (n = 75)	34% (n = 127)	46% (n = 171)
Passenger Car	27% (n = 6085)	37% (n = 8294)	36% (n = 8097)
Pedalcyclist	18% (n = 306)	42% (n = 722)	40% (n = 683)
Pedestrian	30% (n = 3674)	34% (n = 4256)	36% (n = 4482)
Personal Conveyance	19% (n = 64)	38% (n = 125)	43% (n = 141)
Regular Motorcycle	29% (n = 2602)	39% (n = 3460)	32% (n = 2818)

5.3. Crash location

5.3.1. Intersections. Half of the fatal e-scooter crashes identified in FARS were either at an intersection or intersection-related (see Table 8 Junction and intersection-related crashes). Rates of e-scooter (47%) and pedalcycle (38%) crashes are not significantly different (Fisher's Exact Test $p = 0.31$) from each other. However, the combined rate of fatal pedalcycle and e-scooter intersection-related crashes is significantly higher (Chi-Squared Test $p < 2.2 \times 10^{-16}$) than that for pedestrians (26%). This distinction makes

Table 6
Traffic fatalities by daylight conditions.

Victim's Mode	At Dusk	In Daylight	In the Dark	At Dawn	Unknown
E-scooter	5.9% (n = 1)	12% (n = 2)	82% (n = 14)	0% (n = 0)	0% (n = 0)
Moped/Motorized Bicycle	0.7% (n = 1)	40% (n = 57)	57% (n = 82)	2.1% (n = 3)	0% (n = 0)
Motor Scooter	2.4% (n = 9)	51% (n = 192)	45% (n = 166)	1.6% (n = 6)	0% (n = 0)
Passenger Car	2% (n = 453)	51% (n = 11446)	45% (n = 10081)	2.1% (n = 484)	0.48% (n = 109)
Pedalcyclist	1.8% (n = 31)	47% (n = 810)	48% (n = 827)	2.1% (n = 36)	0.41% (n = 7)
Pedestrian	1.9% (n = 236)	20% (n = 2516)	75% (n = 9367)	1.8% (n = 222)	0.57% (n = 71)
Personal Conveyance	3.9% (n = 13)	37% (n = 121)	58% (n = 191)	0.91% (n = 3)	0.61% (n = 2)
Regular Motorcycle	3.8% (n = 342)	58% (n = 5140)	37% (n = 3252)	1.2% (n = 104)	0.53% (n = 47)

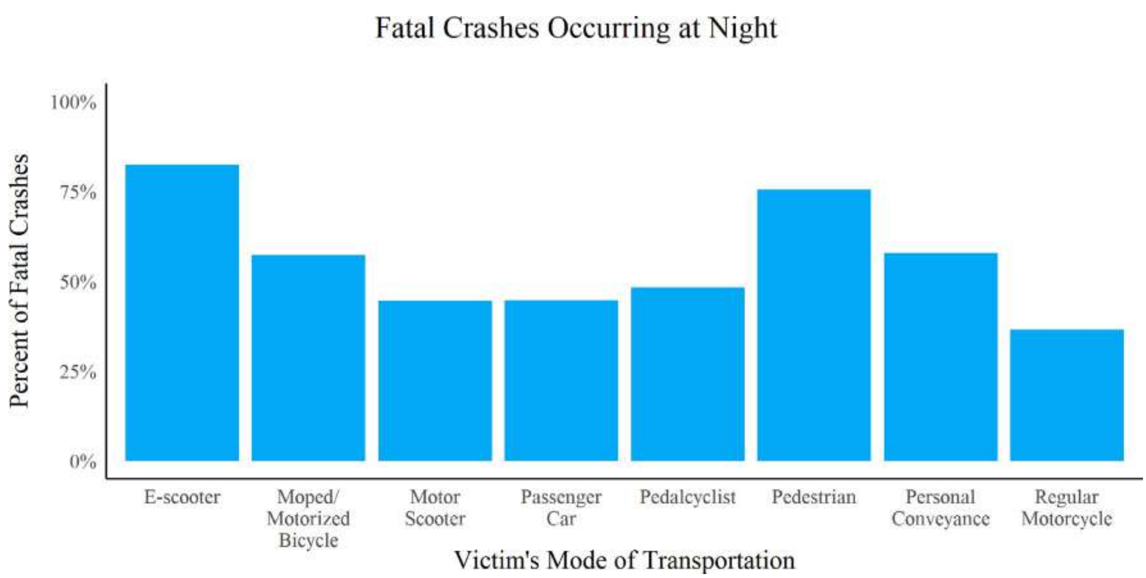


Fig. 4. Percent of fatal crashes occurring at night by mode.

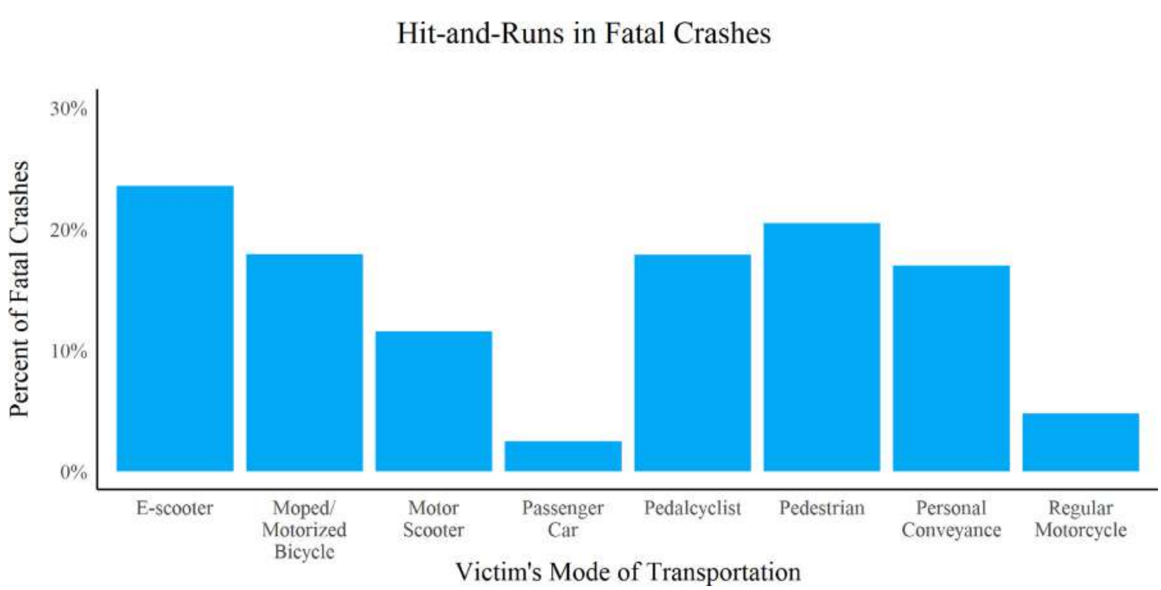


Fig. 5. Rate of hit-and-runs in fatal crashes.

Table 7
Rate of hit-and-runs in fatal crashes.

Victim's Mode	Not Hit-and-Run	Hit-and-Run
E-scooter	76% (n = 13)	24% (n = 4)
Moped/Motorized Bicycle	82% (n = 87)	18% (n = 19)
Motor Scooter	88% (n = 245)	12% (n = 32)
Passenger Car	97% (n = 12666)	2.5% (n = 326)
Pedalcyclist	82% (n = 1405)	18% (n = 306)
Pedestrian	80% (n = 9868)	20% (n = 2544)
Personal Conveyance	83% (n = 274)	17% (n = 56)
Regular Motorcycle	95% (n = 5046)	4.8% (n = 254)

sense: unlike pedestrians, e-scooter riders and pedalcyclists often navigate intersections with the flow of traffic, so the traffic risks they experience would be similar.

Related to the high rate of intersection-related crashes, many of the e-scooter crashes analyzed involved traffic control. While none of them involved a stop sign, they did involve traffic signals at almost triple the rate for pedestrians and pedalcyclists (as shown in Table 9). Fisher's Exact Test yields $p = 0.005$ when comparing the rate of e-scooter fatalities at traffic signals to that for pedestrians and pedalcyclists, which is not statistically significant at the more conservative alpha-level of 0.0033 as required by Bonferroni's multiple testing correction. Further data and analysis may be able to more definitively confirm or deny this connection.

5.3.2. Crosswalks. E-scooter and personal conveyance user fatalities were roughly twice as likely to have been killed at a crosswalk than a pedestrian (Chi-Squared Test $p = 3.39 \times 10^{-4}$), and nearly-three to four times more likely to have been killed in a crosswalk than a pedalcyclist (Chi-Squared Test $p = 1.95 \times 10^{-10}$) (see Table 10). As an additional measure to control for differences in crosswalk use between urban and rural areas, this comparison was made only in crashes that NHTSA classified as in urban areas. They define this as census areas of 50,000 or more people.

5.3.3. Point of impact. The most likely point of impact for a fatal crash, across all modes, is the front of the other motor vehicle. This is followed by an impact to the right side for fast moving VRUs. More powerful, motorized VRUs have elevated rates of rear and left side impacts, such as motorcycles (20% and 16%, respectively) and motor scooters (10% and 9%). By combining motorcycles and motor scooters into a single group and performing a Chi-Squared Test, we see that this high rate of right-side impacts is significantly high compared to other VRUs ($p < 2.2 \times 10^{-16}$). Table 11 lists the point of impact to the other vehicle in crashes that involved at least one other motor vehicle (e.g., single-vehicle motorcycle accidents are not represented in the table).

VRUs experience right-side impact crashes at different rates. E-scooters and mopeds combined appear to have a higher rate of right-side impacts than other small micromobility (pedestrians, pedalcyclists, and personal conveyance), but the difference (Chi-Squared Test $p = 0.018$) is not statistically significant with $\alpha = 0.0033$. We also note that the VRUs in this data appear vulnerable to right-side-impact (sometimes referred to as "right hook") crashes in the same approximate order of speed: motorcycles, then motor scooters, then e-scooters, then mopeds/motorized bicycles, then pedalcycles, then personal conveyance, and then pedestrians.

5.4. Characterization of factors. The characteristics of e-scooter fatalities are not a clear or precise match to the characteristics of pedestrians, pedalcyclists, or motorcyclists (see Table 12). E-scooters share some characteristics, but not all, with each mode. E-scooters share crash circumstances (likelihood to be struck in dark conditions, hit-and-runs, and for the victim to be intoxicated with alcohol) most closely with pedestrians, but demographically, e-scooter victims stand alone. The closest demographically similar mode is motorcyclists (young and predominantly male) but e-scooters are significantly more likely to be under 40 years of age. Motorcyclists also share with e-scooters and pedestrians a high rate of alcohol involvement among their fatalities. The relationship between e-scooter and pedalcyclist fatalities is also unique. E-scooter and cyclist fatalities are similarly likely to be male, hit-and-run, and at an intersection, but pedalcyclist fatalities are much less likely to be under 40, intoxicated by alcohol, killed at night, or struck in a crosswalk.

6. Discussion

6.1. Crash demographics

As discussed, males typically have historically made up the majority of traffic fatalities (Chang, 2008). Certain modes, like motorcycles and pedalcycles, have historically seen males compose 80–90% or more of their fatalities. This strong majority is in line with the results we have found for e-scooters, of 94% male. While this is more imbalanced than either e-scooter injury studies or e-scooter user surveys have found, this is not necessarily a contradic-

Table 8
Junction and intersection-related crashes.

Victim's Mode	Driveway Related	Entrance/Exit Ramp	Intersection Related	No Junction
E-scooter	5.9% (n = 1)	5.9% (n = 1)	47% (n = 8)	41% (n = 7)
Moped/Motorized Bicycle	7.7% (n = 11)	0% (n = 0)	36% (n = 51)	57% (n = 81)
Motor Scooter	6.2% (n = 23)	0.8% (n = 3)	44% (n = 165)	48% (n = 179)
Passenger Car	2.9% (n = 651)	1.9% (n = 427)	26% (n = 5777)	67% (n = 15118)
Pedalcyclist	3.7% (n = 64)	0.41% (n = 7)	38% (n = 654)	56% (n = 961)
Pedestrian	2.1% (n = 261)	1.4% (n = 171)	26% (n = 3236)	68% (n = 8409)
Personal Conveyance	6.4% (n = 21)	0% (n = 0)	45% (n = 148)	48% (n = 159)
Regular Motorcycle	8.2% (n = 727)	3.2% (n = 285)	34% (n = 3027)	52% (n = 4661)

Table 9
Traffic controls present at fatal crashes.

Victim's Mode	No Control	Other/Unknown Control	Traffic Signal	Stop Sign
E-scooter	53% (n = 9)	5.9% (n = 1)	41% (n = 7)	0% (n = 0)
Moped/Motorized Bicycle	73% (n = 115)	4.5% (n = 7)	15% (n = 23)	7.6% (n = 12)
Motor Scooter	69% (n = 288)	4.8% (n = 20)	16% (n = 67)	10% (n = 43)
Passenger Car	76% (n = 18747)	6% (n = 1495)	9.3% (n = 2315)	9% (n = 2218)
Pedalcyclist	74% (n = 1270)	3.3% (n = 56)	17% (n = 299)	5.2% (n = 89)
Pedestrian	81% (n = 10046)	4.6% (n = 571)	13% (n = 1632)	1.6% (n = 204)
Personal Conveyance	73% (n = 240)	2.7% (n = 9)	19% (n = 64)	5.2% (n = 17)
Regular Motorcycle	74% (n = 7222)	6.2% (n = 611)	10% (n = 1023)	9.4% (n = 923)

Table 10
Locations of non-motorists in fatal crashes.

Victim's Mode	Crosswalk	Intersection	Travel Lane
E-scooter	24% (n = 4)	18% (n = 3)	59% (n = 10)
Pedalcyclist	9.7% (n = 130)	22% (n = 292)	63% (n = 841)
Pedestrian	15% (n = 1564)	4.4% (n = 450)	73% (n = 7453)
Personal Conveyance	22% (n = 67)	13% (n = 38)	62% (n = 173)

tion with the literature. Previous research on sex ratio of injuries and fatalities in transportation has established that transportation injuries often have a more balanced sex ratio than transportation fatalities (Santamariña-Rubio et al., 2014).

One possible reason for young men to be overrepresented in e-scooter fatalities is that e-scooter users are generally young, typically between the ages of 16 to 35 (Laa & Leth, 2020) and traffic fatalities in this age group are overwhelmingly male (Chang, 2008). Traffic fatalities under the age of 16 or over the age of 40 exhibit less of a gender disparity (Chang, 2008). One underlying explanation for this is that both age and gender have a substantial

Table 11
Point of impact to other party's vehicle.

Victim's Mode	Front	Rear	Right Side	Left Side	Other/Unknown
E-scooter	78% (n = 14)	5.6% (n = 1)	11% (n = 2)	0% (n = 0)	5.6% (n = 1)
Moped/Motorized Bicycle	78% (n = 86)	3.6% (n = 4)	8.2% (n = 9)	4.5% (n = 5)	5.5% (n = 6)
Motor Scooter	64% (n = 181)	9.9% (n = 28)	14% (n = 40)	9.2% (n = 26)	2.5% (n = 7)
Passenger Car	75% (n = 10849)	12% (n = 1752)	4.1% (n = 601)	6.9% (n = 1001)	2.1% (n = 299)
Pedalcyclist	80% (n = 1410)	2.4% (n = 42)	6.4% (n = 113)	3.8% (n = 67)	7.1% (n = 125)
Pedestrian	80% (n = 10576)	3.4% (n = 454)	3.5% (n = 469)	2.7% (n = 355)	11% (n = 1421)
Personal Conveyance	85% (n = 285)	1.5% (n = 5)	3.6% (n = 12)	2.7% (n = 9)	7.2% (n = 24)
Regular Motorcycle	40% (n = 2301)	20% (n = 1167)	19% (n = 1129)	16% (n = 930)	5.1% (n = 296)

Table 12
Summary of E-scooter variables.

Category	Factor	E-Scooter Rate	Similar to	Higher than the rates for
Demographics	% Male	94%**	Pedalcyclists, Motorcyclists	Pedestrians, Passenger Cars
	% Between ages of 20–40	76%**	None	Personal Conveyance, Pedestrians
Circumstances	% Alcohol involvement among fatalities	41%	Pedestrians, Motorcyclists	Pedalcyclists
	% In dark conditions	82%**	Pedestrians	Motorcyclists, Pedalcyclists
	% Hit-and-run	24%**	Pedestrians, Pedalcyclists, Personal Conveyance	Passenger Cars, Motorcycles
Location	% At intersection	47%**	Pedalcyclist, Personal Conveyance	Pedestrian
	% At traffic signal	41%*	None	Passenger Cars
	% Fatality in crosswalk	24%**	Personal conveyance	Pedestrians, Pedalcyclists
Impact	% Struck by front of car	78%	All VRUs	Motorcycles
	% Struck by right side of car	11%*	Mopeds	Pedestrians, Pedalcyclists

* Statistically Significant at $\alpha = 0.05$.
** Statistically Significant at $\alpha = 0.0033$.

impact on risk-taking behavior, with men and younger drivers associated with higher risk (Turner & McClure, 2003).

6.2. Crash circumstances

Dark conditions may have a confounding effect with alcohol involvement and with the likelihood of hit and run crashes. Previous research has found that most hit and runs (78%) occur at night (Benson et al., 2021) which is also when drivers are most likely to leave the scene of a crash (Solnick & Hemenway, 1995). Research has also suggested drivers are more likely to leave the scene if they are impaired, which may play a role in the elevated rate of hit-and-runs at night and on weekends, and among drivers with previous arrests for driving while intoxicated (Solnick & Hemenway, 1994). A 2008 study by Tay et al. (2009) found that drivers are much more likely to leave the scene of a fatal crash if the victim is a pedestrian; the researchers suggested that drivers may be motivated by a belief they will not be apprehended because pedestrian collisions are less noticeable than two-car collisions. The similar rate of hit-and-runs between pedestrians, personal conveyance, and e-scooter users, suggests that this effect might extend to the latter groups. Because the likelihoods of all three variables are not independent from each other, this effect is only speculative. However, these preliminary results still suggest some points of interest.

Pedestrian and e-scooter fatalities alone had a supermajority occur in dark conditions. Nighttime visibility may be especially important to them because they may lack a sufficiently powerful light source to see or be seen easily in the dark. In contrast, larger VRUs such as motorcycles (37%) or motor scooters (45%) may be easier to see at night because of their size or because of an increased amount of lighting present on the vehicle. However, these may also be confounded by the underlying rates of exposure, which could be very different by mode.

6.3. Crash locations and configurations

E-scooter riders are notable for their hybridization (i.e., their ability to switch between pedestrian and cyclist behavior) and their fatalities so far reflect this pattern. Both fatalities where the e-scooter rider was acting like a pedestrian (e.g., killed in crosswalk), and fatalities where the e-scooter rider was acting like a cyclist (e.g., killed by a right-hook) have occurred in the data. In this way, e-scooter riders unite some of the most dangerous characteristics of both pedestrian and cyclist modes of transportation. Like pedestrians, e-scooter riders have a small visual profile and may be particularly difficult to see at night.

Like cyclists, e-scooter riders are expected to travel on the shoulder of the road in most U.S. states (Unagi, 2021) and maneuver with traffic. If an e-scooter rider is not spotted by a motor vehicle (and they may be traveling in the motor vehicle's blind spot or be otherwise difficult to detect), they are in danger of being hit from behind (if the two trajectories overlap) or experiencing a right-impact crash if the motor vehicle traveling alongside them makes a right-turn into the e-scooter rider's path. The rate of right-side-impact (''right hook'') crashes was also elevated, albeit non-significantly in the e-scooter available. Both traffic signals and ''right hook'' crash risk may be important areas for future research.

Half of the fatal e-scooter crashes identified in FARS were either at an intersection or intersection-related. Since all these cases involved a motor-vehicle collision, this is consistent with the findings of Cicchino, Kulie, and McCarthy (2021) that most motor-vehicle and e-scooter collisions occurred at intersections. Shah et al. (2021) also found that most of their sample of e-scooter crashes that involved a motor vehicle occurred at intersections, and that this rate was similar for their bicycle crashes. Our findings for fatal crashes were similar, although the overall proportion of intersection-related crashes was lower.

One possible complication to the interaction between e-scooters and motor vehicles at an intersection is their unusually large difference in travel speed, also referred to as ''closing'' speed. A high closing speed means the motor vehicle overtakes the second party more quickly. This provides a smaller window of opportunity for the first driver to recognize and react if their line of travel will put them in conflict. In these cases, there is less time for the car driver to recognize and react to their presence, even if they are in the driver's path. Low visibility conditions, such as nighttime, could exacerbate this risk. In contrast, motorcycles often travel at speeds closer to cars, giving the driver more time to react. These differences in speed may also make it difficult for drivers to correctly assess the speed of the other party and respond appropriately. In addition, the nuances of predictable behavior vary by mode: while sidewalk and crosswalk use are expected for pedestrians, drivers may find it unexpected behavior from e-scooters or cyclists (Sumner, 2016). The elevated rate of e-scooter users and people on personal conveyance fatally struck at crosswalks may be a complex result of this effect.

9.4. Limitations

This research is limited by the available data in several ways. First, the sample size of known e-scooter fatalities involving motor vehicles is small. This means that only the most obvious patterns can be tested and be expected to show statistical significance. Local differences or regional subtleties cannot be meaningfully examined with the limited national data. It also means that, in many cases, it is more useful to group e-scooters with a similar mode of transportation to more clearly establish a pattern. This data also only involves fatalities where a motor vehicle was involved *and* that were identified in media sources. Not all e-scooter deaths involve a motor vehicle or otherwise meet the criteria to be included in NHTSA's FARS data. It is also likely that not all e-scooter fatalities were publicized by the media. In that case, since we relied on media reports to identify and label e-scooter fatalities, it is possible that additional e-scooter fatalities were recorded in the FARS data that we did not identify. These would have been therefore misclassified as a different type of non-motorist, most likely as a person on personal conveyance or else as a pedestrian. Any other limitations or inaccuracies from the NHTSA FARS data would also have been propagated forward into our data and our analysis in this paper. The most pertinent risk from this is the possibility that different modes have different degrees of accuracy or reporting. For example, e-scooter fatalities may have undergone more scrutiny and more testing for alcohol involvement than pedestrian fatalities, leading to a greater rate of discovery in the e-scooter fatalities. This could cause an apparent difference in the rate of alcohol involvement as an artifact of the reporting. However, the current rates of alcohol involvement were not significantly different between comparable modes.

Additionally, without detailed usage data, it is difficult to speculate about normalized risk per mile. These comparisons are limited to looking at the proportion of fatalities sharing similar characteristics or circumstances, rather than adjusting for exposure. For example, it may be a coincidence that pedestrians and e-scooters share a similar proportion of fatalities occurring at night, because it is possible that a confounding factor like usage is disguising the real effect, for example, pedestrians could have twice the risk at night per mile, but travel half as many miles as e-scooters: the effects would cancel out. This limits the strength of any speculation about the causes of these differences being inherent to the mode until more precise and larger datasets become available.

10. Conclusions

Compared to other modes of transportation, e-scooter rider fatalities are unusually likely to be young and male. In these demographics, they are most similar to motorcyclists, but e-scooters are still significantly more likely to be under the age of 40. The crash circumstances of e-scooters (such as dark conditions, hit-and-run, and victim alcohol intoxication) were statistically similar to pedestrians. However, the locations of e-scooter crashes are less analogous to existing modes. No other mode has a similarly high rate of intersection and crosswalk fatalities. E-scooter fatalities also had an elevated rate of traffic signals present at the crash and right-side impacts to the motor vehicles, although neither were significant at the conservative threshold of $\alpha = 0.0033$.

11. Practical Applications

The shared e-scooter is a unique and evolving mode of transportation. While existing modes of transportation such as motor vehicles, pedalcyclists, and pedestrians are well-understood and have a wealth of established safety literature surrounding them, these lessons do not necessarily transfer to a new mode with different characteristics. E-scooter user fatalities have important characteristics in common with motorcyclist, pedestrian, and pedalcyclist fatalities, and the evidence suggests that they may share important risk factors. By analyzing e-scooter fatalities in comparison with these more common and well-researched modes, we gain further insight into the sources of the safety issues and what best practices or policy interventions borrowed from other modes may be useful for e-scooters.

It is important to stress to users and policymakers that e-scooters are a distinct mode of transportation. The circumstances of fatal e-scooter collisions have much in common with pedestrians and pedalcyclists. Yet, e-scooter fatalities are younger than either group. E-scooter safety campaigns may wish to focus on the demographics most at risk for fatalities, particularly young men. In other areas, such as the rates of crashes in intersections and crosswalks, e-scooters and personal conveyance users appear distinct. Evidence-based safety solutions must reconcile these distinctions and consider the sources of e-scooter fatality risk before these sources of risk can be mitigated by policy interventions.

It may not be safe to assume that the same policies designed to protect pedestrians will necessarily reduce e-scooter fatalities as well: for example, the elevated rate of e-scooter crosswalk fatalities raises questions about the safety of e-scooter crosswalk use. Other lessons from pedestrian and cycling safety appear better suited to transfer to this new domain. For example, the importance of nighttime visibility and sobriety suggest themselves as continuing themes that apply similarly to both e-scooters and pedestrians.

Funding

This research was financially supported by the MITRE Innovation Program, an independent research and development program run by the MITRE Corporation, a not-for-profit systems engineering firm that manages federally funded research and development centers (FFRDCs) for multiple U.S. government agencies.

Conflict of 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

The authors would like to thank Dr. Valerie Gawron for her continued insight and guidance and Claudette Bishop for assistance with revisions.

References

Abcarian, R. (2018). Column: The Bird electric scooter conundrum: So fun, so exhilarating, so dangerous. *Los Angeles Times*. https://www.latimes.com/local/abcarian/la-me-abcarian-bird-scooters-20180206-story.html.

Almannaa, M. H., Ashqar, H. I., Elhenawy, M., Masoud, M., Rakotonirainy, A., & Rakha, H. (2021). A comparative analysis of e-scooter and e-bike usage patterns: Findings from the City of Austin, TX. *International Journal of Sustainable Transportation, 15*(7), 571–579. https://doi.org/10.1080/15568318.2020.1833117.

Alund, N. (2019). Scooter rider who died in crash had more than twice the legal limit of alcohol in system, police report says. *The Tennessean*. https://www.tennessean.com/story/news/2019/06/21/nashville-scooter-rider-brady-gaulke-killed-had-high-alcohol-blood-level-police-says/1522338001/.

Asbridge, M., Mann, R., Cusimano, M. D., Tallon, J. M., Pauley, C., & Rehm, J. (2014). Cycling-related crash risk and the role of cannabis and alcohol: A case-crossover study. *Preventive Medicine, 66*, 80–86. https://doi.org/10.1016/j.ypmed.2014.06.006.

Austin Public Health and City of Austin. (2019). Dockless Electric Scooter Injuries Study. https://austintexas.gov/sites/default/files/files/Health/Epidemiology/APH_Dockless_Electric_Scooter_Study_5-2-19.pdf.

Badeau, A., Carman, C., Newman, M., Steenblik, J., Carlson, M., & Madsen, T. (2019). Emergency department visits for electric scooter-related injuries after introduction of an urban rental program. *The American Journal of Emergency Medicine, 37*(8), 1531–1533. https://doi.org/10.1016/j.ajem.2019.05.003.

Beck, S., Barker, L., Chan, A., & Stanbridge, S. (2020). Emergency department impact following the introduction of an electric scooter sharing service. *Emergency Medicine Australasia, 32*(3), 409–415. https://doi.org/10.1111/1742-6723.13419.

Bekhit, M. N. Z., Le Fevre, J., & Bergin, C. J. (2020). Regional healthcare costs and burden of injury associated with electric scooters. *Injury, 51*(2), 271–277. https://doi.org/10.1016/j.injury.2019.10.026.

Benson, A. J., Tefft, B. C., Arnold, L. S., & Horrey, W. J. (2021). Fatal hit-and-run crashes: Factors associated with leaving the scene S0022437521001055. *Journal of Safety Research*. https://doi.org/10.1016/j.jsr.2021.08.007.

Benson, A., Arnold, L., Tefft, B., & Horrey, W. (2018). *Hit-and-run crashes: Prevalence, contributing factors and countermeasures*. AAA Foundation https://aaafoundation.org/hit-and-run-crashes-prevalence-contributing-factors-and-countermeasures/.

BIRD. (2019). A Look at e-scooter Safety: Examining risks, reviewing responsibilities, and prioritizing prevention. https://www.bird.co/wp-content/uploads/2019/12/Bird-Safety-Report-April-2019.pdf.

BIRD. (2020). Bird Rental Agreement, Waiver of Liability, and Release. https://www.bird.co/agreement/.

Carmel, M. (2019). *Teenagers mourn 16-year-old killed in scooter crash in downtown Boise*. Idaho Press https://www.idahopress.com/news/local/teenagers-mourn-16-year-old-killed-in-scooter-crash-in-downtown-boise/article_7bc47ad9-8b30-522a-ae9f-d8cc9bb86fcb.html.

CBS Austin. (2019). Police ID 21-year-old scooter rider who died after being hit by Uber driver. *CBS Austin*. https://cbsaustin.com/news/local/police-identify-scooter-rider-hit-killed-by-lyft-driver-in-downtown.

Cervantes, E., & Stickney, R. (2018). Family Mourns Man Killed in Scooter Collision, Company Says Scooter Was Used 'Without Permission'. *NBC San Diego* https://www.nbcsandiego.com/news/local/bird-scooter-should-not-have-been-used-at-time-of-fatal-collision/166181/.

Chang, D. (2008). Comparison of Crash Fatalities by Sex and Age Group. National Highway Traffic Safety Administration's National Center for Statistics and Analysis. Report DOT HS 810 853. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/810853.

Cho, A., DiMargo, C., & Swalec, A. (2018). Man Killed Near Dupont Circle in DC's First Fatal Scooter Crash. *NBC Washington* https://www.nbcwashington.com/news/local/man-struck-by-suv-in-dupont-circle/155337/.

Cicchino, J. B., Kulie, P. E., & McCarthy, M. L. (2021). Severity of e-scooter rider injuries associated with trip characteristics. *Journal of Safety Research, 76*, 256–261. https://doi.org/10.1016/j.jsr.2020.12.016.

Cosgrove, J. (2019). Driver charged in hit-and-run death of scooter rider in Hollywood. *Los Angeles Times* https://www.latimes.com/local/lanow/la-me-ln-driver-charged-in-hit-and-run-death-of-scooter-rider-20190416-story.html.

Creaser, J. I., Ward, N. J., Rakauskas, M. E., Shankwitz, C., & Boer, E. R. (2009). Effects of alcohol impairment on motorcycle riding skills. *Accident Analysis & Prevention, 41*(5), 906–913. https://doi.org/10.1016/j.aap.2009.04.007.

Dill, J. (2019). The e-scooter gender gap. https://jenniferdill.net/2019/02/01/the-e-scooter-gender-gap/.

Electric Scooter Guide. (2021). Electric Scooter Database and Performance Testing. https://electric-scooter.guide/comparisons/compare-electric-scooters/.

Eustace, D., & Wei, H. (2010). The role of driver age and gender in motor vehicle fatal crashes. *Journal of Transportation Safety & Security, 2*(1), 28–44. https://doi.org/10.1080/19439961003590811.

Federal Highway Administration. (2017). 2017 National Household Travel Survey, U.S. Department of Transportation, Washington, DC. https://nhts.ornl.gov.
Fielding, J. (2019). E-scooters present a growing public-health challenge. *The Hill* https://thehill.com/opinion/healthcare/472401-e-scooters-present-a-growing-public-health-challenge.
Fleischer, J., Yarborough, R., & Jones, S. (2019). Cities look for ways to curb e-scooter injuries. *NBC4 Washington*. National Broadcasting Company. https://www.nbcwashington.com/investigations/cities-look-for-ways-to-curb-e-scooter-injuries/2051464/.
Griggs, B., & Lou, M. (2019). A mother whose 5-year-old son died after they rode an electric scooter together has fled to Mexico, police say. *CNN*. Cable News Network. https://www.cnn.com/2019/04/30/us/woman-flees-country-son-lime-scooter-death-trnd/index.html.
Griswold, A. (2020). At least 29 people have died in electric scooter crashes since 2018. *Quartz* https://qz.com/1793164/at-least-29-people-have-died-in-electric-scooter-crashes/.
Hansen, Z. (2019). Family of man killed in Atlanta's first e-scooter wreck wants them off roads. *The Atlanta Journal-Constitution* https://www.ajc.com/news/crime–law/family-man-killed-atlanta-first-scooter-wreck-wants-them-off-roads/PZDb97YI6NMNXIwraVkb0I/.
Haworth, N., Schramm, A., & Twisk, D. (2021). Comparing the risky behaviours of shared and private e-scooter and bicycle riders in downtown Brisbane, Australia. *Accident Analysis & Prevention, 152*, 105981. https://doi.org/10.1016/j.aap.2021.105981.
James, O., Swiderski, J., Hicks, J., Teoman, D., & Buehler, R. (2019). Pedestrians and E-Scooters: An Initial Look at E-Scooter Parking and Perceptions by Riders and Non-Riders. *Sustainability, 11*(20), 5591. https://doi.org/10.3390/su11205591.
Jones, T. (2019). 4th person killed while riding a scooter in metro Atlanta. *WBS-TV* https://www.wsbtv.com/news/local/south-fulton-county/man-on-scooter-hit-killed-in-4th-scooter-related-death-in-metro-atlanta/973874796/.
KHQ Q6 News. (2019). Man killed in Lime Scooter hit-and-run crash in Hillyard identified. KHQ Q6 News. NBC. https://www.khq.com/news/man-killed-in-lime-scooter-hit-and-run-crash-in-hillyard-identified/article_67f7a8a4-f03b-11e9-bb2d-b753b34521c0.html.
Kim, J.-K., Ulfarsson, G. F., Shankar, V. N., & Kim, S. (2008). Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accident Analysis & Prevention, 40*(5), 1695–1702. https://doi.org/10.1016/j.aap.2008.06.005.
Kleinertz, H., Ntalos, D., Hennes, F., Nüchtern, V. J., Frosch, K.-H., & Thiesen, M. D. (2021). Accident mechanisms and injury patterns in e-scooter users. *Deutsches Ärzteblatt International, 118*(8), 117–121. https://doi.org/10.3238/arztebl.m2021.0019.
Kobayashi, L. M., Williams, E., Brown, C. V., Emigh, B. J., Bansal, V., Bradiee, J., ... Doucet, J. (2019). The e-merging e-pidemic of e-scooters. *Trauma Surgery & Acute Care Open, 2019*(4), e000337.
Laa, B., & Leth, U. (2020). Survey of E-scooter users in Vienna: Who they are and how they ride. *Journal of Transport Geography, 89*, 102874. https://doi.org/10.1016/j.jtrangeo.2020.102874.
Liew, Y., Wee, C., & Pek, J. (2020). New peril on our roads: A retrospective study of electric scooter-related injuries. *Singapore Medical Journal, 61*(2), 92–95. https://doi.org/10.11622/smedj.2019083.
Lime. (2021). User Agreement. http://v1.li.me/user-agreement.
Lohrmann, D. (2019). Scooter Accidents, Death Raise Concern in Fort Lauderdale, Fla. *Government Technology*. https://www.govtech.com/transportation/Scooter-Accidents-Death-Raise-Concern-in-Fort-Lauderdale-Fla.html.
Marrero, T. (2019). Man hit by semi-trailer truck while riding rented electric scooter has died. *Tampa Bay Times* https://www.tampabay.com/news/publicsafety/man-hit-by-semi-trailer-truck-while-riding-rented-electric-scooter-has-died-20190628/.
Mayhew, L. J., & Bergin, C. (2019). Impact of e-scooter injuries on Emergency Department imaging. *Journal of Medical Imaging and Radiation Oncology, 63*(4), 461–466. https://doi.org/10.1111/1754-9485.12889.
Mitchell, G., Tsao, H., Randell, T., Marks, J., & Mackay, P. (2019). Impact of electric scooters to a tertiary emergency department: 8-week review after implementation of a scooter share scheme. *Emergency Medicine Australasia, 31*(6), 930–934. https://doi.org/10.1111/1742-6723.13356.
Morgan Lewis. (2021). State and Local Survey of Laws Regulating Escooter Sharing Services. Morgan, Lewis, and Blockius LLP. https://www.morganlewis.com/-/media/files/publication/morgan-lewis-title/white-paper/2021/mlb-e-scooter-white-paper.pdf.
Moskowitz, H., & Burns, M. (1990). Effects of alcohol on driving performance. *Alcohol Health & Research World, 14*(1), 12+ https://link.gale.com/apps/doc/A9353143/AONE?u=mlin_oweb&sid=googleScholar&xid=5cf13ea2.
National Association of City and Transportation Officials. (2019). Shared Micromobility in the U.S.: 2019. https://nacto.org/shared-micromobility-2019/.
National Highway Traffic Safety Administration. (1993). Addressing the safety issues related to younger and older drivers [Report to Congress]. https://one.nhtsa.gov/people/injury/olddrive/pub/yorept.html.
National Highway Traffic Safety Administration. (2020a). NHTSA Field Crash Investigation 2019 Coding and Editing Manual. Report No. DOT HS 813 042. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813042.
National Highway Traffic Safety Administration. (2020b). Traffic Safety Facts: Bicyclists and Other Cyclists. Report No. DOT HS 812 884. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812884.
National Highway Traffic Safety Administration. (2021a). Fatality Analysis Reporting System, Analytical User's Manual 1975-2019. Report No. DOT HS 813 023. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813023.
National Highway Traffic Safety Administration. (2021b). Traffic Safety Facts: Pedestrians. Report No. DOT HS 813 079. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813079.
National Safety Council (2018). Policy Statement Vulnerable Road Users. https://www.nsc.org/getattachment/d5babee6-582d-4e66-804f-8d06f9b021a4/t-vulnerable-road-users-147#:~:text=The%20term%20vulnerable%20road%20user,of%20protection%20against%20such%20collisions.
National Science Foundation (2020). Award Abstract # 2028098 Collaborative Research: RAPID: Maintain Mobility and Reduce Infection Through a Resilient Transit and Micromobility System. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2028098&HistoricalAwards=false.
Nellamattathil, M., & Amber, I. (2020). An evaluation of scooter injury and injury patterns following widespread adoption of E-scooters in a major metropolitan area. *Clinical Imaging, 60*(2), 200–203. https://doi.org/10.1016/j.clinimag.2019.12.012.
News 12 Staff. (2019). Friends gather to mourn Elizabeth teen killed while riding E-scooter. *News 12 New Jersey*. https://newjersey.news12.com/officials-16yearold-boy-killed-in-elizabeth-while-riding-escooter-colliding-with-tow-truck-41353986.
Oxley, J., Lenné, M., & Corben, B. (2006). The effect of alcohol impairment on road-crossing behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour, 9*(4), 258–268. https://doi.org/10.1016/j.trf.2006.01.004.
Penney and Associates (2021). Are electric rideshare scooters safe? *PR Newswire* https://www.prnewswire.com/news-releases/are-electric-rideshare-scooters-safe-301348749.html.
Perry, F. (2020). Why we have a love-hate relationship with electric scooters. *BBC Future Planet* https://www.bbc.com/future/article/20200608-how-sustainable-are-electric-scooters.
Portland Bureau of Transportation. (2018). News release: PBOT releases results of e-scooter user survey. https://www.portlandoregon.gov/transportation/article/700917.
Pozen, L. (2019). Woman out on date night with her husband hit, critically injured riding scooter. https://www.wsbtv.com/news/local/atlanta/scooter-rider-hit-killed-by-car-in-midtown-atlanta-police-say/972063999/.
Prince, C. (2019a). Driver arrested in Atlanta's first deadly e-scooter accident. *The Atlanta Journal-Constitution* https://www.ajc.com/news/crime–law/just-driver-arrested-atlanta-first-deadly-scooter-accident/nQGgEEnEAcMYODal4J0iiK/.
Prince, C. (2019b). No charges for CobbLinc bus driver in deadly e-scooter accident. *The Atlanta Journal-Constitution* https://www.ajc.com/news/breaking-news/breaking-charges-for-cobblinc-bus-driver-deadly-scooter-accident/l0HgZx0NqP5W50Ixyo9JSL/.
R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/.
SAE International. (2019). Surface Vehicle Recommended Practice. J3194. Taxonomy and Classification of Powered Micromobility Vehicles. https://www.sae.org/standards/content/j3194_201911/.
Sanders, R. L., Branion-Calles, M., & Nelson, T. A. (2020). To scoot or not to scoot: Findings from a recent survey about the benefits and barriers of using E-scooters for riders and non-riders. *Transportation Research Part A: Policy and Practice, 139*, 217–227. https://doi.org/10.1016/j.tra.2020.07.009.
Santamariña-Rubio, E., Pérez, K., Olabarria, M., & Novoa, A. M. (2014). Gender differences in road traffic injury rate using time travelled as a measure of exposure. *Accident Analysis & Prevention, 65*, 1–7. https://doi.org/10.1016/j.aap.2013.11.015.
Shah, N. R., Aryal, S., Wen, Y., & Cherry, C. R. (2021). Comparison of motor vehicle-involved e-scooter and bicycle crashes using standardized crash typology. *Journal of Safety Research, 77*, 217–228. https://doi.org/10.1016/j.jsr.2021.03.005.
Shaheen, S., & Cohen, A. (2019). Micromobility policy toolkit: Docked and dockless bike and scooter sharing. https://doi.org/10.7922/G2TH8JW7.
Smith, C. S., & Schwieterman, J. P. (2018). E-scooter scenarios: Evaluating the potential mobility benefits of shared dockless scooters in Chicago. https://trid.trb.org/view/1577726.
Solnick, S. J., & Hemenway, D. (1994). Hit the bottle and run: The role of alcohol in hit-and-run pedestrian fatalities. *Journal of Studies on Alcohol, 55*(6), 679–684. https://doi.org/10.15288/jsa.1994.55.679.
Solnick, S. J., & Hemenway, D. (1995). The hit-and-run in fatal pedestrian accidents: Victims, circumstances and drivers. *Accident Analysis & Prevention, 27*(5), 643–649. https://doi.org/10.1016/0001-4575(95)00012-O.
Stunson, M. (2019). Man riding Lime electric scooter dies following crash on Richmond Road in Lexington. *Lexington Herald Leader* https://www.kentucky.com/news/local/counties/fayette-county/article237018434.html.
Sukut, J. (2019). Billings man dies in scooter wreck in Denver. *Billings Gazette* https://billingsgazette.com/news/local/billings-man-dies-in-scooter-wreck-in-denver/article_ab2f63b9-1e96-554c-9060-be80934cec9c.html.

Sumner, J. (2016). *Bicycling complete book of road cycling skills: Your guide to riding faster, stronger, longer, and safer* (Revised and updated 2nd edition). Rodale.

Tay, R., Barua, U., & Kattan, L. (2009). Factors contributing to hit-and-run in fatal crashes. *Accident Analysis & Prevention, 41*(2), 227–233. https://doi.org/10.1016/j.aap.2008.11.002.

Tillemann, L., & Feasley, L. (2018). E-scooters could save the city. *Wired* https://www.wired.com/story/e-scooter-micromobility-infographics-cost-emissions/.

Todd, J., Krauss, D., Zimmermann, J., & Dunning, A. (2019). Behavior of electric scooter operators in naturalistic environments. 2019-01–1007. https://doi.org/10.4271/2019-01-1007.

Trivedi, T. K., Liu, C., Antonio, A. L. M., Wheaton, N., Kreger, V., Yap, A., ... Elmore, J. G. (2019). Injuries associated with standing electric scooter use. *JAMA Network Open, 2*(1), e187381.

Turner, C., & McClure, R. (2003). Age and gender differences in risk-taking behaviour as an explanation for high incidence of motor vehicle crashes as a driver in young males. *Injury Control and Safety Promotion, 10*(3), 123–130. https://doi.org/10.1076/icsp.10.3.123.14560.

Unagi, Inc. (2021). The 2021 Comprehensive Guide to Electric Scooter Laws. https://unagiscooters.com/articles/the-2021-comprehensive-guide-to-electric-scooter-laws/.

United States Consumer Product Safety Commission (2014). Which Helmet for Which Activity? https://www.cpsc.gov/safety-education/safety-guides/sports-fitness-and-recreation-bicycles/which-helmet-which-activity.

University of North Carolina Chapel Hill. (2021). Collaborative Sciences Center for Road Safety. E-Scooter Fatalities. https://www.roadsafety.unc.edu/wp-content/uploads/2021/02/escooter_fatalities_Feb_2021.pdf.

Wallman B., & Maines J. (2019). 74 accidents and 1 death later, Fort Lauderdale weighs new e-scooter law. *South Florida Sun Sentinel.* https://www.sun-sentinel.com/local/broward/fort-lauderdale/fl-ne-electric-scooters-legal-20190624-v2o5hd6omrfaxa3532x2sildve-story.html.

Weisstein, Eric W. (n.d.) Bonferroni Correction. MathWorld. Wolfram Research, Inc. https://mathworld.wolfram.com/BonferroniCorrection.html.

Yakowicz, W. (2018). 14 Months, 120 Cities, $2 Billion: There's Never Been a Company Like Bird. Is the World Ready? *Inc.com.* https://www.inc.com/magazine/201902/will-yakowicz/bird-electric-scooter-travis-vanderzanden-2018-company-of-the-year.html.

Yang, H., Ma, Q., Wang, Z., Cai, Q., Xie, K., & Yang, D. (2020). Safety of micromobility: Analysis of E-Scooter crashes by mining news reports. *Accident Analysis & Prevention, 143*, 105608. https://doi.org/10.1016/j.aap.2020.105608.

Elizabeth Karpinski is a Lead Data Scientist at the MITRE Corporation. Her focus is on mathematical representations of complex systems and probabilistic analysis. Elizabeth has been leading micromobility research projects at MITRE on the safety and infrastructure requirements needed to support multimodal transport. She received a bachelor's degree in Industrial Engineering and a master's in Operations Analytics & Management both from Worcester Polytechnic Institute.

Ellie Bayles is a Senior Data Scientist at the MITRE Corporation where she specializes in mathematical modeling and statistical computing. She holds a bachelor's degree in statistics from Cornell University.

Lisa Daigle is a Lead Health Policy Analyst at The MITRE Corporation. Her primary research interest is in health policy at the federal, state, and local government levels. She has a Master of Science in Health Communication from Boston University and a Master of Arts in Political Science from American University.

Dan Mantine is a Lead Project Management Specialist at the MITRE Corporation. His background includes 40 years of experience in Civil Engineering and transportation, as well as construction and design. In his work at MITRE, he serves sponsors in project and program management roles and performs portfolio analysis. Dan holds a bachelor's in Civil Engineering from Pennsylvania State University, and an MBA from George Washington University.

Journal of Safety Research 84 (2023) 371–383

Contents lists available at ScienceDirect

Journal of Safety Research

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

Consumer demand for partial driving automation and hands-free driving capability

Alexandra S. Mueller [a,*], Jessica B. Cicchino [a], Joseph V. Calvanelli Jr [b]

[a] Insurance Institute for Highway Safety, Arlington VA, United States
[b] Opinion America Group, United States

ARTICLE INFO

Article history:
Received 25 May 2022
Received in revised form 7 September 2022
Accepted 30 November 2022
Available online 9 December 2022

Keywords:
Public opinion
Lane centering
Level 2
Driver monitoring
Auto lane change

ABSTRACT

Introduction: It is often assumed that consumers want partial driving automation in their vehicles, yet there has been little research on the topic. Also unclear is what the public's appetite is for hands-free driving capability, automated (auto)-lane-change functionality, and driver monitoring that helps reinforce proper use of these features. *Method:* Through an internet-based survey of a nationally representative sample of 1,010 U.S. adult drivers, this study explored consumer demand for different aspects of partial driving automation. *Results:* Eighty percent of drivers want to use lane centering, but more prefer versions with a hands-on-wheel requirement (36%) than hands-free (27%). More than half of drivers are comfortable with different driver monitoring strategies, but comfort level is related to perceptions of feeling safer with it given its role in helping drivers use the technology properly. People who prefer hands-free lane centering are the most accepting of other vehicle technologies, including driver monitoring, but some also indicate an intent to misuse these features. The public is somewhat more reluctant to accept auto lane change, with 73% saying they would use it, and more often prefer it to be driver-initiated (45%) than vehicle-initiated (14%). More than three quarters of drivers want auto lane change to have a hands-on-wheel requirement. *Conclusion:* Consumers are interested in partial driving automation, but there is resistance to more sophisticated functionality, especially vehicle-initiated auto lane change, in a vehicle that cannot technically drive itself. *Practical applications:* This study confirms the public's appetite for partial driving automation and possible intention for misuse. It is imperative that the technology be designed in ways that deter such misuse. The data suggest that consumer information, including marketing, has a role to play to communicate the purpose and safety value of driver monitoring and other user-centric design safeguards to promote their implementation, acceptance, and safe adoption.

1. Introduction

Many of the vehicles one can buy today come equipped with driver assistance features. Some driving support is so sophisticated that it can give consumers the impression that the vehicle can drive itself. At present, however, there are no driverless or self-driving vehicles on the consumer market. Partial driving automation, also known as Level 2 systems (SAE International, 2021), is currently the most advanced vehicle technology available for purchase in North America. Most of these systems are designed to operate on highways or limited-access roads, and they provide continuous speed, headway monitoring, and steering support through the combined use of adaptive cruise control and sustained lane centering.

Partial driving automation requires the driver to supervise the road and the vehicle, and drivers must be able to intervene rapidly when these systems get into situations they cannot cope with (SAE International, 2021). On-road testing has shown that these systems can struggle to provide support under fairly benign conditions (American Automobile Association, Inc. [AAA], 2020; Insurance Institute for Highway Safety [IIHS], 2018; Kim, Song, & Doerzaph, 2022); for example, they can have trouble steering the vehicle within the lane when traveling on hills or in curves or detecting stopped vehicles. In comparison with crash avoidance features, such as automatic emergency braking, which has empirical support for mitigating and preventing crashes (e.g., Cicchino, 2017), no clear crash-reduction effectiveness has been established for partial driving automation (Goodall, 2021; Highway Loss Data Institute [HLDI], 2021a, 2021b). Rather, these systems are often

* Corresponding author at: Insurance Institute for Highway Safety, 4121 Wilson Blvd, Suite 600, Arlington, VA 22203.

E-mail address: amueller@iihs.org (A.S. Mueller).

https://doi.org/10.1016/j.jsr.2022.11.012
0022-4375/

marketed as driver convenience features, as they are intended to make driving easier or more comfortable.

Unfortunately, human limitations make the partially automated nature of the technology an issue when it comes to its safe and proper use. People have difficulty supervising a task they are not actively or physically involved in, and this difficulty is exemplified when the driving task is supported by partial driving automation (Banks, Eriksson, O'Donoghue, & Stanton, 2018; Biondi, Lohani, Hopman, Mills, Cooper, & Strayer, 2018; Gaspar & Carney, 2019). In addition, there is growing evidence that this technology increases the opportunity for distracted driving (Banks et al., 2018; Dunn, Dingus, Soccolich, & Horrey, 2021; Kim et al., 2022; Reagan et al., 2021), and driver distraction increases crash risk (Dingus et al., 2016). Most of these systems monitor the driver for behavior that indicates they are out of the loop; for example, steering wheel torque or capacitive touch sensors are typically used to detect when the driver's hands are off the wheel and cameras are used to detect when the driver's eyes or head position are directed away from the road. There has been a recent push for automakers to adopt eye-gaze or head-pose monitoring strategies (AAA Inc., 2022; IIHS, 2022; Preston, 2021) because what the driver is paying attention to tends to be more reliably correlated with where the driver is looking than what the driver's hands are doing (for a review, see El Khatib, Ou, & Karray, 2020), yet driver hand monitoring remains important for detecting manual distraction (Halin, Verly, & Van Droogenbroeck, 2021). Although there are societal, legal, and privacy concerns around driver monitoring (Ghazizadeh & Lee, 2018; Horrey, Lesch, Dainoff, Robertson, & Noy, 2012), little is known about the driving public's attitudes concerning driver monitoring and whether they vary based on the design of the partial automation itself.

While most commercially available systems require drivers to keep their hands on the wheel, some allow drivers to take their hands off the wheel for extended periods. Systems that permit hands-free driving typically utilize eye-gaze or head-pose tracking to help ensure drivers fulfill their supervisory roles, but most do not monitor what the driver's hands are doing when hands-free driving is engaged. A concern is that hands-free driving might increase the likelihood of drivers doing non-driving-related manual activities, such as eating or texting. If the driver's hands are otherwise occupied, their reaction time to take the wheel will be slowed (Wang, Zheng, Kaizuka, & Nakano, 2019), even if they are attentively supervising and can see that steering intervention is required. While it remains to be demonstrated what the appeal of hands-free driving is among the driving public, it is also an open question whether consumer attitudes and expectations for different driver monitoring strategies vary based on the appeal of certain system functionalities, such as hands-free driving.

The increasing functionality of partial automation may also be exacerbating consumer misunderstanding about the driver's responsibilities and the technology's limitations (Banks & Stanton, 2016; Mueller, Reagan, & Cicchino, 2021). Some of these systems have automated (auto)-lane-changing functionalities, where the vehicle will make a lane change on its own without the driver needing to steer. Despite this capability, the driver still has to make sure the maneuvers are safe to perform and to intervene when necessary to prevent crashes. It is difficult to convey the limits of partial driving automation to consumers when it can perform maneuvers that are inherently riskier than just steering within the lane. Some auto-lane-change-assistance features require driver input, for example by use of the turn stalk or button press, as a form of verification that the driver is in the loop before the system will perform the lane-change maneuver. This is known as driver-initiated auto lane change. Other versions of this feature, known as vehicle-initiated auto lane change, do not require any driver input and the vehicle can make the decisions to change lanes on its own. Vehicle-initiated auto lane change makes it difficult for even attentive drivers to anticipate these actions, let alone be able to determine whether it is safe for the maneuver to be performed before it happens. The complexity of the issue is amplified by the fact that some automakers offer auto-lane-change functionality as hands-free (e.g., Toyota's Teammate Advanced drive and General Motors's Super Cruise; Toyota, n.d.; General Motors, n.d.).

1.1. Study objectives

The marketing of partial driving automation frequently assumes that features such as hands-free driving and auto lane change are what consumers want. However, while the appeal of fully self-driving technology has received much attention, few studies have explored what the public appetite is for partial driving automation that exists in production vehicles today (e.g., Daziano, Sarrias, & Leard, 2017; Lee, Gershon, Reimer, Mehler, & Coughlin, 2021). The goal of this study was to provide clarity on the subject and to determine whether preferences for hands-free driving support are ubiquitous or whether there are differences between lane centering and auto lane change and, if so, why. Another aim of the investigation was to determine whether consumers accept and see the value of driver monitoring technologies that are commonly equipped with partial driving automation to ensure its proper use. Furthermore, do those attitudes vary between features that offer hands-free driving capability and those that require drivers to keep their hands on the wheel? We conducted a nationally representative survey of U.S. drivers to answer these questions.

2. Method

2.1. Procedure

The survey was conducted online from September to October in 2021 and was hosted on the Voxco platform. Individuals were recruited to participate by email invitation from the Lucid Online Marketplace, which is a community composed of hundreds of suppliers with a diverse set of recruitment and sourcing methodologies. Respondents were informed that the survey was about understanding public opinion concerning commercially available driver assistance technologies and that it would take approximately 15 min to complete. They provided written informed consent to participate before gaining access to the survey instrument. Each respondent who qualified and completed the survey was paid $5 by the marketplace supplier from whom they were recruited within the Lucid Online Marketplace. The study protocol was deemed exempt by Advarra, an independent IRB company.

2.2. Sample

Quotas were used to match the sample to the age and gender distributions of the U.S. population aged 21 and older using 2010 Census data as the target basis because 2020 estimates were not yet available at the time of the survey; the 2010 age and gender distributions were similar to the 2020 middle series estimates of the U.S. resident population (United States Census Bureau, 2020). As most respondents were expected to be unfamiliar with the topics discussed in the survey, a pilot testing phase was necessary to refine survey language; however, the data collected from the pilot phase ($n = 101$) were not used for the study analysis. The final sample consisted of 1,010 U.S. residents aged 21 years and older who typically drove at least 1 day per week on highways. Six hundred and eighty other individuals agreed to participate but were determined to be ineligible—nine of whom were younger than 21 years, 294 did not drive at all, 85 drove less than 1 day per week, and 292 drove less than 1 day per week on highways. An additional 34 respondents were removed due to quality control issues (based on response consistency, response quality, and completion time criteria) and 775 other individuals started but did not complete the survey.

2.3. Survey instrument

Definitions about a technology's purpose, functional capabilities, and limitations were provided to respondents before they were asked about their opinions and preferences for that given technology. Lane-centering assistance was also described in the context of partial driving automation, in that it could be activated simultaneously with adaptive cruise control. To prevent respondents confusing lane centering with lane departure prevention, both features were defined independently to distinguish the temporary nature of lane departure prevention from the continuous steering support of lane centering. Hands-free and hands-on-wheel driving requirements were defined for lane centering and auto lane change. Driver-initiated and vehicle-initiated versions of auto lane change were characterized separately. The survey included sections on lane centering, driver monitoring, lane-change assistance, self-driving car appeal, and demographics and driving habits.

2.3.1. Lane centering

Respondents were asked if they would want to use lane-centering assistance, and if so, what type they would prefer (hands-free, hands-on-wheel, no preference). Five-point Likert scales assessed whether hands-free assistance would make driving more or less stressful, safe, and boring, and make the driver more or less comfortable, tired, distracted, and likely to do non-driving-related activities, compared with hands-on-wheel lane centering. Participants were additionally asked with 5-point Likert scales about their likelihood of buying or leasing their next vehicle with hands-free or hands-on-wheel lane centering, if cost was not an issue.

2.3.2. Driver monitoring

Using 5-point Likert scales, respondents were asked how comfortable they would be with different driver monitoring strategies for hands-on-wheel and hands-free versions of lane centering, as well as how safe they would feel knowing that the vehicle would be monitoring them to help ensure the technology was being used as designed.

2.3.3. Auto lane change

Respondents were asked about the degree of confidence they had in their ability to make manual lane changes on the highway and the degree of stress they tend to experience when making those maneuvers with 5-point Likert scales. They were then asked about vehicle technologies that can assist with lane changing. Blind spot detection alerts the driver if there is another vehicle in their blind spot when they want to change lanes. Respondents were asked whether they want to use blind spot detection, driver-initiated auto lane change, and vehicle-initiated auto lane change. Every respondent was then asked to specify whether they preferred to use hands-free or hands-on-wheel versions of driver- and vehicle-initiated auto lane change and why. Using 5-point Likert scales, respondents were asked about their willingness to purchase or lease their next vehicle with hands-free and hands-on-wheel versions of driver-initiated and vehicle-initiated auto lane change.

2.3.4. Self-driving technology appeal

Respondents were asked "How appealing would it be for you to own or regularly use a self-driving vehicle in the future? Self-driving means that the vehicle itself would control all the safety–critical functions, even allowing the vehicle to travel without a passenger if required. In other words, the vehicle would be able to drive itself anytime, anywhere, and under any conditions. You would be able to get into the vehicle, instruct it where you would like to travel to, and the vehicle would then carry out your desired route with no further intervention required from you. There might not even be a steering wheel or speed controls in the vehicle." Degree of appeal was captured through a 5-point Likert scale.

2.4. Data analysis

To simplify interpretation, Likert-scale data were grouped into broader categories. Differences in survey responses by preference for lane centering (referred to as lane-centering preference group), preference for auto lane change (auto-lane-change preference group), and self-driving car appeal were examined using chi-square tests. A critical p value of 0.05 was used to determine statistical significance and actual p values were reported for statistically significant results. Response categories were collapsed for select comparisons reported in the *Results* section, and the data for all response categories for those survey items can be found in Appendix A.

3. Results

3.1. Sample

Table 1 shows the distribution of sample demographics for age, gender, education, income, U.S. Census region, and weekly high-

Table 1
Sample demographics.

	Percent (N = 1,010)
Age (years)	
21 to 34	26
35 to 64	51
65 and older	23
Gender	
Male	49
Female	51
Other	< 1
Education level	
High school diploma or less	23
Some college education, associate degree, or trade school	34
Bachelor's degree	24
Some graduate education	4
Graduate or professional degree	15
Income	
Less than $50,000	45
$50,000 to $74,999	20
$75,000 to $99,999	13
$100,000 to $149,999	13
$150,000 to $199,999	6
$200,000 or more	3
Region	
Northeast	17
Midwest	21
South	39
West	24
Highway driving exposure (average days per week)	
5 days or more per week	58
3 to 4 days per week	24
1 to 2 days per week	19

Note. Percentages may not sum to 100 due to rounding.

way driving exposure. The average age of respondents was 46 years (*SD* = 17, min = 21, max = 91). The majority drove on the highway 5 or more days per week on average.

3.2. Lane centering

Eighty percent of the sample wanted to use at least some form of lane centering. More respondents preferred hands-on-wheel (36 %) than hands-free lane centering (27 %), and 18 % had no preference between the two types. Sixteen percent did not want to use any form of lane centering and 4 % were unsure.

3.2.1. Expectations for hands-free vs hands-on-wheel lane centering

As shown in Fig. 1, compared with when using hands-on-wheel lane centering, most (62 %) of the sample said that using the hands-free version would make driving more stressful, would make them more likely to do non-driving-related activities (e.g., eat, drink, text, groom, converse with a passenger) (61 %), and more distracted (56 %) and comfortable (50 %). Nearly half said it would make driving safer (47 %) or more boring (46 %), and 41 % said it would make them more tired.

Differences were noted among the lane-centering preference groups. A larger percentage of respondents who preferred hands-free or hands-on-wheel versions of lane centering, as well as those who did not want to use the feature, said hands-free lane centering would make driving more stressful, X^2 (4, 1010) = 69.16, p <.0001, or would make them more distracted, X^2 (4, 1010) = 15.70, p = 0.003, compared with those who had no preference or were unsure. Those who preferred hands-free lane centering also most often reported that the feature would make them more comfortable, X^2 (4, 1010) = 136.51, p <.0001, and make driving safer, X^2 (4, 1010) = 149.71, p <.0001; these opinions were reported least often among those who did not want to use any version of lane centering. Additionally, drivers who preferred hands-free lane centering most often reported it would make driving more boring, X^2 (4, 1010) = 36.47, p <.0001, and would make them more tired, X^2 (4, 1010) = 49.81, p <.0001.

3.2.2. Conceptual appeal and likelihood of purchasing lane centering

Assuming price was not an issue, a larger proportion of the sample was willing (defined by being at least moderately likely) to buy or lease their next vehicle with hands-on-wheel lane centering than a hands-free version, as shown in Fig. 2. Willingness for hands-free (X^2 (4, 1010) = 229.81, p <.0001) and hands-on-wheel lane centering (X^2 (4, 1010) = 167.42, p <.0001) varied as a function of preference group. Over three quarters of respondents who preferred hands-free lane centering were willing to buy either a hands-free or hands-on-wheel lane-centering feature. More than two thirds of those who preferred hands-on-wheel lane centering or who had no preference were willing to buy a hands-on-wheel version, but less than half of them were willing to buy a hands-free version. Less than half of respondents who were unsure or who did not want any lane centering were willing to buy either type of lane-centering assistance, but willingness to purchase or lease a vehicle with a hands-on-wheel version was higher among both groups than for a hands-free one.

3.3. Driver monitoring

Attitudes concerning driver monitoring strategies. In the context of using a hands-on-wheel lane-centering feature, the majority of the sample was at least somewhat comfortable with all three driver monitoring strategies: 70 % with steering wheel sensors to monitor driver hands, 59 % with camera monitoring of driver hands, and 57 % with camera monitoring of driver gaze. Similar percentages of respondents were at least somewhat comfortable with camera monitoring of driver hands (58 %) and driver gaze (57 %) in the context of hands-free lane centering.

Comfort with driver monitoring varied by lane-centering preference group for steering wheel sensor monitoring of driver's hands (X^2 (4, 1010) = 242.38, p <.0001), camera monitoring of driver's hands (X^2 (4, 1010) = 207.62, p <.0001), and camera monitoring of driver's gaze (X^2 (4, 1010) = 183.88, p <.0001) when using hands-on-wheel lane centering as well as camera monitoring of driver's hands (X^2 (4, 1010) = 177.32, p <.0001) and camera monitoring of driver's gaze (X^2 (4, 1010) = 193.11, p <.0001) when using

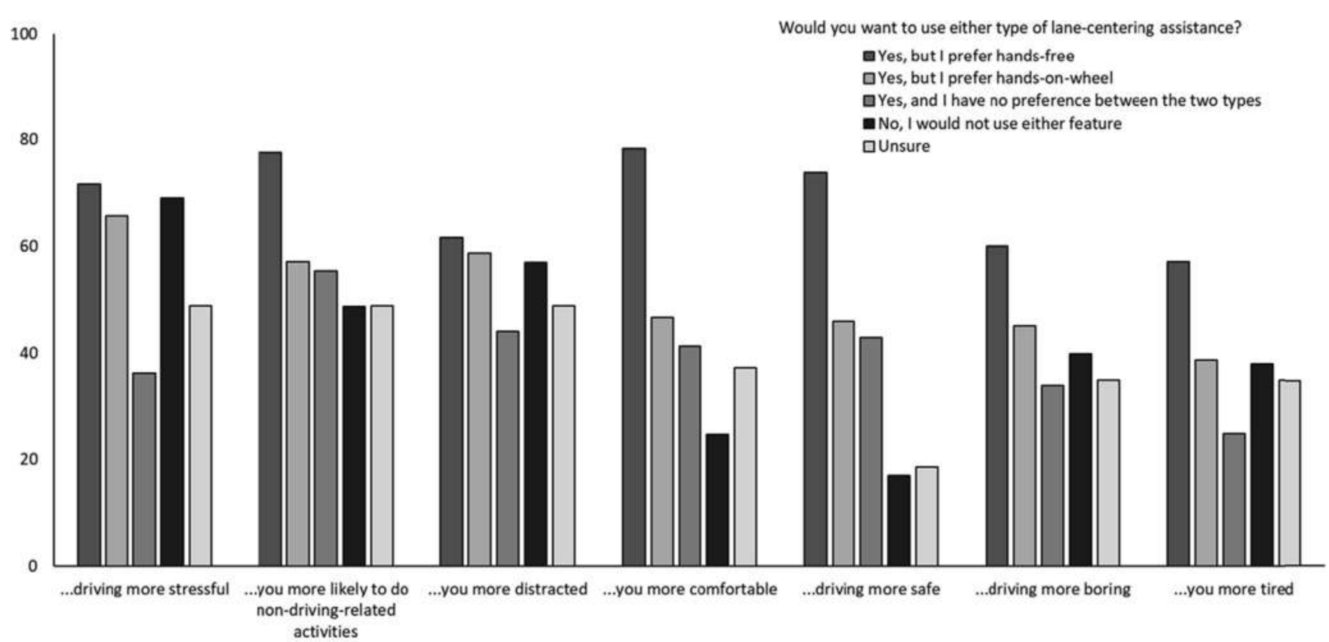


Fig. 1. Percent of respondents by lane-centering preference group who agree that various outcomes would be greater when using hands-free than hands-on-wheel lane centering.

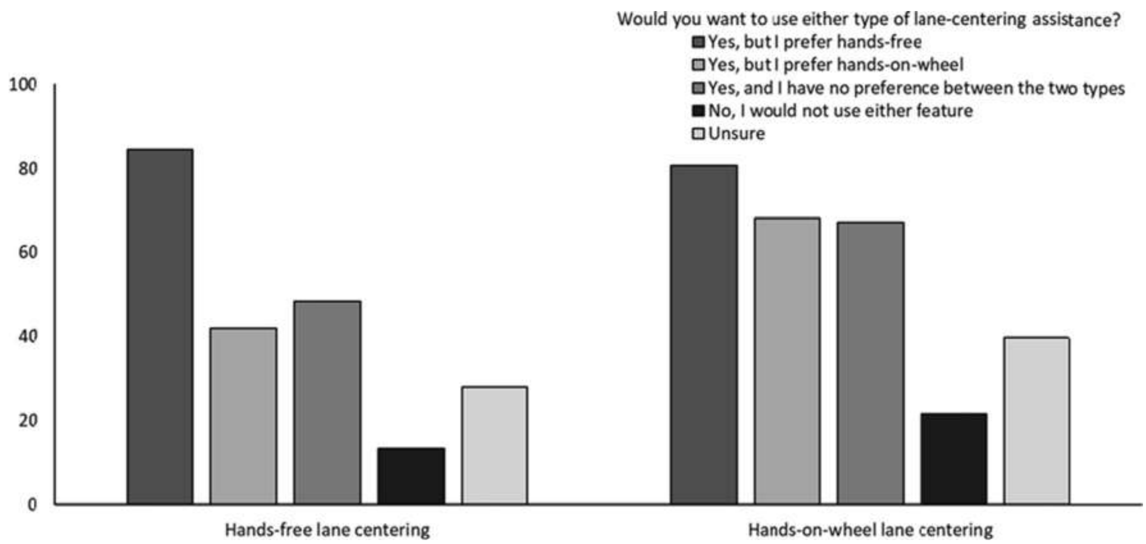


Fig. 2. Percent of respondents at least moderately likely to buy or lease a vehicle with hands-free or hands-on-wheel lane centering per lane-centering preference group.

hands-free lane centering. The patterns observed for the overall sample were primarily driven by respondents who wanted some form of lane centering, although more respondents who preferred hands-free lane centering were extremely comfortable with all forms of driver monitoring for both types of lane centering compared with the other preference groups (Fig. 3). Conversely, almost half of the respondents who did not want to use any lane-centering assistance reported being at least somewhat uncomfortable with driver monitoring when using either type of lane centering, especially camera-based monitoring.

3.3.1. Perceived safety of driver monitoring to ensure proper system use

While most respondents said that they would feel safer knowing that the vehicle was monitoring them to make sure that the feature was being used as it was designed to be used (Table 2), there were differences by lane-centering preference group, X^2 (4, 1010) = 185.99, p <.0001. Among those who did not want lane centering, only 27 % reported they would feel safer with driver monitoring. In contrast, the majority of respondents who wanted hands-free lane centering said they would feel much safer with driver monitoring.

Differences were also seen with how safe driver monitoring would make people feel based on how comfortable they were with the different driver monitoring strategies (Table 3). People who were at least somewhat comfortable with different driver monitoring strategies reported feeling safer than those who were neutral or uncomfortable with driver monitoring. The pattern of the interaction between feeling comfortable and safe with driver monitoring was similar among the different strategies in the context of

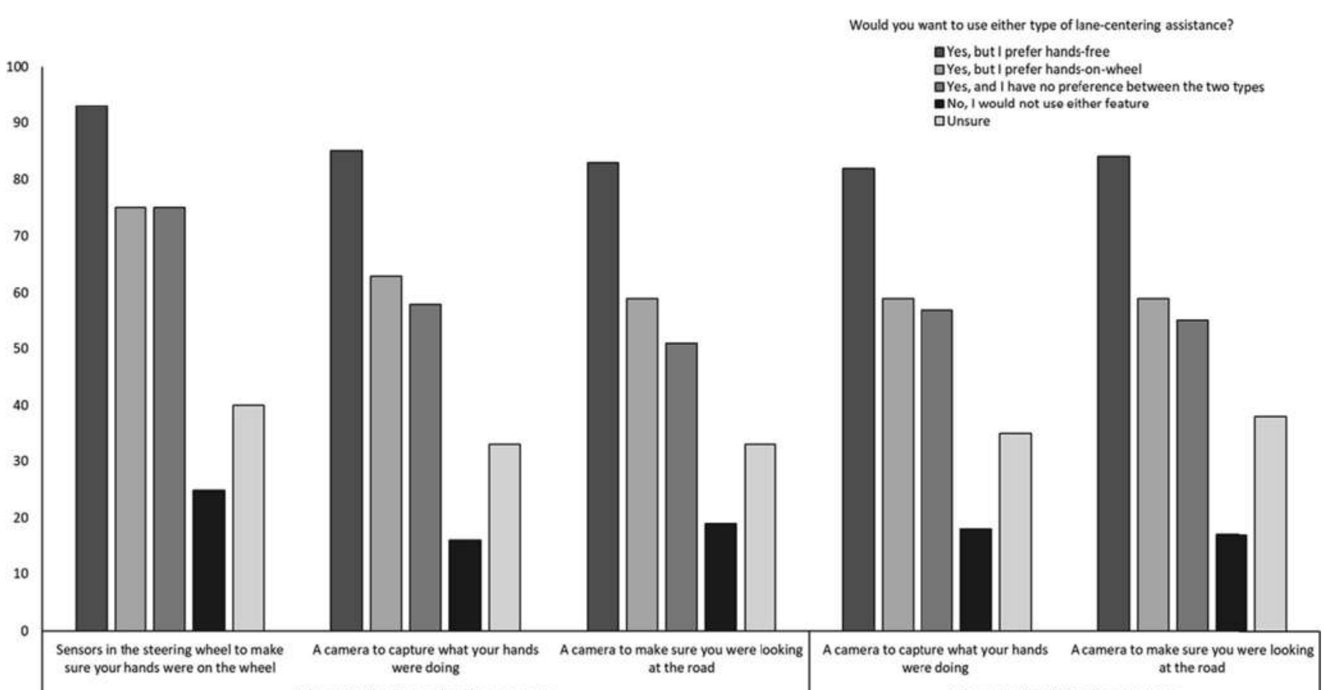


Fig. 3. Percent of respondents per lane-centering preference group reporting to be comfortable with driver monitoring strategies for hands-free and hands-on-wheel lane centering.

Table 2
Perceived safety of driver monitoring as a function of lane-centering preference (in percent).

Would you feel safer knowing that the vehicle is monitoring you to make sure you are using the feature as it was designed to be used?	Would you want to use either type of lane-centering assistance?					
	Yes, but I prefer hands-free	Yes, but I prefer hands-on-wheel	Yes, and I have no preference between the two types	No, I would not use either feature	Unsure	Total
	(n = 268)	(n = 364)	(n = 177)	(n = 158)	(n = 43)	(n = 1,010)
Much safer	60	32	30	9	16	35
Somewhat safer	28	40	36	18	35	33
Neither more nor less safe	9	22	28	47	42	25
Somewhat less safe	1	3	3	11	2	4
Much less safe	< 1	2	2	16	5	4

Note. Percentages may not sum to 100 due to rounding.

Table 3
Perceived safety of driver monitoring as a function of being at least somewhat comfortable with different driver monitoring strategies (in percent).

Would you feel safer knowing that the vehicle is monitoring you to make sure you are using the feature as it was designed to be used?	In the case of hands-on-wheel lane-centering assistance, how comfortable would you be if the vehicle monitored you through…						In the case of hands-free lane-centering assistance, how comfortable would you be if the vehicle monitored you through…			
	sensors in the steering wheel to make sure your hands were on the wheel		a camera to capture what your hands were doing		a camera to make sure you were looking at the road		a camera to capture what your hands were doing		a camera to make sure you were looking at the road	
	At least somewhat comfortable	Neutral or not comfortable	At least somewhat comfortable	Neutral or not comfortable	At least somewhat comfortable	Neutral or not comfortable	At least somewhat comfortable	Neutral or not comfortable	At least somewhat comfortable	Neutral or not comfortable
	(n = 711)	(n = 299)	(n = 597)	(n = 413)	(n = 575)	(n = 435)	(n = 583)	(n = 427)	(n = 577)	(n = 433)
Much safer	47	6	53	9	56	7	55	8	55	9
Somewhat safer	38	21	37	27	34	30	35	29	35	29
Neither more nor less safe	12	54	8	48	8	46	8	47	9	46
Somewhat less safe	2	8	2	7	1	8	1	7	1	7
Much less safe	1	11	<1	9	1	9	<1	9	<1	9

Note. Percentages may not sum to 100 due to rounding.

using hands-on-wheel and hands-free versions of lane centering. Specifically, the interaction persisted with respect to steering wheel sensors to monitor the driver's hands (X^2 (4) = 345.12, p <.0001) and cameras to monitor the driver's hands (X^2 (4) = 371.12, p <.0001) and face (X^2 (4) = 393.19, p <.0001) in hands-on-wheel lane-centering systems as well as cameras to monitor hands (X^2 (4) = 380.41, p <.0001) and face (X^2 (4) = 359.81, p <.0001) in hands-free lane-centering systems.

3.4. Lane-changing assistance

3.4.1. Manual lane-changing ability and lane-change assistance

Most of the sample (67 %) indicated that they found changing lanes on highways at least somewhat stressful, yet 97 % also said they were at least somewhat confident in their ability to perform those maneuvers. Even so, most (88 %) reported that they would like the vehicle to warn them if there is another vehicle in their blind spot when they want to change lanes (i.e., a blind spot detection feature), whereas only 5 % were unsure if they would like and 8 % said they would not like such a feature. Fewer respondents had a desire for auto-lane-change assistance. When asked whether they would want to use either driver-initiated or vehicle-initiated auto lane change, 73 % said they would use some form of auto lane change, with 45 % indicating they would prefer to use driver-initiated auto lane change, and far fewer (14 %) preferring vehicle-initiated auto lane change. Few had no preference (13 %) or were unsure (5 %) if they would use either type of assistance, and 23 % said they would not use either type.

3.4.2. Attitudes toward hands-free and hands-on-wheel auto lane change

Respondents were asked whether they would prefer hands-free or hands-on-wheel requirements separately for driver-initiated and vehicle-initiated versions of auto lane change. Approximately three quarters preferred auto lane change to require the driver to keep their hands on the wheel, regardless of whether the feature was driver-initiated (77 %) or vehicle-initiated (75 %). Only 14 % preferred the hands-free version for driver-initiated and vehicle-initiated auto lane change, and the remainder said they were unsure.

Those who preferred hands-free or hands-on-wheel driver- and vehicle-initiated auto lane change were asked to select reasons for their preference from a list of options, with multiple responses allowed (Table 4). The most common reasons were that they thought it would improve their driving comfort and make driving safer and less stressful. However, more respondents who preferred hands-on-wheel versions gave the reason that it would make driving safer than those who preferred hands-free versions. While respondents who preferred hands-free versions more often indicated that the reason for their preference was to have more opportunity to do non-driving-related activities, this reason was not selected by most of the sample regardless of hands-free or hands-on-wheel preference.

3.4.3. Conceptual appeal and likelihood of purchasing auto lane change

Over half of the sample reported they would be at least moderately likely to get a form of auto lane change on their next vehicle if

Table 4
Reasons why respondents preferred driver-initiated or vehicle-initiated auto lane change to allow hands off the wheel or require drivers to keep their hands on the wheel (in percent).

Reason	Driver-initiated auto lane change		Vehicle-initiated auto lane change	
	Prefer it to allow you to have your hands off the wheel	Prefer it to require you to keep your hands on the wheel	Prefer it to allow you to have your hands off the wheel	Prefer it to require you to keep your hands on the wheel
	(n = 137)	(n = 774)	(n = 138)	(n = 755)
Would make me more comfortable	58	53	46	51
Would make driving less stressful	41	43	40	44
Would make driving safer	30	46	33	45
Would make me less tired	29	20	22	23
Would make me less distracted	16	23	11	22
Would make driving less boring	16	16	20	15
Would give me more opportunity for non-driving-related activities, such as eating, texting, conversing with a passenger, etc.	15	4	14	5

Note. Multiple responses were allowed, and preferences were asked separately for driving-initiated and vehicle-initiated auto lane change. Participants who were unsure about hands-free vs hands-on-wheel requirement preferences for driver-initiated auto lane change (n = 99, 10 % of the sample) and vehicle-initiated auto lane change (n = 117, 12 % of the sample) were not asked about why they preferred hands-free or hands-on-wheel requirements.

price were not an issue, and they were more willing to purchase or lease hands-on-wheel than hands-free versions (Fig. 4). The proportion of respondents that were at least moderately willing to buy or lease their vehicles with various implementations of auto lane change differed based on their preferences to use driver- or vehicle-initiated versions (to buy hands-free driver-initiated: X^2 (4, 1010) = 231.93, p <.0001; to buy hands-on-wheel driver-initiated: X^2 (4, 1010) = 212.59, p <.0001; to buy hands-free vehicle-initiated: X^2 (4, 1010) = 219.90, p <.0001; to buy hands-on-wheel vehicle-initiated: X^2 (4, 1010) = 237.91, p <.0001). Respondents who indicated any preference for driver-initiated or vehicle-initiated auto lane change indicated a greater willingness for their next vehicles to be equipped with these systems compared to respondents who were unsure about these features and/or who do not want to use auto lane change at all. Unlike all the other preference groups, however, those who wanted vehicle-initiated auto lane change were similarly willing to have either hands-on or hands-free driver-initiated and vehicle-initiated versions in their next vehicle.

3.5. Hands-free driving preferences for different driver support features

Differences were observed in hands-free and hands-on-wheel preferences for driver-initiated (X^2 (8, 1010) = 106.43, p <.0001) and vehicle-initiated auto lane change (X^2 (8, 1010) = 110.33, p <.0001) as a function of lane-centering preferences. As shown in Table 5, over 80 % of respondents who wanted to use either hands-free or hands-on-wheel versions of lane centering preferred hands-on-wheel versions of driver- and vehicle-initiated auto lane change compared with hands-free versions of those features. Smaller majorities of the other lane-centering preference groups indicated a preference for hands-on-wheel versions of auto lane change.

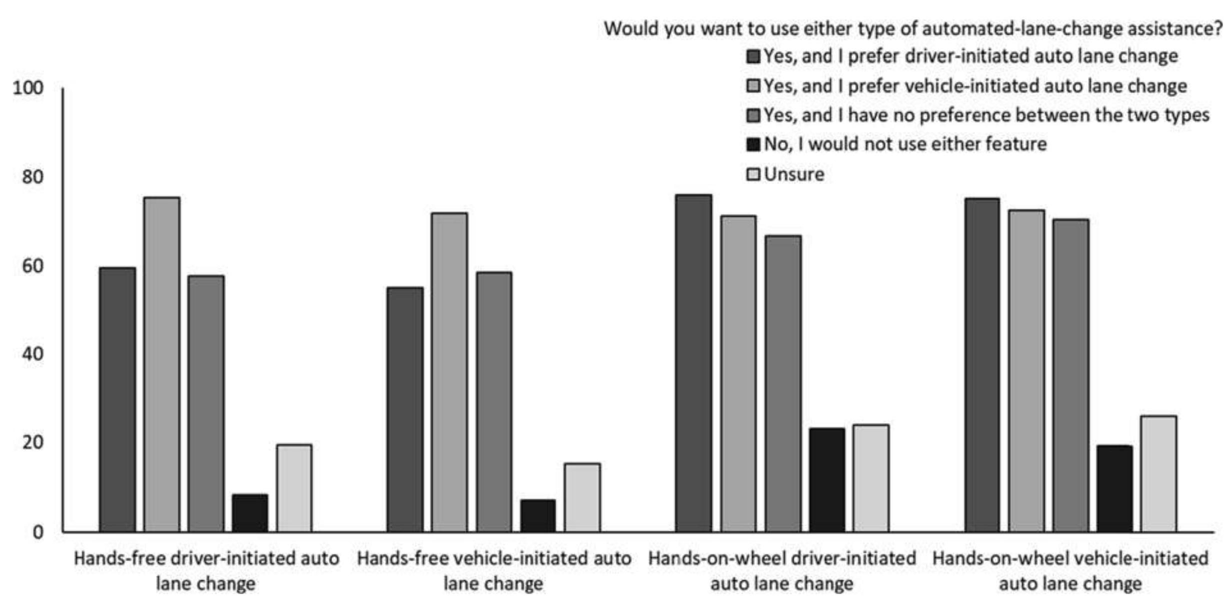


Fig. 4. Percent of respondents at least moderately likely to buy or lease a vehicle with hands-free or hands-on-wheel versions of driver-initiated and vehicle-initiated auto lane change, per auto-lane-change preference group.

Table 5
Hands-free vs hands-on-wheel driver-initiated and vehicle-initiated auto-lane-change preferences based on respondent preferences for lane centering (in percent).

Would you prefer a [driver-initiated or vehicle-initiated, separately asked] automated-lane-change assistance feature to require that you have your hands on the wheel the whole time or would you prefer it to be hands-free?		Would you want to use either type of lane-centering assistance?				
		Yes, but I prefer hands-free	Yes, but I prefer hands-on-wheel	Yes, and I have no preference between the two types	No, I would not use either feature	Unsure
		(n = 268)	(n = 364)	(n = 177)	(n = 158)	(n = 43)
Driver-initiated auto lane change	Prefer it to allow you to have your hands off the wheel	16	11	21	10	5
	Prefer it to require you to keep your hands on the wheel	82	84	64	68	60
	Unsure	2	5	15	22	35
Vehicle-initiated auto lane change	Prefer it to allow you to have your hands off the wheel	16	12	20	8	7
	Prefer it to require you to keep your hands on the wheel	81	82	66	63	58
	Unsure	3	7	14	29	35

Note. Percentages may not sum to 100 due to rounding.

3.6. Self-driving technology appeal

Thirty-five percent of the sample said they found self-driving technology extremely appealing, 19 % said moderately appealing, 14 % said somewhat appealing, 8 % said slightly appealing, and 23 % said not at all appealing. Respondents who indicated they found self-driving technology appealing also indicated a desire for other vehicle technologies (Fig. 5).

Lane-centering preferences (X^2 (16, 1010) = 412.23, p <.0001), feeling more or less safe with driver monitoring (X^2 (16, 1010) = 320.31, p <.0001), and the appeal of auto lane change (X^2 (16, 1010) = 428.40, p <.0001) varied with self-driving technology appeal. Respondents who found self-driving technology to be extremely appealing were most likely to prefer hands-free lane centering and report feeling much safer with driver monitoring. These respondents also most often wanted to use auto lane change, although like respondents in other groups, they preferred it to be driver-initiated rather than vehicle-initiated. Those said that self-driving technology was not at all appealing were most likely to not want to use any form of lane centering or auto lane change, and most frequently reported that driver monitoring would make them feel neither more nor less safe.

4. Discussion

While this study confirms that there is an appetite for partial driving automation, it also shows that the appeal of specific features varies. Lane centering was revealed to be more appealing than auto lane change, and vehicle-initiated auto lane change was overwhelmingly the least appealing of all. Moreover, hands-on-wheel requirements were widely preferred over hands-free driving capability for all these features, especially for both types of auto lane change. Driver monitoring was generally seen as acceptable, although there was less enthusiasm for camera-based monitoring than for steering wheel sensor monitoring, which is to be expected given consumer concerns about privacy and user autonomy (Ghazizadeh & Lee, 2018; Horrey et al., 2012). Even so, attitudes and expectations around driver monitoring corresponded with the perception of being safer with it, given its purpose to help prevent drivers from misusing the technology. This finding suggests that conveying the safety value of driver monitoring may be a key component for consumer education and acceptance (Abraham, Reimer, & Mehler, 2018; Koppel, Charlton, Fildes, & Fitzharris, 2008; Trimble, Baker, Russell, & Tidwell, 2020). Education around driver monitoring and other safeguards may promote

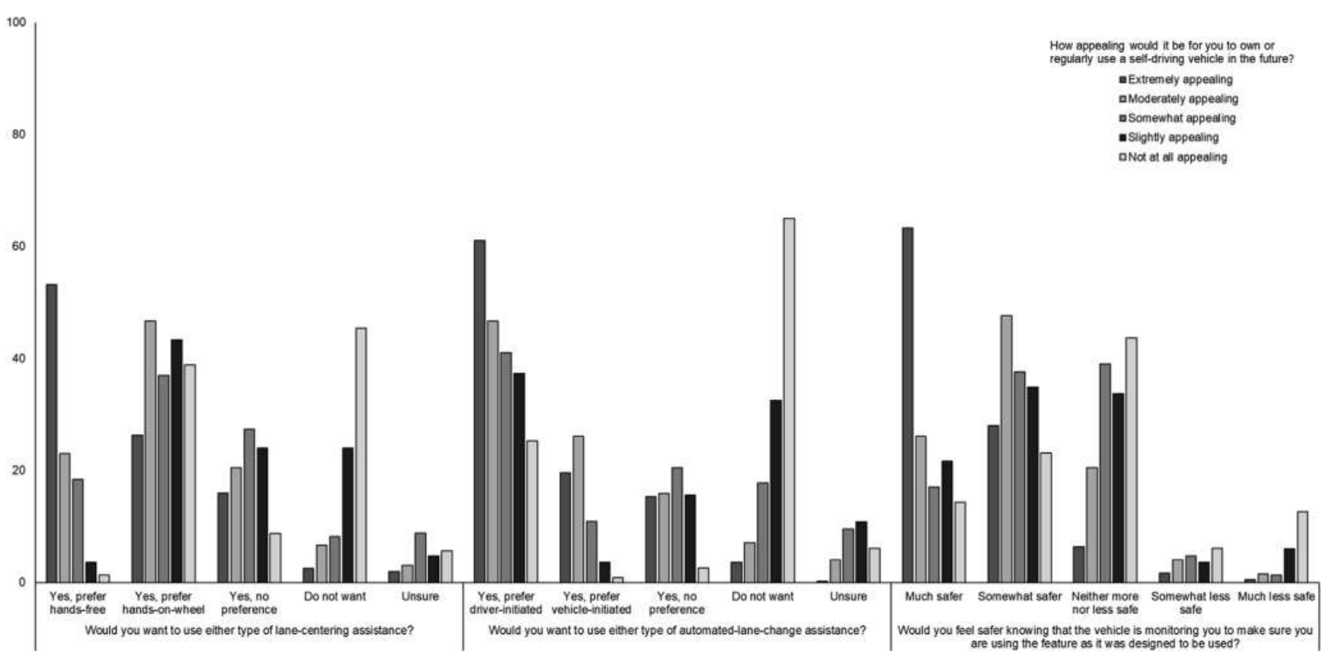


Fig. 5. Percent of respondents per self-driving technology appeal rating with preferences for partial driving automation and safety ratings for driver monitoring.

adoption of in-vehicle technologies that help to prevent driver distraction and inattention more generally, outside the context of partial driving automation.

Individual differences were also observed in how appealing the features were. Respondents who found lane centering appealing tended to find other vehicle technologies also appealing, such as auto lane change and self-driving technology, which is consistent with previous research (e.g., Abraham, Lee, Brady, Fitzgerald, Mehler, Reimer, & Coughlin, 2017). Those people were likewise more accepting of driver monitoring strategies and were more likely to say that they would feel safer with it. Interestingly, respondents who preferred hands-free lane centering were overwhelmingly the most accepting of all driver monitoring strategies, and they were also most willing to buy or lease a vehicle with either hands-on or hands-free versions of lane centering. The rest of the sample was far more willing to buy or lease their next vehicle with hands-on lane centering than a hands-free version. Although respondents were less enthusiastic to have their next vehicle equipped with auto lane change, there was a general preference for it to be driver-initiated and require the driver to keep their hands on the wheel.

One of the aims of this study was to investigate the reasons behind preferences for hands-free or hands-on-wheel versions of driving automation. Compared with hands-on-wheel lane centering, respondents who wanted a hands-free version said it would make driving more comfortable and safer than other drivers. In an interesting display of cognitive dissonance, most of those respondents also said hands-free lane centering would make driving more stressful, boring, and tiring as well as increase distraction. Driver drowsiness and distraction are known crash risk factors (Dingus et al., 2016), and many other respondents shared these concerns too. When asked to choose between hands-free or hands-on-wheel versions of driver-initiated and vehicle-initiated auto lane change, the most common reasons for selecting either were to make driving more comfortable, less stressful, and safer—although more respondents identified "to make driving safer" as a reason for wanting a hands-on-wheel feature than a hands-free one. Evidently, more sophisticated functionality can lead some people to assume that these features have safety benefits, even though the survey never described them as such, and no data currently support those assumptions for partial driving automation in general (e.g., Goodall, 2021; HLDI, 2021a, 2021b). It is nevertheless curious that the vast majority of those respondents who preferred hands-free lane centering (as well as the rest of the sample) preferred hands-on-wheel versions of both types of auto lane change, which indicates some degree of understanding about the risk, driver responsibility, and maneuver complexity of the different functionalities.

Related to the common expectation that hands-free lane centering would increase driver distraction, most of the sample also acknowledged that it would make the driver more likely to do non-driving-related activities than hands-on-wheel lane centering. This attitude was particularly reflected among respondents who wanted hands-free lane centering. Furthermore, a subset of respondents who preferred hands-free auto lane change said they wanted that functionality to be hands-free because they wanted to do non-driving-related activities. These findings are relevant because, when these features were first described in the survey, respondents were informed that vehicles with these features are not autonomous and require the driver to constantly supervise and intervene whenever necessary. This would suggest that informing consumers about the driver's responsibilities and the system's limitations does not necessarily prevent the intention of misuse, which raises concerns for consumer education efforts. That said, many of these respondents were also the most accepting of all types of driver monitoring when using any form of lane centering, which suggests that they may also be more willing to use systems that have safeguards built in, such as attention reminders and emergency escalation countermeasures (IIHS, 2022; Mueller et al., 2022). However, automakers must ensure that these safeguards are reliable and robust to inform and reinforce mental models about the driver's responsibilities and system limits (Cummings & Bauchwitz, 2022). Should systems not have adequate safeguards, there is the potential for hands-free driving to exacerbate the risk of driver disengagement, which has already been contributing to crashes involving the misuse of hands-on-wheel partial driving automation (e.g., National Transportation Safety Board, 2017, 2019, 2020).

Design philosophies that promote cooperative steering between the driver and the lane-centering support play an important role in keeping the driver engaged in the driving task. Information feedback between a driver's hands and gaze behavior helps to coordinate anticipatory steering control (Navarro, Hernout, Osiurak, & Reynaud, 2020). A shared control design philosophy is moreover beneficial for conveying the driver's autonomy in the driver-vehicle interaction (Wen, Kuroki, & Asama, 2019). Designs that reinforce the driver's role and responsibilities in the interaction helps to prevent misperceptions around the use of the system as being "driver versus machine," and instead encourage the driver to work *with* the machine. More research is nevertheless needed to understand how hands-free driving capability affects a driver's mental model about the system limitations and their role and responsibilities when using it (Carsten & Martens, 2019).

4.1. Limitations

Unlike a lot of automaker advertising, the current survey emphasized the limitations of partial automation features and the driver's role when using them. The naming of these systems alone has been shown to mislead consumers about system capability (Teoh, 2020). It is possible that the skepticism exhibited by some of the sample could have been informed by the pragmatic descriptions used to introduce the features of interest.

This study captured the conceptual appeal of the technology, but it cannot predict purchasing behavior or actual use of these features. The intention to use a technology does not guarantee actual use (Turner, Kitchenham, Brereton, Charters, & Budgen, 2010), and the appeal of a technology can change once the individual has had experience using it (Kidd & Reagan, 2018). Moreover, owning or having regular access to vehicles equipped with partial driving automation can change driving habits and technology use (Hardman, 2021). While surveys such as this are useful for investigating the driving public's attitudes, expectations, and intentions concerning the latest advanced vehicle technologies, purchasing behavior itself is necessary to capture the uptake rate, and post-purchase/use research remains key to understanding consumer usage patterns (Melnicuk, Birrell, Thompson, Mouzakitis, & Jennings, 2019).

5. Conclusions

Prior to this study, little was known about the public's appetite for commercially available partial driving automation in the United States, especially with hands-free driving and automated-lane-changing capability. Results indicate that while some consumers find hands-free driving appealing, most prefer driver support features that require the drivers to keep their hands on the wheel. Although people are generally comfortable with driver monitoring, their acceptance level seems to be related to their belief that it improves safety by ensuring proper system use. There is an indication that consumers who prefer hands-free lane centering may be more likely to do non-driving-related activities when using it, yet those people are also the most comfortable with all types of driver monitoring strategies and their purpose. Overall, drivers appear to understand the differences between lane centering and auto lane change, and the latter is less appealing, especially in the form of vehicle-initiated auto lane change.

6. Practical applications

While the intention to misuse a technology does not necessarily mean it will occur, behavioral observation research indicates that driver disengagement can increase over time while using partial driving automation (Banks et al., 2018; Dunn et al., 2021; Kim et al., 2022; Reagan et al., 2021). The results from this study suggest that informing consumers about the limitations of partial driving automation does not necessarily deter the intention to misuse it. Design safeguards with responsible application of automated functionality are necessary to impose functional 'guardrails' that minimize opportunity for misuse. One of the mechanisms supporting these safeguards is driver monitoring. This study's data show that the public's acceptance of it appears to be connected to the understanding of its purpose in helping drivers use the technology properly. Notably, this acceptance does not appear to differ between the context of using hands-free or hands-on-wheel partial automation. Therefore, in conjunction with design safeguards, consumer information should convey the purpose of driver monitoring to promote its acceptance and adoption.

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 study was funded by the Insurance Institute for Highway Safety. The authors are grateful to Marel Montgomery, Teresa Schwemmer, Gary Blender, Amber Malmstead, David Manninen, and Mary Robinson for their help with data collection; Sarah Karush for her advice on survey design; Eric Teoh for statistical analysis guidance; and Teresa O'Connell for assistance with editing.

Appendix A. Supplemental analyses

See Tables A1-A4.

Table A1
Percent of respondents by lane-centering preference group for various outcomes that would be expected to occur when using hands-free compared with hands-on-wheel lane centering.

Compared with hands-on-wheel lane centering, hands-free lane centering would:		Would you want to use either type of lane-centering assistance?					
		Yes, but I prefer hands-free	Yes, but I prefer hands-on-wheel	Yes, and I have no preference between the two types	No, I would not use either feature	Unsure	Total
		(n = 268)	(n = 364)	(n = 177)	(n = 158)	(n = 43)	(n = 1,010)
Make driving more or less stressful	Much more	50	27	10	37	23	32
	Somewhat more	22	38	26	32	26	30
	Same as hands-on-wheel	7	18	36	20	40	20
	Somewhat less	10	14	20	9	9	13
	Much less	10	3	7	1	2	5
Make you more or less comfortable	Much more	45	13	14	13	2	21
	Somewhat more	34	33	27	11	35	29
	Same as hands-on-wheel	11	19	34	16	28	20
	Somewhat less	6	25	21	26	19	19
	Much less	4	9	3	33	16	11
Make you more or less tired	Much more	32	12	6	13	14	17
	Somewhat more	25	26	19	25	21	24
	Same as hands-on-wheel	19	36	41	46	42	34
	Somewhat less	13	19	24	11	16	17
	Much less	11	7	10	4	7	8
Make driving more or less safe	Much more	43	16	13	5	7	21
	Somewhat more	31	30	30	12	12	27
	Same as hands-on-wheel	15	24	33	27	47	25
	Somewhat less	8	23	23	28	28	20
	Much less	3	8	2	27	7	8
Make driving more of less boring	Much more	33	14	8	18	14	19
	Somewhat more	27	31	25	22	21	27
	Same as hands-on-wheel	15	30	39	40	49	30
	Somewhat less	12	20	19	13	9	16
	Much less	13	5	8	7	7	8
Make you more or less distracted	Much more	33	20	8	34	21	24
	Somewhat more	28	39	36	23	28	33
	Same as hands-on-wheel	17	22	27	32	40	24
	Somewhat less	10	15	22	7	9	13
	Much less	11	5	7	4	2	7
Make you more or less likely to do non-driving-related activities	Much more	46	23	18	23	16	28
	Somewhat more	32	34	37	25	33	33
	Same as hands-on-wheel	14	28	33	34	37	26
	Somewhat less	4	9	7	6	7	7
	Much less	4	6	5	12	7	6

Note. Percentages may not sum to 100 due to rounding.

Table A2
Percent of respondents likely or unlikely to buy or lease a vehicle with hands-free or hands-on-wheel lane centering per lane-centering preference group.

If cost were not a problem, how likely would you be to buy or lease your next vehicle with:		Would you want to use either type of lane-centering assistance?					
		Yes, but I prefer hands-free	Yes, but I prefer hands-on-wheel	Yes, and I have no preference between the two types	No, I would not use either feature	Unsure	Total
		(n = 268)	(n = 364)	(n = 177)	(n = 158)	(n = 43)	(n = 1,010)
Hands-free lane centering	Extremely	61	19	21	6	9	28
	Moderately	24	23	27	8	19	21
	Somewhat	12	21	33	13	28	20
	Slightly	3	18	10	9	16	11
	Not at all	<1	19	10	65	28	20
Hands-on-wheel lane centering	Extremely	49	34	34	6	7	32
	Moderately	32	34	33	15	33	31
	Somewhat	16	19	24	14	33	19
	Slightly	3	9	5	18	14	8
	Not at all	<1	4	3	47	14	10

Note. Percentages may not sum to 100 due to rounding.

Table A3
Percent of respondents per lane-centering preference group reporting to be comfortable or uncomfortable with driver monitoring strategies for hands-free and hands-on-wheel lane centering.

		Would you want to use either type of lane-centering assistance?					
		Yes, but I prefer hands-free	Yes, but I prefer hands-on-wheel	Yes, and I have no preference between the two types	No, I would not use either feature	Unsure	Total
		(n = 268)	(n = 364)	(n = 177)	(n = 158)	(n = 43)	(n = 1,010)
In the case of **hands-on-wheel** lane-centering assistance, how comfortable would you be if the vehicle monitored you through:							
Sensors in the steering wheel to make sure your hands were on the wheel	Extremely comfortable	69	35	38	11	12	40
	Somewhat comfortable	24	40	37	14	28	30
	Neither comfortable nor uncomfortable	4	13	18	26	40	15
	Somewhat uncomfortable	2	9	6	23	16	9
	Extremely uncomfortable	1	2	2	25	5	6
A camera to capture what your hands were doing	Extremely comfortable	56	26	29	8	7	31
	Somewhat comfortable	29	37	29	8	26	28
	Neither comfortable nor uncomfortable	9	14	16	22	37	15
	Somewhat uncomfortable	4	17	19	27	16	15
	Extremely uncomfortable	2	7	7	34	14	10
A camera to make sure you were looking at the road	Extremely comfortable	59	27	25	11	5	32
	Somewhat comfortable	24	32	26	8	28	25
	Neither comfortable nor uncomfortable	9	15	17	18	37	15
	Somewhat uncomfortable	4	14	23	25	16	15
	Extremely uncomfortable	3	11	9	38	14	13
In the case of **hands-free** lane-centering assistance, how comfortable would you be if the vehicle monitored you through:							
A camera to capture what your hands were doing	Extremely comfortable	54	25	28	8	5	30
	Somewhat comfortable	28	34	29	10	30	28
	Neither comfortable nor uncomfortable	10	16	16	21	30	16
	Somewhat uncomfortable	4	15	18	24	26	14
	Extremely uncomfortable	3	10	9	37	9	12
A camera to make sure you were looking at the road	Extremely comfortable	57	26	28	11	12	31
	Somewhat comfortable	27	33	27	6	26	26
	Neither comfortable nor uncomfortable	8	12	15	21	33	14
	Somewhat uncomfortable	3	16	20	19	16	14
	Extremely uncomfortable	5	13	11	44	14	15

Note. Percentages may not sum to 100 due to rounding.

Table A4
Percent of respondents likely or unlikely to buy or lease a vehicle with hands-free or hands-on-wheel versions of driver-initiated and vehicle-initiated auto lane change per auto-lane-change preference group.

If cost were not a problem, how likely would you be to buy or lease your next vehicle with:		Would you want to use either type of automated-lane-change assistance?					
		Yes, but I would prefer for the vehicle to wait for me to tell it when to make the lane change (i.e., driver-initiated)	Yes, but I would prefer for the vehicle to make the lane changes on its own (i.e., vehicle-initiated)	Yes, and I have no preference between the two types	No, I would not use either feature	Unsure	Total
		(*n* = 268)	(*n* = 364)	(*n* = 177)	(*n* = 158)	(*n* = 43)	(*n* = 1,010)
Hands-free driver-initiated auto lane change	Extremely	32	32	34	3	4	25
	Moderately	27	44	24	5	15	23
	Somewhat	19	20	27	12	30	19
	Slightly	10	4	13	19	22	12
	Not at all	11	<1	3	60	28	20
Hands-free vehicle-initiated auto lane change	Extremely	30	29	30	4	2	23
	Moderately	25	43	28	3	13	22
	Somewhat	18	20	27	14	33	19
	Slightly	12	6	10	15	17	12
	Not at all	16	2	4	63	35	24
Hands-on-wheel driver-initiated auto lane change	Extremely	44	35	38	8	7	32
	Moderately	32	37	29	15	17	28
	Somewhat	20	23	28	18	41	22
	Slightly	3	6	3	18	20	8
	Not at all	<1	<1	2	40	15	11
Hands-on-wheel vehicle-initiated auto lane change	Extremely	43	31	38	8	7	31
	Moderately	32	42	33	11	20	28
	Somewhat	16	18	24	17	39	19
	Slightly	6	8	4	18	17	9
	Not at all	3	<1	1	45	17	13

Note. Percentages may not sum to 100 due to rounding.

References

Abraham, H., Lee, C., Brady, S., Fitzgerald, C., Mehler, B., Reimer, B., & Coughlin, J. F. (2017). Autonomous vehicles and alternatives to driving: Trust, preferences, and effects of age. In *Proceedings of the Transportation Research Board 96th Annual Meeting* (TRB'17).

Abraham, H., Reimer, B., & Mehler, B. (2018). Learning to use in-vehicle technologies: Consumer preferences and effects on understanding. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62*(1), 1589–1593. https://doi.org/10.1177/1541931218621359.

American Automobile Association, Inc. (2020). *Evaluation of active driving assistance systems*. https://newsroom.aaa.com/asset/report-fact-sheet-evaluating-active-driver-assistance-systems/.

American Automobile Association, Inc. (2022). *Effectiveness of driver monitoring systems*. https://newsroom.aaa.com/asset/technical-report-effectiveness-of-driver-monitoring-systems.

Banks, V. A., Eriksson, A., O'Donoghue, J., & Stanton, N. A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. *Applied Ergonomics, 68*, 138–145. https://doi.org/10.1016/j.apergo.2017.11.010.

Banks, V. A., & Stanton, N. A. (2016). Keep the driver in control: Automating automobiles of the future. *Applied Ergonomics, 53*(Part B), 389–395. https://doi.org/10.1016/j.apergo.2015.06.020.

Biondi, F., Lohani, M., Hopman, R., Mills, S., Cooper, J.M., & Strayer, D. (2018). 80 mph and out-of-the-loop: Effects of real-world semi-automated driving on driver workload and arousal. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62*, 1878–1882. https://doi.org/10.1177/1541931218621427.

Carsten, O., & Martens, M. (2019). How can humans understand their automated cars? HMI principles, problems and solutions. *Cognition, Technology & Work, 21*, 3–20. https://doi.org/10.1007/s10111-018-0484-0.

Cicchino, J. B. (2017). Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. *Accident Analysis & Prevention, 99*(Part A), 142–152. https://doi.org/10.1016/j.aap.2016.11.009.

Cummings, M. L., & Bauchwitz, B. (2022). Safety implications of variability in autonomous driving assist alerting. *IEEE Transactions on Intelligent Transportation Systems, 23*, 12039–12049. https://doi.org/10.1109/TITS.2021.3109555.

Daziano, R. A., Sarrias, M., & Leard, B. (2017). Are consumers willing to pay to let cars drive for them? Analyzing response to autonomous vehicles. *Transportation Research Part C: Emerging Technologies, 78*, 150–164. https://doi.org/10.1016/j.trc.2017.03.003.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences, 113*(10), 2636–2641. https://doi.org/10.1073/pnas.1513271113.

Dunn, N., Dingus, T., Soccolich, S., & Horrey, W. (2021). Investigating the impact of driving automation systems on distracted driving behaviors. *Accident Analysis & Prevention, 156*. https://doi.org/10.1016/j.aap.2021.106152 106152.

El Khatib, A., Ou, C., & Karray, F. (2020). Driver inattention detection in the context of next-generation autonomous vehicles design: A survey. *IEEE Transactions on Intelligent Transportation Systems, 21*(11), 4483–4496. https://doi.org/10.1109/TITS.2019.2940874.

Gaspar, J., & Carney, C. (2019). The effect of partial automation on driver attention: A naturalistic driving study. *Human Factors, 61*(8), 1261–1276. https://doi.org/10.1177/0018720819836310.

General Motors. (n.d). How to use Super Cruise [web page]. https://www.cadillac.com/support/vehicle/driving-safety/driver-assistance/super-cruise/how-to.

Ghazizadeh, M., & Lee, J. (2018). Modelling driver acceptance: From feedback to monitoring and mentoring systems. In M. A. Regan, T. Horberry, & A. Stevens (Eds.), *Driver acceptance of new technology: Theory, measurement and optimisation* (pp. 51–70). CRC Press.

Goodall, N. J. (2021, October 21). A methodology for normalizing safety statistics of partially automated vehicles. Virginia Transportation Research Countil, University of Virginia. https://doi.org/10.31224/osf.io/m8j6g.

Halin, A., Verly, J., & Van Droogenbroeck, M. (2021). Survey and synthesis of state of the art in driver monitoring. *Sensors, 21*(16), 5558. https://doi.org/10.3390/s21165558.

Hardman, S. (2021). Investigating the decision to travel more in a partially automated electric vehicle. *Transportation Research Part D, 96*. https://doi.org/10.1016/j.trd.2021.102884 102884.

Highway Loss Data Institute (2021a). Advanced driver assistance systems on the 2017–19 Nissan Rogue. *HLDI Bulletin, 38*(1), 1–16.

Highway Loss Data Institute. (2021b). 2013–17 BMW collision avoidance features—A 2021 update. *HLDI Bulletin, 38*(26), 1–30.

Horrey, W., Lesch, M., Dainoff, M., Robertson, M., & Noy, I. (2012). On-board safety monitoring systems for driving: Review, knowledge gaps, and framework. *Journal of Safety Research, 43*(1), 49–58. https://doi.org/10.1016/j.jsr.2011.11.004.

Insurance Institute for Highway Safety. (2018). IIHS examines driver assistance features in road, track tests. *Status Report, 53*(4), 3–6. https://www.iihs.org/news/detail/iihs-examines-driver-assistance-features-in-road-track-tests.

Insurance Institute for Highway Safety. (2022). *IIHS creates safeguard ratings for partial automation*. https://www.iihs.org/news/detail/iihs-creates-safeguard-ratings-for-partial-automation.

Kidd, D., & Reagan, I. (2018). System attributes that influence reported improvement in drivers' experiences with adaptive cruise control and active lane keeping after daily use in five production vehicles. *International Journal of Human-Computer Interaction, 35*, 972–979. https://doi.org/10.1080/10447318.2018.1561786.

Kim, H., Song, M., & Doerzaph, Z. (2022). Is driving automation used as intended? Real-world use of partially automated driving systems and their safety consequences. *Transportation Research Record, 2676*(1), 30–37. https://doi.org/10.1177/03611981211027150.

Koppel, S., Charlton, J., Fildes, B., & Fitzharris, M. (2008). How important is vehicle safety in the new vehicle purchase process? *Accident Analysis & Prevention, 40*(3), 994–1004. https://doi.org/10.1016/j.aap.2007.11.006.

Lee, C., Gershon, P., Reimer, B., Mehler, B., & Coughlin, J. F. (2021). Consumer knowledge and acceptance of driving automation: Changes over time and across age groups. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 65*(1), 1395–1399. https://doi.org/10.1177/1071181321651116.

Melnicuk, V., Birrell, S., Thompson, S., Mouzakitis, A., and Jennings, P. (2019). How Acceptable Is It to Monitor Driver State? Using the UTAUT Model to Analyse Drivers' Perceptions Towards the System. In: Stanton, N. (Ed.) Advances in Human Aspects of Transportation: Proceedings of the AHFE 2018. International Conference on Applied Human Factors and Ergonomics. (pp. 579–590). Springer International Publishing. https://doi.org/10.1007/978-3-319-93885-1_53.

Mueller, A. S., Reagan, I. J., & Cicchino, J. B. (2021). Addressing driver disengagement and proper system use: Human factors recommendations for Level 2 driving automation design. *Journal of Cognitive Engineering and Decision Making, 15*, 3–27. https://doi.org/10.1177/1555343420983126.

National Transportation Safety Board. (2017). C*ollision between a car operating with automated vehicle control systems and a tractor- semitrailer truck near Williston, Florida, May 7, 2016* (Highway Accident Report NTSB/HAR-17/02).

National Transportation Safety Board. (2019). *Collision between car operating with partial driving automation and truck- tractor semitrailer: Delray Beach, Florida, March 1, 2019* (Highway Accident Brief NTSB/HAB-20/01).

National Transportation Safety Board. (2020). Collision between a sport utility vehicle operating with partial driving automation and a crash attenuator: Mountain View, California, March 23, 2018 (Report No. Highway Accident Report NTSB/HAR-20/01).

Navarro, J., Hernout, E., Osiurak, F., & Reynaud, E. (2020). On the nature of eye-hand coordination in natural steering behavior. *PLoS one, 15*(11), e0242818. https://doi.org/10.1371/journal.pone.0242818.

Preston, B. (2021). Auto industry and lawmakers call for driver monitoring systems to improve safety. *Consumer Reports.* https://www.consumerreports.org/car-safety/callfor-driver-monitoring-to-improve-car-safety-a3949626347/.

Reagan, I., Teoh, E., Cicchino, J., Gershon, P., Reimer, B., Mehler, B., & Seppelt, B. (2021). Disengagement from driving when using automation during a 4-week field trial. *Transportation Research Part F, 82*, 400–411. https://doi.org/10.1016/j.trf.2021.09.010.

SAE International. (2021). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles* (SAE Standard J3016, Report No. J3016_202104). https://doi.org/10.4271/J3016_202104.

Teoh, E. R. (2020). What's in a name? Drivers' perceptions of the use of five SAE Level 2 driving automation systems. *Journal of Safety Research, 72*, 145–151. https://doi.org/10.1016/j.jsr.2019.11.005.

Toyota. (n.d.). Your new partner for the road [web page]. https://www.toyota.com/mirai/teammate/.

Trimble, T. E., Baker, S., Russell, S., & Tidwell, S. (2020). *Consumer driving automation system education: A learning and retention assessment* (Report No. 20-UO-089). Virginia Tech Transportation Institute, National Surface Transportation Safety Center for Excellence.

Turner, M., Kitchenham, B., Brereton, P., Charters, S., & Budgen, D. (2010). Does the technology acceptance model predict actual use? A systematic literature review. *Information and Software Technology, 52*(5), 463–479. https://doi.org/10.1016/j.infsof.2009.11.005.

United States Census Bureau. (2020). *National demographic analysis tables: 2020*. https://www.census.gov/data/tables/2020/demo/popest/2020-demographic-analysis-tables.html.

Wang, Z., Zheng, R., Kaizuka, T., & Nakano, K. (2019). Relationship between gaze behavior and steering performance for driver–automation shared control: A driving simulator study. *IEEE Transactions on Intelligent Vehicles, 4*(1), 154–166. https://doi.org/10.1109/TIV.2018.2886654.

Wen, W., Kuroki, Y., & Asama, H. (2019). The sense of agency in driving automation. *Frontiers in Psychology, 2691*. https://doi.org/10.3389/fpsyg.2019.02691.

Alexandra Mueller is a senior research scientist at the Insurance Institute for Highway Safety. Dr. Mueller specializes in human factors research with a focus on advanced driver assistance systems and driving automation. She received her bachelor's and master's degrees from the University of Guelph and her doctorate from the University of Western Ontario. Dr. Mueller completed postdoctoral fellowships at McMaster University and the Cincinnati Children's Hospital Medical Center.

Jessica Cicchino is Vice President, Research at the Insurance Institute for Highway Safety. Since joining the Institute in 2012, she has authored papers on topics such as the effectiveness and use of advanced driver assistance systems, automated traffic enforcement, and the safety of older drivers, pedestrians, bicyclists, motorcyclists, and child passengers. Dr. Cicchino holds a B.A. from Vassar College and a Ph.D. in psychology from Carnegie Mellon University.

Joe Calvanelli Jr. is President of OpinionAmerica Group and has over 4 decades of experience in survey and market research. Mr. Calvanelli is also a Founding Partner and former Co-Director of the Fairleigh Dickinson University PublicMindTM Poll -- an opinion research center that to this day continues to conduct polling and survey research on politics, social policy, pop-culture, and consumer and economic trends. Mr.Calvanelli has a BS in Marketing and Management from Rider College (1978) and an MBA from Fairleigh Dickinson University (1985).

Journal of Safety Research 84 (2023) 18–23

Contents lists available at ScienceDirect

Journal of Safety Research

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

Crash characteristics for classic/historic vehicles and comparisons to newer vehicles

Daniel Kielminski [a], Elise Atkinson [b], Diane Peters [b], Seann Willson [a], Theresa Atkinson [b,*]

[a] Orthopaedic Surgery, McLaren Hospital, 401 S. Ballenger Hwy, Flint, MI 48532, United States
[b] Kettering University, 1700 University Ave., Flint, MI 48504, United States

ARTICLE INFO

Article history:
Received 26 April 2022
Received in revised form 18 July 2022
Accepted 17 October 2022
Available online 28 October 2022

Keywords:
Fatality
Classic car
Model year
Crash
Driver

ABSTRACT

Introduction: Older vehicles, commonly referred to as "classic," "vintage," or "historic" vehicles (CVH), share the roadways with newer vehicles. Older vehicles lacking safety systems likely come with an increased risk of fatality, however there is no study examining the typical conditions for crashes involving CVH. *Method:* This study utilized information from crashes occurring in 2012 to 2019 to estimate fatal crash rates for vehicles grouped by model year deciles. Data from crashes documented in the National Highway Traffic Safety Administration's (NHTSA) FARS and GES/CRSS data sets were utilized to examine roadway, temporal, and crash types for passenger vehicles produced in 1970 or earlier (CVH). *Results*: These data show CVH crashes are rare (<1% of crashes), but carry a relative risk of fatality from 6.70 (95th CI: 5.44–8.26) for impacts with other vehicles, which was the most common crash, to 9.53 (7.28–12.47) for rollovers. Most crashes occurred in dry weather, typically during summer, in rural areas, most frequently on two lane roads, and in areas with speed limits between 30 and 55 mph. Factors associated with fatality for occupants in CVH included alcohol use, lack of seat belt use, and older age. *Conclusions and Practical Applications:* Crashes involving a CVH are a rare event but have catastrophic consequences when they do occur. Regulations that limit driving to daylight hours may lower the risk of crash involvement, and safety messaging to promote belt use and sober driving may also help. Additionally, as new "smart" vehicles are developed, engineers should keep in mind that older vehicles remain on the roadway. New driving technologies will need to safely interact with these older, less safe vehicles.

1. Introduction

Over the past decades, vehicle fatalities in the United States have generally declined, not only in absolute fatalities but also when weighted by population levels and by mile driven (National Center for Statistics and Analysis, 2020a; Weast, 2018). Changes in traffic fatalities have been attributed to factors including vehicle design, road design, laws, and changes in driver behavior (Farmer & Lund, 2015; Stewart, 2020; Santaella-Tenorio et al., 2017). Although vehicle safety continues to improve, there are older model vehicles lacking safety systems that continue to share the roadway with newer vehicles. One such group is the so-called "classic," "vintage," or "historic" vehicles (CVH). Across the United States, there are clubs and organizations for owners to share their love of these vehicles, but there are limited recommendations for how to enjoy these vehicles while reducing the risk of severe injury or fatality. There are no national standards in the United States that regulate access to roads for these CVH, however, according to Hagerty, a company that specializes in insuring these CVH vehicles, 36 of the 50 states do impose some driving restrictions. Typically, these either limit driving to certain days or enforce a mileage per year cap (Fitzgerald, 2019). Eleven states require some sort of safety inspection (Fitzgerald, 2019). Other countries including India, France, and England have recently adopted regulations intended to limit the use of older vehicles, primarily due to concerns regarding vehicle emissions. However these, for the most part, exclude the very old "classic" vehicles. Going forward, as vehicles and traffic systems change due to vehicle electrification and smart driving technologies, it is possible that more governments may consider restricting roadway access for older vehicles. These may consider performance concerns related to either environmental or roadway safety.

While there is no study specifically examining these very old vehicles, there is some information in the literature that demonstrates an increased risk of fatality for occupants of older vehicles

* Corresponding author.
E-mail address: tatkinso@kettering.edu (T. Atkinson).

https://doi.org/10.1016/j.jsr.2022.10.004
0022-4375/

involved in crashes. A 2009 study found both severe injury and fatality occurrence were higher in older model vehicles, but these were 1994 model year (MY) or newer (Ryb et al., 2009). A study of single-vehicle crashes in 2003–2010 in Australia indicated fatality and injury were more frequent in vehicles produced before 1996, but the earliest vehicles included were 1991 MY (Anderson & Searson, 2015). A 2012 study of United States crashes focused on vehicles produced in a similar period (1990–2009) found increased HARM (a composite score based on injuries and the cost to treat injury) for older vehicles for all types of crashes (Eigen et al., 2012). More recently, Høye (2019) examined older vehicles in crashes in Norway and found that the risk of fatality and significant injury increases with vehicle age. The data also suggested that rollovers and impacts with fixed objects like trees pose an especially high fatality risk to an occupant in these older vehicles. While these data demonstrate, not surprisingly, that older vehicles are less safe in crashes than newer vehicles, little is known regarding the frequency and typical conditions for crashes involving CVH vehicles.

The current study seeks to estimate the frequency of crashes involving older vehicles within the United States and estimate the rate of fatal crashes, with special attention to crashes involving CVH vehicles, defined as those with model years of 1970 or earlier. It will also document common road and environmental factors associated with CVH vehicle crashes. The CVH group definition used in this study is intended to identify a vehicle group that is primarily composed of vehicles driven for pleasure, excluding those older vehicles that are driven due to economic pressures, which may have different crash involvement characteristics (Høye, 2019). These data will help vehicle enthusiasts and governmental groups better understand crash frequency and fatality risk for CVH vehicles in comparison to risks in modern vehicles. This information can be used in the development of safety recommendations and future vehicle regulations.

2. Methods

This study utilized information from crashes occurring in the United States from 2012 to 2019 to estimate yearly fatal crash rates for vehicles grouped by model-year deciles. Data from crashes occurring in 2016 to 2019 was used to examine event factors associated with classic/vintage/historic (CVH) vehicles. The data were extracted from three data sets maintained by the National Highway Traffic Safety Administration (NHTSA): Fatality Analysis Reporting System (FARS), the National Automotive Sampling System-General Estimates System (GES), and Crash Report Sampling System (CRSS), which are available online at https://www.nhtsa.gov/file-downloads?p=nhtsa/downloads/. FARS is a NHTSA data source aggregating annual fatal crash data for all qualifying fatality crashes within the 50 states and the District of Columbia. Crashes must involve a motor vehicle traveling on a public roadway and must have resulted in the death of at least one motorist or non-motorists within 30 days of the collision to be documented in the dataset. This dataset was used to estimate fatality risk and examine the most severe types of crashes for CVH vehicles. Both the GES and CRSS data sets are stratified proportional samples of crashes in the United States. The CRSS data set replaced the GES system in 2016 and follows a similar, but not the same, data sampling strategy (National Center for Analysis and Statistics, 2020b). Crashes selected for inclusion in these data sets provide information from police reports describing the events during the crash, roadway conditions, vehicle damage, and occupant information including restraint use. These data sets document roughly 50,000 cases each year. Both GES and CRSS data sets provide weighting factors, based on the probability of selecting the crash for inclusion in the sample, and these weights allow researchers to make estimates of yearly numbers of crashes. For a crash event to be a candidate for documentation in the dataset it must include either police-reported fatalities, injury, or property damage. Information from the GES and CRSS data provides an estimate of yearly numbers of crashes for the crash rate estimates in the current study and descriptions of the crash events. All crashes involving passenger vehicles were extracted from each data set where passenger vehicles were defined using the vehicle body type codes ("cars": 1–10; "SUVs": 14, 15; "vans": 20, 21, and "trucks": 34).

The risk of fatality in CVH vehicles was compared to the risk in newer vehicles using relative risk. Risk is defined as the number of events divided by the number of exposures. The exposure level was defined using the weighted CRSS crash counts (Teoh & Lund, 2011). The relative risk measure was selected as the risk descriptor to control for potential bias inherent in using the CRSS crash counts as the exposure measure. For example, there may be bias as this approach assumes that cars are on the road as frequently as they are involved in crashes, with crash involvement equaling exposure.

All statistical analysis utilized the 2016 through 2019 CRSS data. A chi-square test (or Fishers Exact test when case counts were below five for CHV) was used to identify a statistically significant difference in frequencies of events between CVH and newer vehicles for the fatal crash data. The surveyfreq tool in SAS (SAS Institute, Cary, NC) was utilized to account for the CRSS stratified sample design when calculating frequencies and 95th confidence intervals for the broader sample of all types of crashes involving CVH (National Center for Analysis and Statistics, 2020b). A Rao-Scott chi-square test was utilized to account for sample design in making comparisons of proportions from the CRSS data.

3. Results

Crashes involving CVH vehicles represent a small but constant event each year, constituting between 0.02 and 0.06% of yearly crashes (Fig. 1).

Fatal crashes involving these types of vehicles occurred in all states except Maine, Connecticut, and Delaware during the 2012–2019 timeframe (Fig. 2), with the highest number of crashes occurring in California (98) and Texas (31). However, the highest rate of involvement per population occurred in South Dakota at 0.56 crashes per million residents per year.

While these were rare events, a larger proportion of the crashes resulted in a fatality as compared to newer vehicles (Fig. 3).

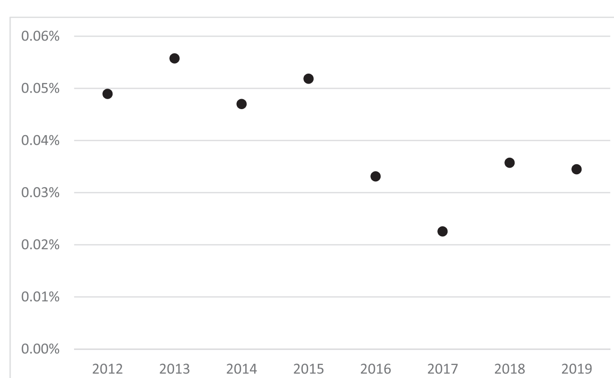


Fig. 1. Percentage of crashes involving CVH vehicles. Note that 2012–2015 is taken from NASS-GES and the 2016 – 2019 data is taken from the replacement national surveillance data set CRSS, and the sample design for these two data are slightly different.

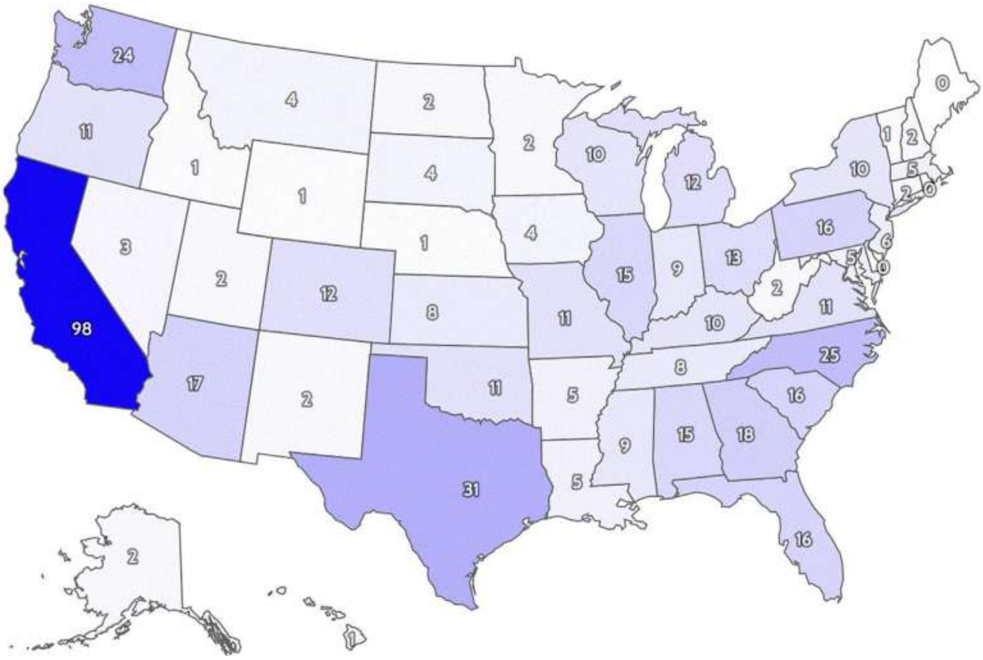


Fig. 2. Locations and counts of fatal CVH vehicle crashes from 2012 to 2019.

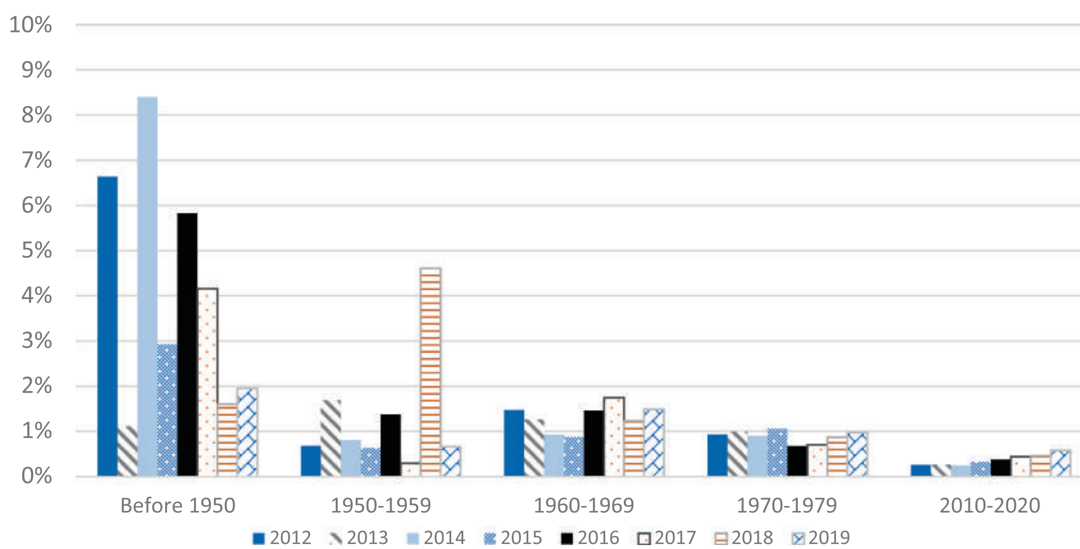


Fig. 3. Proportion of crashes resulting in a fatality in the vehicle (vehicles with a fatality due to the crash in each decile group from FARS / all crashes in decile group as estimated from GES or CRSS).

The risk of fatality for vehicle crashes involving a CVH vehicle was increased in comparison to vehicles in the 2010–2020 model year group, based yearly average fatality and crash counts across the eight years. Note that the proportion of fatal crashes varies year to year in the CVH vehicles due to differences in both the number of these vehicles included in the GES or CRSS sample set and the low frequency of these crashes, which makes yearly estimates highly sensitive to small differences. However, fatality rates were more consistent for the 1960 and 1970 deciles where the sample size tended to be larger where, for example, the 1950 s group had an estimated crash mean of approximately 700 cases per year versus 9,000 for the 1970 s group.

A more detailed analysis of crashes focused on crashes occurring in 2016–2019. The data from FARS, for all crashes producing a fatality in at least one of the vehicles involved in the event, included a total of 234 CVH and 151,322 newer vehicles. Note that a subset of these crashes included a fatality in the vehicle of interest and were included in the fatality cases shown in Fig. 3 and subsequent fatality analysis. The data used to describe the overall crash frequency (CRSS) included a weighted estimate of 15,656 (127 raw count) CVH and 4,095,173 (3,033,363 raw count) newer vehicles. These data indicated that the most frequent crash type for CVH vehicles was impact with another vehicle, which accounted for 76.8% [95th CI: 66.5–87.2%] of crashes. In comparison, these crashes accounted for 84.0% of crashes for newer vehicles, which was not significantly different ($p = 0.1159$), suggesting that vehicle to vehicle interactions are an important crash mode for these very old vehicles. The largest difference in crash type frequency was in impacts with parked cars where these occurred more frequently in CVH vehicles as compared to newer vehicles, at 7.6% [0 0.0 – 17.6%] of crashes for CVH versus 2.0% (p = 0.0270). Rollover crashes were estimated to occur at similar rates for both CVH and newer vehicles (0.6% [0.0–1.32%] of crashes vs 1.0%, p = 0.4198) but rollovers accounted for a significantly higher proportion of the fatal crashes in CVH vehicles: 22.4% of fatal crashes for CVH versus 10.1% for newer (p < 0.001). Overall, a significantly higher proportion of vehicle crashes in CVH vehicles resulted in “apparent” or more serious injury (based on assessment in the police record as indicated in the CRSS data), as compared to newer model-year vehicles (25.6% [17.6–33.6%] vs 18.5%; p = 0.0376), however detailed injury data are not collected in the data sets, so it is not possible to make more detailed injury assessments.

The relative risk of fatality in a CVH vehicle in comparison to newer vehicles in rollovers, impacts with narrow objects, and impacts with other vehicles ranged from 6.7 to 12.1 (Table 1). However, these were rare events accounting for less than 1% of fatalities for all vehicles (regardless of model year) involved in similar crashes each year.

In considering the road and environmental factors associated with CVH vehicle crashes, most occurred in dry weather, with the proportion of dry weather crashes similar to that for newer vehicles, at 85.5% [75.5–95.5%] of crashes for CVH versus 82.6% for newer vehicles (p = 0.5774). The types of roads for the CVH vehicle crashes were also similar to those for newer vehicles. For example, a higher proportion of crashes occurred on two lane roads, at 54.4% [42.7–66.0%] versus 46.3% for newer vehicles, and a lower proportion on roads with four or more lanes (23.7% [10.9–36.5%] vs 32.2%), but these differences were not significant (p = 0.1565 and 0.1699, respectively). No CVH vehicle crash in the surveillance sample of CRSS occurred on a roadway with a speed limit above 65 mph. Most crashes occurred on roads with speed limits between 30 and 55 mph (95.2% [89.8–100.0%]), as compared to 88.3% of crashes in newer vehicles (p = 0.0848). An estimated 72.1% [57.9–86.4%] of CVH crashes occurred in rural areas, and an estimated 2.68% of these rural crashes resulted in a fatality versus 1.14% for urban crashes. The CVH vehicle crashes tended to occur in the summer months, with an additional peak in February, which differed from the relatively flat distribution of crashes across months for the newer vehicles (Fig. 4).

Most fatal CVH vehicle crashes occurred in the late afternoon/evening with an additional peak at midnight (Fig. 5).

When considering factors related to the occupants of the vehicle, they demonstrate characteristics known to increase the risk of fatality in the crash event. For example, alcohol was a factor in 4.0% [0–8.4%] of crashes for CVH vehicles, which was not significantly different from 2.5% for crashes involving newer vehicles (p = 0.3702). However, pooling 2016–2019 data for all crashes producing a fatality in the CVH vehicle showed that 24.4% of these crashes involved driver alcohol use, which was significantly higher than 18.1% in similar crashes involving newer vehicles (p = 0.017). Seat belt use was also significantly lower in fatal crashes involving CVH at 47.7% versus 74.2% for the newer vehicles (p < 0.001), but in the CVH group belt use was at 85% in those who survived the crash. The data set does not indicate whether the seat belt was present or not (many older vehicles were not designed with belts) and there is no information in FARS to indicate whether an after-market seat belt was in use. In 20.0% of CVH crashes that produced a fatality, speed played a role per the police report. While this was higher than the proportion of 17.4% for similar crashes with newer vehicles, this difference was not significant (p = 0.228). The mean age of occupants in the CVH vehicles involved in fatal crashes was 52.9 ± 18.8 years, which was older than the mean overall age of 38.4 ± 21.1 years (mean of yearly means) for occupants in all fatal crashes. Across all fatal crashes, the average age of the occupants who died because of the crash was older than the average for those who survived (43.8 vs 35.2 years), and this same trend was observed when the CVH vehicles were considered (57.6 vs 42.9 years).

Table 1
Relative risk of fatality in a crash for CVH vs newer vehicles.

Crash Type	RR	95th CI	% Fatal Crashes
Tree	12.10	15.54–16.40	0.3%
Rollover	9.53	7.28–12.47	0.4%
Other Vehicle	6.70	5.44–8.26	0.2%
Utility/Light pole	9.67	5.44–17.19	0.5%
Parked Car	1.67	0.42–6.68	0.2%
Other	4.42	2.94–6.64	1.2%

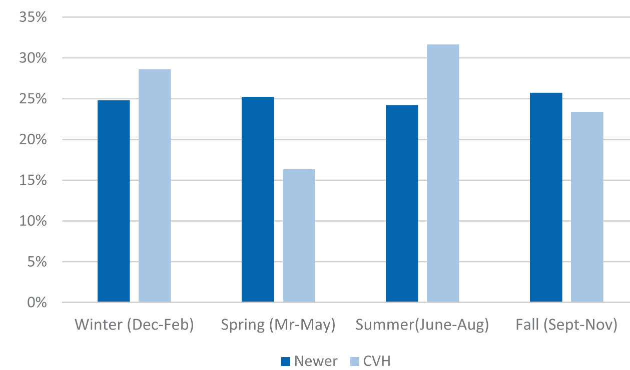


Fig. 4. Estimated proportion of crashes by season from CRSS data.

4. Discussion

The present study aims to examine the frequency of crashes involving CVH vehicles in the United States, with special attention to the type of crash and roadway conditions associated with these events. Crashes involving CVH most frequently involve impact with another vehicle and occur in rural areas in summer months on dry pavement. Occupants of these vehicles experience injuries due to the crash more frequently than occupants in newer vehicles. While crashes involving CVH are rare events, these crashes lead to a higher relative risk of fatality as compared to newer vehicles. Alcohol, lack of seatbelt use, and age were factors associated with fatalities in a CVH crash event.

The current study suggests that crashes involving CVH are more frequent during the summer months, with most crashes occurring during dry weather. This is consistent with the idea of leisure use of CVH, where drivers can select optimal times for their use. Also, the CRSS data suggest that users tend to select rural roadways and locations with speed limits of 55 or lower, which perhaps drivers perceive as safer. Note that this study used crash event observations as a surrogate for exposure, as there are no available data to fully account for roadway exposure for these vehicles. This may result in some bias in the data, however the observations related to the crash events were consistent year to year. Some state and local laws limit CVH vehicle use, which may contribute to the relatively low incidence of these crashes.

The results from this study are consistent with other vehicle safety research. A study in 2019 looked at alcohol usage during motor-vehicle collisions and survivability. This study showed that alcohol is an independent risk factor for mortality, with every age group faring worse in a motor-vehicle collision when compared to sober drivers (Culhane et al., 2019). The current study found that alcohol use was high in the CVH crashes that produced a fatality. Previous work indicated that alcohol was a variable involved in

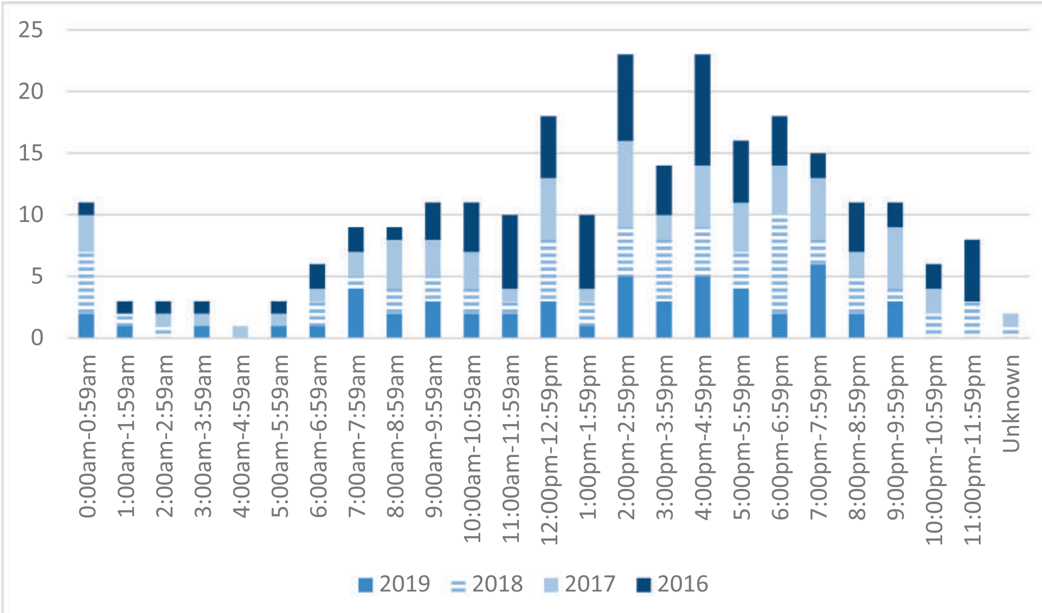


Fig. 5. Time of day for CVH vehicle fatal crash.

increasing the severity of driver injuries across multiple types of crashes (Li et al., 2019). Another factor associated with fatality is the lack of seat-belt usage. The observed use in CVH crashes that produced a fatality (47.7%) was far lower than the national estimated use rate of 90% (National Center for Statistics and Analysis, 2021). Lower usage may be due to lower risk perception in drivers of these vehicles (Sheveland et al., 2020) or may be related to alcohol use (Marco et al., 2020). In 2014, self-report seat belt usage in the United States was 86.9% overall with the highest (88.8%) in urban counties and the lowest (74.7%) in rural counties (Beck et al., 2014). Rurality was associated with increasing severity of injuries for drivers involved in MVCs and reasons cited for this include lower traffic densities, fewer traffic control facilities, and law enforcement (Li et al., 2019). Several other factors may contribute to the increased number of fatalities in rural/open roads such as lower belt use, high speeds, roadway characteristics, and lack of appropriate trauma care in a reasonable timeframe (Beck et al., 2014). Speed was a factor in CVH fatal crashes at a rate similar to that in newer vehicles. A recent study confirmed that a majority of older drivers occasionally exceed the posted speed limit (Cull et al., 2020). Another factor associated with fatal CVH vehicle crashes was the higher average age of occupants. Rolison and Moutari (2018) focused on drivers 60 and older and found drivers 70 and older had higher risk of crash involvement, even after accounting for risk exposure. Another recent study demonstrated a similar finding where more severe injuries were associated with drivers 65 and older, with these outcomes attributed to chronic medical conditions and vision, cognition, or mobility impairments (Li et al., 2019). Involvement in motor-vehicle collision and impaired motion perception in older adults have also been linked (Swain et al., 2021). Additionally, Høye (2017) found that drivers of older vehicles tended to show higher risk-taking behavior including drunk driving, speeding, and being unbelted compared to drivers of newer vehicles. This terrible triad of factors are those generally associated with crash fatalities (Shyhalla, 2014), but the current study did not explore effects of combined risk factors.

The data presented in the current study suggest that favorable driving conditions do not necessarily prevent crashes. However, as fatal crash events were observed to occur more frequently in evening and late-night times, an effective risk reduction approach might involve limiting the CVH vehicle roadway access to daylight hours. The data also indicate that driver choices may contribute to fatality risk in these crashes. It may also be helpful to increase public awareness of the risks associated with use of so called "classic" cars and promote safe driving behaviors. The study findings can also be extrapolated to consider the future roadway. This study demonstrated that significant numbers of older vehicles, including CVH, continue to utilize the U.S. roadway. In the near future, vehicles with autonomous driving and vehicle to vehicle communication technology will become more common. This poses many questions and possibilities regarding their interaction with other non-autonomous vehicles sharing the roadways. CVH have different shapes and may behave differently on the road in comparison to their modern counterparts due to differences in vehicle handling capabilities and driver factors. When looking forward, it is difficult to predict how driver-assisted technology may behave when encountering a CVH on the road. It may be possible to design systems that identify older vehicles and alter driving factors to reduce the risk of interactions and crashes with these vehicles, but this will only occur if manufacturers intentionally consider these events in the development of autonomous driving systems.

There are several limitations associated with our study. Firstly, the CVH vehicles were treated as a group without considering differences based on the specific vehicle type or model year. However, the study did not seek to attribute fatality risk to a particular type of vehicle, rather it sought to estimate fatality risk in comparison to newer vehicles. Secondly, it only considered crashes in the United States. However, it demonstrated decreasing risk for each decile of vehicle production moving towards 2020, consistent with other studies. Additionally, this study utilized national estimates for numbers of crashes as a surrogate for exposure, as no national survey of CVH usage is available. Using CRSS data is also a limitation because a vehicle involved in a crash must be towed away to qualify for the dataset, potentially creating bias. For example, an older vehicle that was not designed considering crash performance may be more likely to require towing after a crash and this may therefore increase the likelihood of that vehicle being sampled. The relative risk measure helps control for this concern as it has both new cars and old cars compared on the same scale. Another limitation is that the study could only report proportions of crashes resulting in "no apparent injury" compared to newer vehicles, rather than injury data. Detailed injury data for very old vehicles is not collected in U.S. federal surveillance data sets. Information involving injuries would be a useful assessment because crash survivors may have injuries associated with significant morbidity, which can affect daily living activities resulting in increased healthcare spending. A retrospective study from Victoria, Australia investigated the most common injuries in adults aged 65 or older involved in motor-vehicle accidents. The older adults were found to have a higher incidence of chest wall injuries such as rib fractures (Yee et al., 2006). Older patients from this study were also found to have a higher average ICU stay compared to younger patients as well as a higher in-hospital mortality rate (Yee et al., 2006).

5. Conclusions and Practical Applications

Overall, this study suggests that even though crashes involving a CVH are rare events, when these crashes occur they carry a high relative risk of fatality. The study found multiple occupant-based risk factors including older age of the occupants, alcohol use, and lack of seatbelt use. Given the very high relative risk of fatality associated with CVH crashes, it may be worthwhile to aim safety messaging toward their drivers and/or limit their access to the roadway. Based on crash data from recent years, these classic vehicles will continue to coexist with other, new vehicles on public roadways. As smart driving technologies are developed, it may be possible to explore new, innovative approaches that may help alleviate crash risks when operating CVH vehicles.

Funding

This study was performed without direct support.

Acknowledgments

The authors would like to thank Brian Smith, BS and Jack Mason, BS who helped with a preliminary analysis of classic car crashes and contributed their ideas to this study.

References

Anderson, R. W., & Searson, D. J. (2015). Use of age-period-cohort models to estimate effects of vehicle age, year of crash and year of vehicle manufacture on driver injury and fatality rates in single vehicle crashes in New South Wales, 2003–2010. *Accident Analysis & Prevention, 75*, 202–210. https://doi.org/10.1016/j.aap.2014.11.013. Epub 2014 Dec 6 PMID: 25485731.

Beck, L. F., Downs, J., Stevens, M. R., & Sauber-Schatz, E. K. (2014). Rural and Urban Differences in Passenger-Vehicle-Occupant Deaths and Seat Belt Use Among Adults - United States, 2014. Morbidity and Mortality Weekly report. *Surveillance Summaries (Washington, D.C. : 2002), 66*(17), 1–13. https://doi.org/10.15585/mmwr.ss6617a1. PMID: 28934184; PMCID: PMC5829699.

Culhane, J., Silverglate, B., & Freeman, C. (2019). Alcohol is a predictor of mortality in motor vehicle collisions. *Journal of Safety Research, 71*, 201–205. https://doi.org/10.1016/j.jsr.2019.10.002. Epub 2019 Nov 17 PMID: 31862031.

Cull, A. W., Porter, M. M., Nakagawa, S., Smith, G. A., Rapoport, M. J., Marshall, S. C., ... Mazer, B. (2020). Speeding and Speed Modification of Older Drivers: Does Vehicle Type Make a Difference? *Canadian Journal on Aging, 39*(3), 385–392. https://doi.org/10.1017/S0714980819000394. Epub 2019 Jun 25 PMID: 32723412.

Eigen, A. M., Digges, K. & Samaha, R. R. (2012). *Safety Changes in the US Vehicle Fleet since Model Year 1990, Based on NASS Data*. In Annals of Advances in Automotive medicine. Association for the Advancement of Automotive Medicine. Annual Scientific Conference. (Vol. 56. pp. 241-251). PMID: 23169134; PMCID: PMC3503428.

Farmer, C. M., & Lund, A. K. (2015). The Effects of Vehicle Redesign on the Risk of Driver Death. *Traffic Injury Prevention, 16*(7), 684–690. https://doi.org/10.1080/15389588.2015.1012584. PMID: 25664589.

Fitzgerald, C. (2019). How classic car-friendly is your state?. Hagerty Media. 2019 July. https://www.hagerty.com/media/archived/how-classic-car-friendly-is-your-state/. Retrieved October 22, 2021.

Høye, A. (2017). In-depth analysis of speed-related road crashes. TØI Report 1569/2017. 2017. Institute of Transport Economics., Oslo.

Høye, A. (2019). Vehicle registration year, age, and weight - Untangling the effects on crash risk. *Accident Analysis & Prevention, 123*, 1–11. https://doi.org/10.1016/j.aap.2018.11.002. Epub 2018 Nov 15 PMID: 30447490.

Li, Z., Wu, Q., Ci, Y., Chen, C., Chen, X., & Zhang, G. (2019). Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accident Analysis & Prevention, 129*, 230–240. https://doi.org/10.1016/j.aap.2019.04.001. Epub 2019 Jun 5 PMID: 31176143.

Marco, C. A., Ekeh, A. P., Hardman, C., Lovell, M., Brent, A., & Akamune, J. (2020). Seat belt use among patients in motor vehicle collisions: Clinical and demographic factors. *American Journal of Emergency Medicine, 38*(6), 1069–1071. https://doi.org/10.1016/j.ajem.2019.158367. Epub 2019 Jul 26 PMID: 31375356.

National Center for Statistics and Analysis. (2020a). Overview of motor vehicle crashes in 2019: Traffic safety facts research note. Report No. DOT HS 813 060. National Highway Traffic Safety Administration. Washington, D.C. US Department of Transportation.

National Center for Analysis and Statistics. (2002b). Crash report sampling system CRSS analytical user's manual 2016-2019. Report No DOT HS 813 022. National Highway Traffic Safety Administration. Washington, D.C. US Department of Transportation.

National Center for Statistics and Analysis. (2021). Seat belt use in 2020 – Overall results (Traffic Safety Facts Research Note. Report No. DOT HS 813 072. National Highway Traffic Safety Administration. . Washington, D.C. US Department of Transportation.

Rolison, J. J., & Moutari, S. (2018). Risk-Exposure Density and Mileage Bias in Crash Risk for Older Drivers. *American Journal of Epidemiology, 187*(1), 53–59. https://doi.org/10.1093/aje/kwx220. PMID: 28605422; PMCID: PMC5860223.

Ryb, G. E., Dischinger, P. C., & Ho, S. (2009). Vehicle model year and crash outcomes: A CIREN study. *Traffic Injury Prevention, 10*(6), 560–566. https://doi.org/10.1080/15389580903271401. PMID: 19916126.

Santaella-Tenorio, J., Mauro, C. M., Wall, M. M., Kim, J. H., Cerdá, M., Keyes, K. M., Hasin, D. S., Galea, S., & Martins, S. S. (2017). US Traffic Fatalities, 1985-2014, and Their Relationship to Medical Marijuana Laws. *American Journal of Public Health, 107*(2), 336–342. https://doi.org/10.2105/AJPH.2016.303577. Epub 2016 Dec 20. PMID: 27997245; PMCID: PMC5227945.

Sheveland, A. C., Luchman, J. N., Mendelson, J., Xie, J., Bleiberg, M. A., Eby, D. W., ... Walton, B. R. (2020). Psychological Constructs Related to Seat Belt Use: A Nationally Representative Survey Study. *Accident Analysis & Prevention, 148*. https://doi.org/10.1016/j.aap.2020.105715. Epub 2020 Oct 8 PMID: 33038864 105715.

Shyhalla, K. (2014). Alcohol involvement and other risky driver behaviors: Effects on crash initiation and crash severity. *Traffic Injury Prevention, 15*(4), 325–334. https://doi.org/10.1080/15389588.2013.822491. PMID: 24471355.

Stewart, T. (2020). Overview of Motor Vehicle Crashes in 2020. Report No. 813 266. National Highway Traffic Safety Administration.

Swain, T. A., McGwin, G., Jr, Wood, J. M., & Owsley, C. (2021). Motion perception as a risk factor for motor vehicle collision involvement in drivers $\geq$ 70 years. *Accident Analysis & Prevention, 151*. https://doi.org/10.1016/j.aap.2020.105956. Epub 2021 Jan 11. PMID: 33444870; PMCID: PMC7878317.

Teoh, E. R., & Lund, A. K. (2011). IIHS side crash test ratings and occupant death risk in real-world crashes. *Traffic Injury Prevention., 12*(5), 500–507. https://doi.org/10.1080/15389588.2011.585671. PMID: 21972861.

Weast, R. (2018). Temporal factors in motor-vehicle crash deaths: Ten years later. *Journal of Safety Research, 65*, 125–131. https://doi.org/10.1016/j.jsr.2018.02.011. Epub 2018 Apr 17 PMID: 29776521.

Yee, W. Y., Cameron, P. A., & Bailey, M. J. (2006). Road traffic injuries in the elderly. *Emergency Medicine Journal, 23*, 42–46. https://doi.org/10.1136/emj.2005.023754.

Daniel Kielminski, MD is a second year orthopaedic surgery resident at McLaren Flint hospital. His interest in motor vehicles and trauma prevention led him to join this project.

Elise Atkinson, BS is a mechanical engineering graduate at Kettering University. She has an interest in social justice and injury prevention. She joined this project to investigate motor vehicle crashes and factors associated with these crashes.

Diane Peters, PhD is a mechanical engineering associate professor at Kettering University. She is interested in future mobility issues and autonomous vehicles. She joined the project to enhance her understanding of vehicles sharing the road with autonomous vehicles.

Seann Willson, MD is an orthopaedic surgeon and faculty instructor for the McLaren Flint orthopaedic residency program. She is interested in classic vehicles and factors associated with serious crashes involving these vehicles. Her practice specializes in adult joint reconstruction but she cares for trauma patients and is interested in injury prevention.

Theresa Atkinson, PhD is a mechanical engineering associate professor at Kettering University. She is interested in injury prevention and orthopaedic trauma. She has published a range of studies examining injuries associated with motor vehicle crashes and child occupant protection. Her interest in the current study resulted from local news stories of severe crashes involving older vehicles.

Journal of Safety Research 84 (2023) 261–272

Contents lists available at ScienceDirect

Journal of Safety Research

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

Differences in perception of the importance of process safety indicators between experts in Iran and the West

Leila Omidi[a,*], Khadijeh Mostafaee Dolatabad[b], Colin Pilbeam[c]

[a] Department of Occupational Health Engineering, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
[b] Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran
[c] Cranfield Safety and Accident Investigation Centre, Cranfield University, Cranfield, UK

ARTICLE INFO

Article history:
Received 30 August 2021
Received in revised form 26 July 2022
Accepted 2 November 2022
Available online 10 November 2022

Keywords:
Process safety
Lagging indicators
Leading indicators
Process industries
Fuzzy best worst method

ABSTRACT

Introduction: The importance of safety in high-risk industries such as oil and gas facilities has been reported previously. Process safety performance indicators can provide insight into improving the safety of process industries. This paper aims to rank the process safety indicators (metrics) by Fuzzy Best-Worst Method (FBWM) using the data gathered through a survey. *Method:* The study uses a structured approach considering the UK Health and Safety Executive (HSE), the Center for Chemical Process Safety (CCPS), and the IOGP (International Association of Oil and Gas Producers) recommendations and guidelines to generate an aggregate set of indicators. It calculates the level of importance of each indicator based on the opinions of experts from Iran and some Western countries. *Results:* The findings of the study demonstrate that some lagging indicators such as the number of times processes do not proceed as planned due to insufficient staff competence and the number of unexpected disruptions of the process due to failure in instrumentation and alarms are important in process industries in both Iran and Western countries. Western experts identified process safety incident severity rate as an important lagging indicator, whereas Iranian experts considered this as relatively unimportant. In addition, leading indicators such as sufficient process safety training and competency, the desired function of instrumentation and alarms, and proper management of fatigue risk play an important role in enhancing the safety performance of process industries. Experts in Iran viewed permit to work as an important leading indicator, while experts in the West focused on fatigue risk management. *Practical Applications:* The methodology used in the current study gives a good view to managers and safety professionals in regard to the most important indicators of process safety and allows them to focus more on important process safety indicators.

1. Introduction

1.1. Background

The importance of safety in high-risk industries such as oil and gas facilities has been reported previously (Askarian et al., 2018; Moradi Hanifi et al., 2019; Omidi et al., 2021; Omidi et al., 2018). Process safety and risk analysis are generally considered to be of paramount significance in preventing fatalities and asset loss due to accidents (Amin et al., 2019). In order to monitor and improve the safety in process facilities and to provide ongoing assurance that major hazard risks are adequately controlled (HSE, 2006), process safety performance indicators (metrics) are applied (Khan et al., 2010).

Process safety performance indicators and the information they provide are required to create a safer process industry. Aggregating existing process safety indicators, sorting them into specific elements, determining their relative importance, and providing a risk score for each may not only help to reduce an over-abundance of indicators but also further reduce losses and improve safety. Reviewing existing indicators to define a small but effective number of indicators can reduce the effort required to collect necessary information (Pasman & Rogers, 2014). Simple and easy-to-use metrics and a small number of the best predictive indicators can improve the effectiveness of the safety management system (Khan et al., 2010; Sultana et al., 2019). In addition, implementing practical and actionable safety metrics in key areas can lead to improvements in performance outcomes and provide important information about the level of safety within the organization (Øien

* Corresponding author.
E-mail addresses: omidil@razi.tums.ac.ir (L. Omidi), k.mostafaee@modares.ac.ir (K.M. Dolatabad), colin.pilbeam@cranfield.ac.uk (C. Pilbeam).

https://doi.org/10.1016/j.jsr.2022.11.002
0022-4375/

et al., 2011; Stough, 2012). The development of process safety indicators can give early warnings and may help prevent major accidents in process industries (Sultana et al., 2019). However, the type of safety performance indicators used and their number vary heavily across industries and between countries.

The levels of safety, accepted levels of risks, and safety regulations are different between countries. Furthermore, the levels of safety culture in high-hazard industries in Western (industrialized) countries are different from the cultural contexts of developing countries. These differences may lead to different risk perceptions and the use of different safety indicators in high-hazard industries in developed and developing countries (Manzey & Marold, 2009).

1.2. Lagging and leading indicators

Two types of process safety indicators (lagging and leading) are identified in the literature (Sultana et al., 2019; Swuste et al., 2016). Lagging indicators are a form of reactive monitoring based on measures of undesired outcomes such as injuries, accidents, near misses, and process safety incidents (CCPS, 2011; HSE, 2006; Louvar, 2010). Lagging indicators need to be monitored but these indicators do not provide adequate forewarning for preventing accidents (Louvar, 2010).

In contrast, leading indicators are a form of active monitoring based on the routine systematic checking of key actions or activities within the risk control systems. They can be considered as measures essential to deliver the desired safety outcome (HSE, 2006). Leading indicators focus on the performance of key work processes, operating discipline, or layers of protection that prevent incidents (CCPS, 2011). These indicators provide an early warning to prevent process accidents (Louvar, 2010). The key characteristics of leading indicators offered in the literature include simplicity with a close connection to outcomes, readily interpretable by different groups in the same way, objectively and reliably measurable, easily and accurately communicated, and broadly applicable across company operations (Sinelnikov et al., 2015; Stough, 2012).

Both leading and lagging indicators provide insights into the level of safety of a system. Leading indicators are associated with potential barrier failures and are proxies for hazards, while lagging indicators are associated with failures after an incident and are proxies of the events (Sultana et al., 2019; Swuste et al., 2016). The development of process safety indicators is an effective strategy to provide early warnings for major accidents and to measure how safety is managed within installations (Sultana et al., 2019).

1.3. Process safety in developing and developed countries

Process safety can affect chemical and manufacturing industries in both developing and developed countries. Major process safety incidents that occurred between the 1970s and the 1990s led to the development of process safety management in developed countries (Besserman & Mentzer, 2017). Developing countries have also addressed and promoted process safety, but more recently. Typically, developed countries have better reporting procedures, process safety metrics, and more developed process regulations, such as the process safety management regulations established by the U.S. Occupational Safety and Health Administration (OSHA, 1992), for preventing and mitigating loss incidents. In contrast, new process safety regulations in developing countries are based on previous regulations in developed countries. These help developing countries use learnings from developed countries to protect workers, the public, and the environment. Moreover, developed countries have better emergency response, infrastructure, more enforcement of regulations, and lower fatality rates than developing countries. The reported job fatality rate per region by the International Labor Organization (ILO) in 2001 for the United Kingdom was 0.84 per 100,000 workers and for India and China was 9.97 and 12.31, respectively (Besserman & Mentzer, 2017). It appears from major hazard incidents records in 2007 that the consequences of major incidents (such as the probability of lethality) are significantly higher in developing countries than in developed countries (Hemmatian et al., 2014). More incident reports in developed countries are due to better reporting procedures. Therefore, developed regions and developing countries are at different points in the evolution of process safety, which provides a basis for comparison (Besserman & Mentzer, 2017).

1.4. Guidelines and recommended practices on process safety indicators

Following the Texas City explosion and fire at the BP site, several organizations such as the UK Health & Safety Executive (UK HSE), the Center for Chemical Process Safety (CCPS), the American Petroleum Institute (API), and the Organization for Economic Cooperation and Development (the OECD) have developed recommendations or guidelines on process safety indicators (Zhen et al., 2019). The UK HSE (2006) framework considers the two types of indicators to provide dual assurance to confirm that the risk control system is operating as intended or process safety risks are being effectively managed. In the CCPS (2008 and 2011) guidelines, three types of process safety performance metrics are described (i.e., lagging metrics, leading metrics, and near-miss metrics). The CCPS (2011) metric recommendations are consistent with the API documents and contain examples of leading metrics and related quantifiable parameters (Swuste et al., 2016; Zhen et al., 2019).

OECD published guidelines on safety performance indicators in two versions; one for industry and the other for public authorities and communities. In these documents, developed by a group of experts, safety metrics are defined and classified into result indicators (reactive or lagging indicators) and activity indicators (proactive or leading indicators) (OECD, 2008a, 2008b).

A recommended practice (RP) for the refining and petrochemical industries was issued by the API (ANSI/API, 2010, 2016). Process safety indicators in the RP are categorized into four tiers. Tiers 1 and 2 (corresponding to lagging indicators) are intended for process safety events and public reporting, and tiers 3 and 4 (corresponding to leading indicators) are related to challenges to safety systems and operating discipline and management system performance for internal use within individual facilities.

The International Association of Oil & Gas Producers (IOGP) provided further guidance on key performance indicators (OGP report no. 456) to support the applicability of the API RP 754 and to reduce and eliminate process risks (IOGP, 2016a; Zhen et al., 2019). Leading indicators in the report are linked to preventive barriers and the lagging indicators are linked to de-escalating barriers. The report provides further guidance on the HSE framework and the ANSI/API RP754 (Swuste et al., 2016; Zhen et al., 2019).

1.5. Prioritization and weighting method

Safety professionals in process industries have different perspectives on safety performance indicators. These lead them to attach different levels of importance to each indicator and to assign different weights to measurements. Assigning different weights to different indicators allows managers and safety professionals to formulate different strategies for improving process safety. The factors considered to be more influential may vary by country, encouraging the adoption of different process safety management strategies.

To accommodate this variation between perspectives, multi-criteria decision-making (MCDM) may be used (Salimi & Rezaei, 2018). During the past decade, MCDM methods have increasingly been used for dealing with uncertainties and solving engineering problems (Antucheviciene et al., 2015). MCDM methods are appropriate where there is uncertainty, for example through vagueness (due to the lack of complete information) or ambiguity (arising from the qualitative judgment of decision-makers) (Guo & Zhao, 2017). Consequently, they are helpful for tackling real-world issues that share these characteristics (Wang & Lee, 2009). The best-worst multi-criteria decision-making method (BWM), as a new MCDM method, was proposed by Rezaei (2015). Unlike other MCDM methods, the BWM obtains the weights of criteria and alternatives with respect to different criteria by using least pairwise comparisons. Extending BWM to the fuzzy environment (fuzzy BWM or FBWM) and the employment of fuzzy information may be a more appropriate way for tackling convoluted decision-making problems under an uncertain environment (Guo & Zhao, 2017; Hafezalkotob & Hafezalkotob, 2017). It is noteworthy that the BWM procedure seems to be much easier, more accurate, and less redundant than the conventional MCDM procedures because the method does not require secondary comparisons (Guo & Zhao, 2017; Rezaei, 2015).

1.6. Research purpose

The aim of this paper is to demonstrate the difference in ranking of process safety indicators between experts in Iran and in the West using FBWM and based on fuzzy preference comparisons. Specifically, the paper will:

i. use a structured approach considering the UK HSE, the CCPS, and the IOGP recommendations and guidelines to aggregate the indicators and to identify a reduced number of suitable indicators for process safety;
ii. capture perceived importance of process safety indicators from experts in Iran and the West;
iii. describe and apply FBWM to evaluate two sets of indicators including lagging and leading indicators;
iv. account for differences in expert perceptions between Iran and the West.

Experts' subjective evaluations of process safety indicators are anticipated to reflect the focus and the level of process safety and related indicators in Iran and Western countries, permitting comparison.

2. Method

The importance of process safety indicators has been addressed in scientific literature and in the reports of national and international organizations (Swuste et al., 2016). This study is based on the UK HSE guideline, the CCPS recommendations, and the IOGP guideline. These guidelines and recommendations consist of process safety indicators that are scientifically designed to consider process sensitivity, measurable values, and monitorable parameters, and contain easy-to-use metrics (Khan et al., 2010).

Process safety indicators classified into leading and lagging indicators were ranked by experts. The experts were experienced staff within the field of process safety and involved in the process industries in Iran (as a developing country) and Western countries (Western Europe and the United States) (as developed countries). Fig. 1 presents the safety practitioners' working experience. Almost 50% of the Iranian respondents had more than 10 years of work experience. Among the Western experts, 60% had more than 25 years of experience.

2.1. The basis for the study of lagging and leading indicators

Definitions for lagging and leading indicators were drawn from the UK HSE, the CPPS, and the IOGP (Fig. 2). In this study, some indicators from the HSE guide such as the number of incidents or unexpected disruption of process due to deficiencies in plant change and permit to work were considered as lagging indicators and the percentage of successful process implementation due to the appropriate inspection/maintenance and the appropriate level of staff competence were regarded as leading indicators (HSE, 2006). Process Safety Total Incident Rate (PSTIR) and Process Safety Incident Severity Rate (PSISR) were considered as lagging metrics in CCPS recommendations (CCPS, 2011), and used here. In addition, three safety performance indicators including fatal accident rate (FAR), total recordable injury rate (TRIR), and lost time injury frequency (LTIF) from the IOGP were considered as other lagging indicators (Fig. 2) (IOGP, 2016b, 2019).

2.2. Procedure

After determining the indicators from related guidelines, these were weighted by experts who have worked in the oil and gas industries in Iran or Western countries, in a comparative study was conducted to weight the indicators by experts who have had

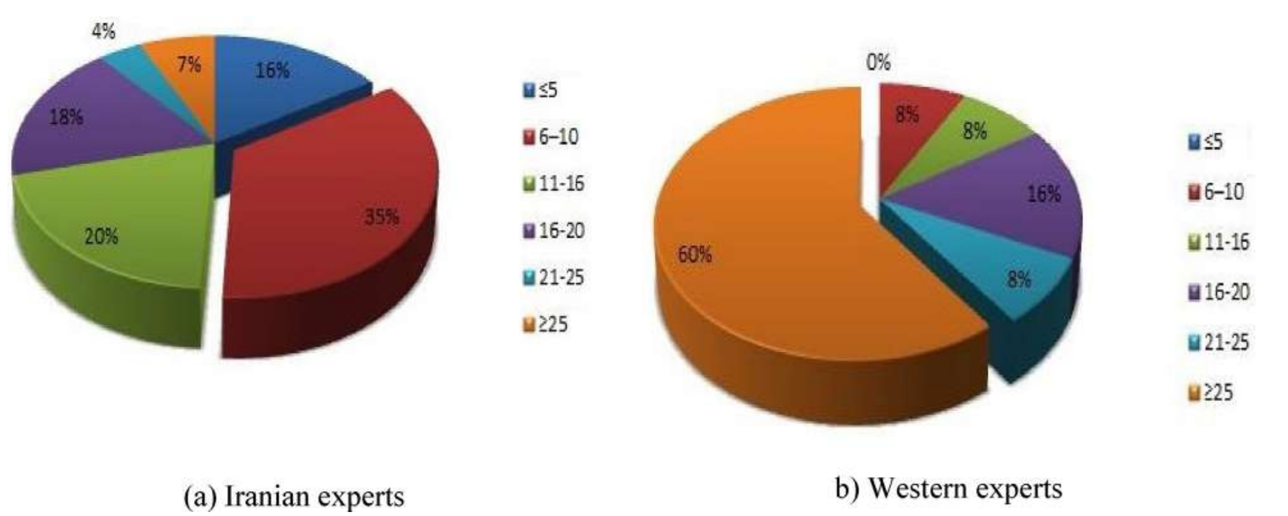


Fig. 1. Distribution of experts based on their work experience.

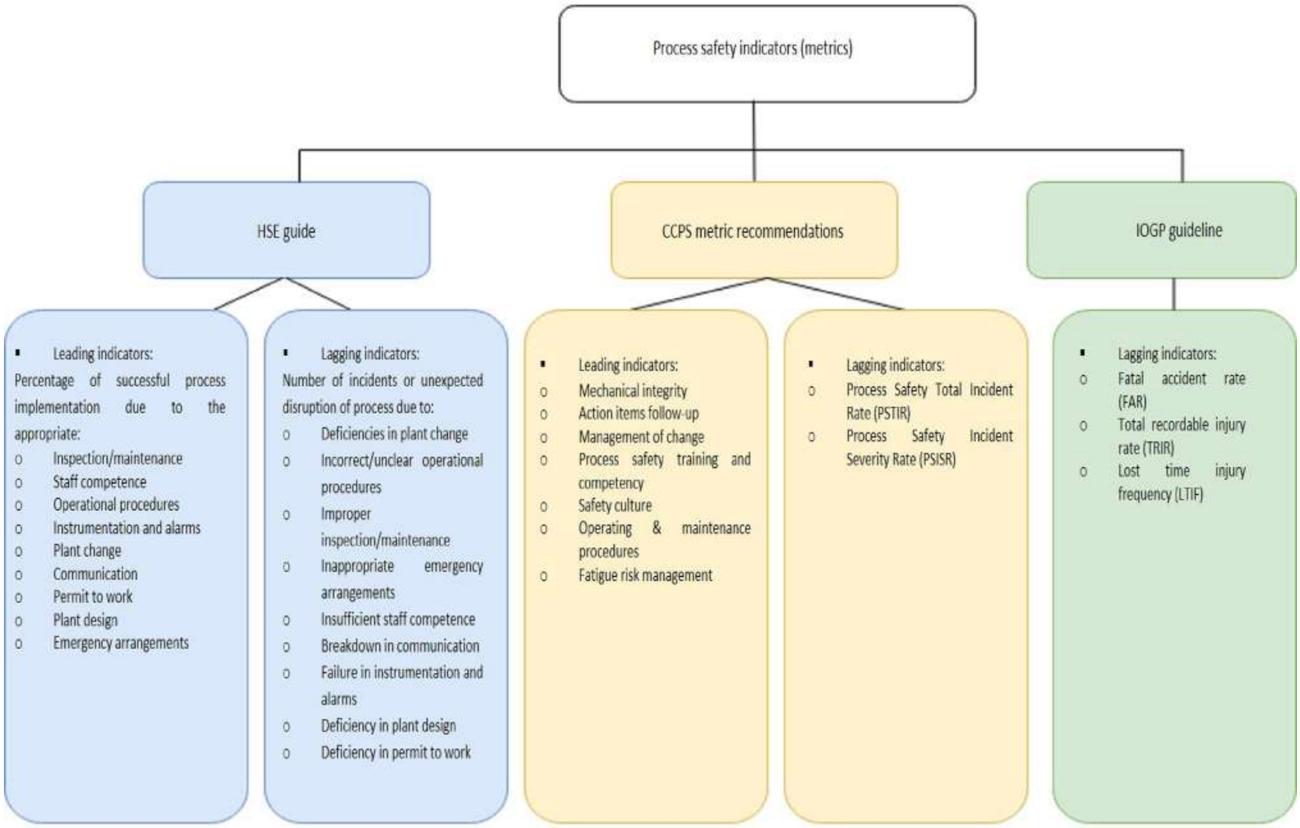


Fig. 2. Process safety indicators incorporated in the survey questionnaire.

either past or current work experience in the context of oil and gas industries in Iran and in Western countries. A questionnaire was developed to gather data in relation to each indicator. The questionnaire was sent by email to respondents. A total of 35 questionnaires were sent to Iranian industrial practitioners, and 32 questionnaires were returned (response rate = 91%). The questionnaire was sent to 23 Western industrial practitioners, and 13 questionnaires were returned (response rate = 56%). Fig. 3 shows the workflow of the approach used in the current study.

2.2.1. Determination of the importance of indicators

FBWM as a pairwise comparison-based method was applied to determine the fuzzy weights of indicators. The procedure of FBWM can be described in a series of steps (Guo & Zhao, 2017; Hafezalkotob & Hafezalkotob, 2017; Rezaei et al., 2017; Rezaei, 2015):

1. Determine the decision criteria system. In the first step, the criteria {C1, C2,…, Cn}that should be used for decision making are considered. In this work, these are process safety indicators.
2. Determine the best (B) and the worst (W) criteria. The best (most important) and the worst (least important) criteria are identified by decision-makers (respondents).
3. Execute the fuzzy preference comparisons for the best criterion. The fuzzy preference of the best criterion over all the other criteria is determined. The linguistic terms of preferences (Table 1) are used to determine the fuzzy preference of the most important (best) criterion over all the criteria. Then, the transformation of obtained fuzzy preference to triangular fuzzy numbers (TFNs) ($a_{Bj} = (a^L_{Bj}, a^M_{Bj}, a^U_{Bj})$) is done according to the transformation rules. The resulting fuzzy Best-to-Others vector would be:

$$A_B = (a_{B1}, a_{B2}, \dots, a_{Bn})$$

where A_B indicates the fuzzy Best-to-Others vector; a_{Bj} indicates the fuzzy preference of the best criterion c_B over criterion j, $j = 1, 2, \dots, n$. Since each criterion is equally important in comparison with itself then the fuzzy preference of the best criterion over itself would be $a_{BB} = (1, 1, 1)$.

4. Execute the fuzzy preference comparisons for the worst criterion. The fuzzy preferences of all the criteria over the worst criterion are extracted using the linguistic variables. The fuzzy preferences of all the criteria over the worst criterion are determined, and the obtained fuzzy preferences are transformed to TFNs ($a_{Bj} = (a^L_{jW}, a^M_{jW}, a^U_{jW})$) according to the transformation rules. The resulting fuzzy Others-to-Worst vector would be:

$$A_W = (a_{1W}, a_{2W}, \dots, a_{nW})$$

where A_W indicates the fuzzy Others-to-Worst vector; a_{iW} indicates the fuzzy preference of criterion i over the worst criterion c_W, $i = 1, 2, \dots, n$. Since in the comparison process the worst criterion is equally important in comparison with itself then the fuzzy preference of the worst criterion to itself is $a_{WW} = (1, 1, 1)$.

5. Find the optimal weights $(w^*_1, w^*_2, \dots, w^*_n)$. The optimal weight for the criterion j (w_j) is the one where for each fuzzy pair of w_B/w_j and w_j/w_W, we have $w_B/w_j = a_{Bj}$ and $w_j/w_W = a_{jW}$. Where w_B indicates the weight of the best criterion and w_j is the weight of the worst criterion. To satisfy these conditions for all j, a solution should be determined where the maximum absolute differences $\left|\frac{w_B}{w_j} - a_{Bj}\right|$ and $\left|\frac{w_j}{w_W} - a_{jW}\right|$ for all j is minimized. The optimization problem for determining the optimal fuzzy weights $\left(w^*_1, w^*_2, \dots, w^*_n\right)$ can be determined as follows.

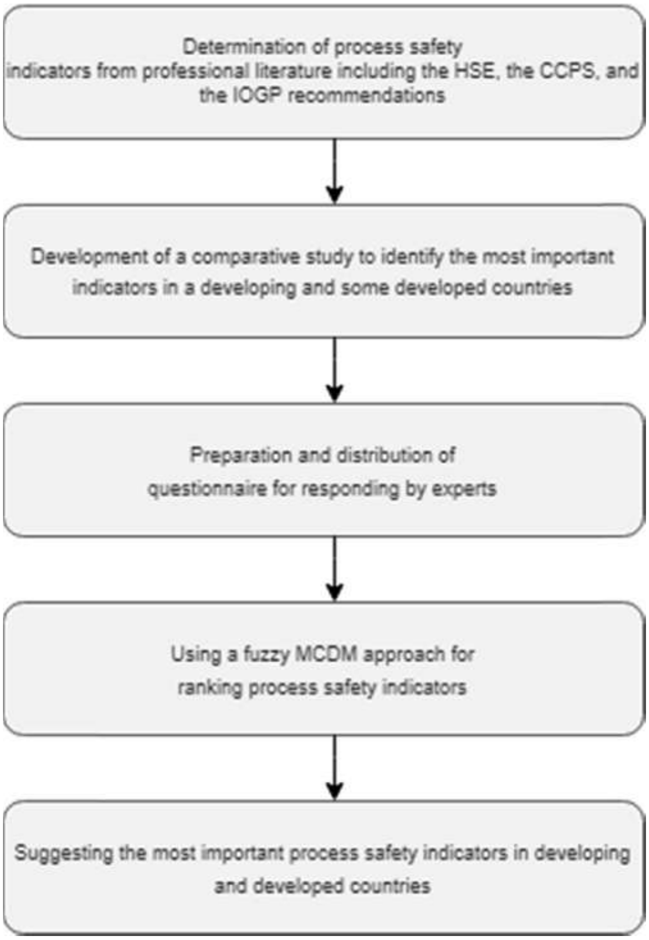


Fig. 3. The workflow of the study comparing the expert opinions regarding lagging and leading indicators.

Table 1
Linguistics variables for evaluating the factors.

Linguistics terms	Membership function
Equally important (EI)	(1,1,1)
Weakly important (WI)	(0.666,1,1.5)
Fairly important (FI)	(1.5,2,2.5)
Very important (VI)	(2.5,3,3.5)
Absolutely important (AI)	(3.5,4,4.5)

$$\min \quad \varepsilon,$$

$$s.t.$$

$$w_B - \varepsilon \lesssim a_{Bj} w_j, \quad \text{for all } j,$$

$$w_B + \varepsilon \gtrsim a_{Bj} w_j, \quad \text{for all } j,$$

$$w_j - \varepsilon \lesssim a_{jW} w_W, \quad \text{for all } j,$$

$$w_j + \varepsilon \gtrsim a_{jW} w_W, \quad \text{for all } j,$$

$$\sum_j w_j = 1,$$

$$w_j \geq 0, \quad \text{for all } j$$

In the above problem, the symbol $\lesssim$ refers to "almost lesser than" which is used to consider fuzzy values in the model. Transferring the fuzzy constraints to the crisp equivalents would lead to the following problem:

$$\min \quad \varepsilon,$$

$$s.t.$$

$$w_B - \varepsilon \leq \left[a_{Bj}^M + (1-\alpha)a_{Bj}^U\right] w_j \quad \text{for all } j,$$

$$w_B + \varepsilon \leq \left[a_{Bj}^M - (1-\alpha)a_{Bj}^L\right] w_j \quad \text{for all } j,$$

$$w_j - \varepsilon \leq \left[a_{jW}^M + (1-\alpha)a_{jW}^U\right] w_W \quad \text{for all } j,$$

$$w_j + \varepsilon \leq \left[a_{jW}^M - (1-\alpha)a_{jW}^L\right] w_W \quad \text{for all } j,$$

$$\sum_j w_j = 1,$$

$$w_j \geq 0, \quad \text{for all } j$$

where, $\alpha (0 \leq \alpha \leq 1)$ indicates a possibility level defined by the decision maker, while a_{Bj}^U, a_{Bj}^M and a_{Bj}^L respectively stand for upper bound, middle value, and lower bound of the triangular fuzzy number describing the fuzzy preference of the best criterion over criterion j. Similarly, a_{jW}^U, a_{jW}^M , a_{jW}^L represent the upper bound, middle value, and lower bound of the triangular fuzzy number describing the fuzzy preference of criterion j over the worst criterion. The optimal fuzzy weights $\left(w_1^*, w_2^*, \ldots, w_n^*\right)$, can be determined by solving the problem.

In addition, in the current study, a hierarchical structure was developed to determine the weight of each leading indicator. For this purpose, three criteria (levels or aspects) consisting of organizational, human, and technical were considered for leading indicators, each of which included sub-criteria (sub-aspects). Organizational criteria included mechanical integrity, action items follow-up, management of change (plant change), safety culture, operating & maintenance procedures (operational procedures), emergency arrangements, and inspection/maintenance. Human criteria included process safety training and competency, fatigue risk management, and communication. Technical criteria included instrumentation and alarms, plant design, and permit to work.

2.2.2. Actionability of the process safety indicators

The actionability (practicability) of each lagging/leading indicator was examined as well. For determining the actionability of each indicator, respondents were requested to determine the actionability of each study indicator based on the available information on the companies or publicly available databases of process industries in their countries. The respondents rated the actionability of each indicator on a five-point scale from very low to very high.

The possible values for actionability (practicability) were described based on the linguistic variables (terms) of decision-makers. The linguistic evaluations were transformed to fuzzy numbers (represented by TFNs). The process of fuzzification and defuzzification were applied to determine the actionability of each indicator in relation to applications in process industries and to compute the score for each indicator based on experts' evaluation. Table 2 presents the description of linguistic variables of actionability.

Table 2
Linguistics variables for actionability.

Linguistics terms	Membership function
Very low	(0,1,1.5)
Low	(0.5,1.5,2.5)
Moderate	(1.5,2.5,3.5)
High	(2.5,3.5,4.5)
Very high	(4.5,5,5)

Table 3
Linguistics variables for the perceived probability.

Linguistics terms	Membership function
Very low	(0,0,0.3)
Low	(0.1,0.3,0.5)
Moderate	(0.3,0.5,0.7)
High	(0.5,0.7,0.9)
Very high	(0.9,1,1)

2.2.3. The score of indicators

The safety score of each indicator was calculated from the perceived importance of the indicator, the perceived probability of incident occurrence due to failure to observe the indicators, and the perceived compliance status of the indicator (Tang et al., 2018b). The perceived importance of indicators was determined using FBWM. The respondents were asked to rate the perceived probability of incident occurrence due to failure to observe the indicators on a five-point scale. The higher the perceived rating of each indicator, the higher level of perceived probability. The perceived compliance status of each indicator was determined based on a numbering system adapted from the traffic light system proposed by the HSE in its Asset Integrity Key Program where red, amber, and green indicate non-compliance, partial compliance (the desired status has not been met), and compliance, respectively (HSE, 2008; Tang et al., 2018a). In the numbering system, "0" was assigned for indicators without data, and "1," "2," and "3" were assigned for non-compliance, partial compliance, and compliance, respectively. The comparison of indicators' performance by the respective performance targets or standards was applied to determine the compliance status.

The weight of each indicator (W_i) was calculated by multiplying the perceived importance of the indicator (I_i) with the perceived probability of incident occurrence due to failure to observe the indicator (P_i) and the safety score of an indicator (a) was obtained by multiplying the number assigned to the compliance level of an indicator (C_i) with the weight of the indicator (W_i), as follows:

$$W_i = I_i \times P_i$$

$$\text{Score of each indicator}, \quad a = W_i \times C_i$$

A higher score represents greater compliance with performance targets.

The possible values for each of the variables related to the perceived probability of incident occurrence due to failure to observe the indicators and the perceived compliance status were described based on the linguistic variables (terms) of decision-makers.

The linguistic evaluations were transformed into fuzzy numbers (represented by TFNs). The process of fuzzification and defuzzification were applied to compute the score for each indicator based on experts' evaluations. Table 3 and Table 4 show the descriptions of linguistic variables of perceived probability and compliance status specified by mathematical explanations (fuzzy membership function). In this work, the average method was applied for the defuzzification of fuzzy outputs.

2.2.4. Fuzzy risk assessment for leading indicators

For leading indicators, the perceived risk level was determined. The level of perceived risk was determined based on experts' judgment. Experts were safety practitioners from Iran and Western countries. Good risk understanding, adequate expertise, and subjective (knowledge-based) judgments about risk based on probabilities are required for risk assessment (Aven & Krohn, 2014; Aven et al., 2011). The comparison arises because perceptions of risk are different between countries (Keown, 1989) and levels of safety are different in the process sectors of Iran and the West.

Table 4
Linguistics variables for the compliance status.

Linguistics terms	Membership function
Without data "0"	(0,0,1.5)
Non-compliance "1"	(0.5,1.5,2.5)
Partial compliance "2"	(1.5,2.5,3.5)
Compliance "3"	(3.5,4,4)

The perceived risk value of the indicator (R_i) is the product of severity (S_i) and likelihood of occurrence (or probability) (P_i) as: $R_i = S_i \times P_i$ (Gul & Guneri, 2016). In the current study, the perceived risk of the indicator was calculated by multiplying the perceived severity of consequences (or outcomes) due to failure to observe the indicator with the perceived probability of incident occurrence due to failure to observe the indicator. Measurement of this perceived probability was done using a five-point scale from 1 = rare to 5 = almost certain. For determining the perceived severity of consequences (or outcomes) due to failure to observe the indicator, the respondents were requested to indicate the perceived severity on a five-point scale from 1 = insignificant to 5 = catastrophic. The acceptability level of the perceived risks was determined based on the risk assessment matrix provided by Gul and Guneri (2016) (Table 5).

In process risk analysis, due to the number of uncertainties, real situations are very often not crisp and deterministic. In these circumstances, a fuzzy logic system (FLS) can be employed (Markowski & Mannan, 2008) to develop a fuzzy risk assessment. This was used here because the categorization of probability and severity in a traditional approach is imprecise and vague and can lead to major uncertainties concerning the risk category.

The steps of FLS, used to assess the perceived risks, are as follows (Markowski & Mannan, 2008, 2009; Yen & Langari, 1999):

1. The fuzzifier transforms crisp inputs into fuzzy inputs. In the fuzzification process, the mapping of the linguistic variables of each risk matrix component including probability, severity, and risk into fuzzy sets is performed in order to activate rules. Input variables for developing fuzzy risk assessment and their domain in a number of fuzzy sets are shown in Table 6. Different forms of a membership function can be used based on the type of input and output variables.
2. Inference engine of the FLS maps input fuzzy sets into fuzzy output sets by a set of rules. It handles the way in which rules are combined. The set of rules for risk assessment is created based on the logic of the traditional risk matrix. IF $\bar{p}_n$ is probability AND $\bar{s}_m$ is severity of consequences THEN risk is $\bar{r}_z$. $\bar{p}_n$, $\bar{s}_m$, and $\bar{r}_z$ represent the fuzzy sets in relation to probability, severity, and risk in a universe of discourse, respectively. The set of 25 knowledge rules (e.g., **IF** Probability is "Possible" and Severity of Consequence is "Moderate" **THEN** Risk Category (Level) is "Intermediate Risk") was generated using the risk matrix consisting of 5 categories of probability, 5 categories of severity, and 5 categories of risk. The Mamdani fuzzy inference system was applied to convert the qualitative rules into quantitative

Table 5
The risk assessment matrix.

	Perceived probability				
Perceived severity	Rare (1)	Unlikely (2)	Possible (3)	Likely (4)	Almost certain (5)
Insignificant (1)	Insignificant risks	Acceptable risks	Acceptable risks	Acceptable risks	Acceptable risks
Minor (2)	Acceptable risks	Acceptable risks	Acceptable risks	Intermediate risks	Intermediate risks
Moderate (3)	Acceptable risks	Acceptable risks	Intermediate risks	Intermediate risks	Significant risks
Major (4)	Acceptable risks	Intermediate risks	Intermediate risks	Significant risks	Significant risks
Catastrophic (5)	Acceptable risks	Intermediate risks	Significant risks	Significant risks	Unacceptable risks

Table 6
Fuzzy sets for risk value in a comparison of expert opinions between Iran and Western countries.

Linguistic variables	Linguistic term (fuzzy set)	Descriptive range	Universe of discourse
Probability	Rare	$0 \leq L \leq 0.3$	$L \in (0, 1)$
	Unlikely	$0.1 \leq L \leq 0.5$	
	Possible	$0.3 \leq L \leq 0.7$	
	Likely	$0.5 \leq L \leq 0.9$	
	Almost certain	$0.7 \leq L \leq 1$	
Severity of consequences	Insignificant	$0 \leq S \leq 1.5$	$S \in (0, 5)$
	Minor	$0.5 \leq S \leq 2.5$	
	Moderate	$1.5 \leq S \leq 3.5$	
	Major	$2.5 \leq S \leq 4.5$	
	Catastrophic	$3.5 \leq S \leq 5$	
Risk category	Insignificant	$0 \leq R \leq 0.45$	$R \in (0, 5)$
	Acceptable	$0 \leq R \leq 1.75$	
	Intermediate	$0.25 \leq R \leq 3.15$	
	Significant	$1.05 \leq R \leq 5$	
	Intolerable	$2.45 \leq R \leq 5$	

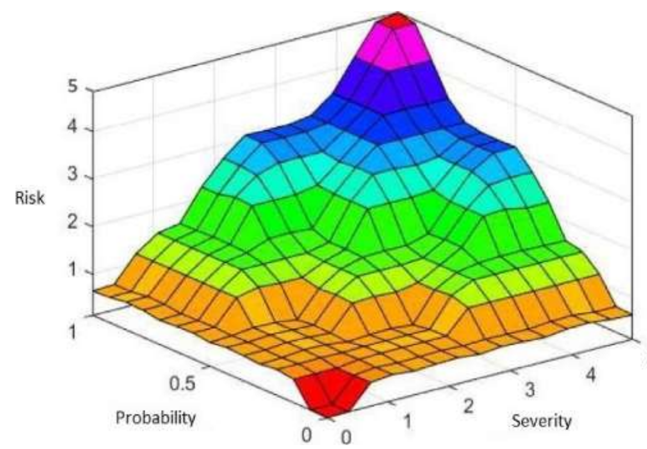


Fig. 4. Risk surface in the current study.

results (Mamdani & Assilian, 1975; Yen & Langari, 1999). After evaluating the rules, the aggregation of the output of different rules was performed. The aggregated output membership function is expressed as follows:

$$\mu_{R^-}(r) = \max_k \left\{ min\mu^k_{P^-}(p_n),\ \mu^k_s,\ \mu^k_{R^-}(r_Z) \right\}$$

where k, n, m, and z are the number of rules, the number of fuzzy probability sets, the number of fuzzy severity sets, and the number of fuzzy risk sets, respectively.

3. Defuzzification is the process of the conversion of the final fuzzy set into a crisp number. In the process, weighting and averaging the outputs from all of the individual fuzzy rules into a crisp numerical output value are carried out. There are various methods for defuzzification. In the current study, the center of area (COA) or the centroid method was used for defuzzification. The defuzzified output applying COA defuzzification method for the risk category (level) can be expressed by the following formula:

$$r_{crisp} = \frac{\int \mu_{\bar{R}}(r) r dr}{\int \mu_{\bar{R}}(r) dr}$$

where r is the output variable (risk category), r_{crisp} denotes the crisp quantity of the output variable and $\int \mu_{\bar{R}}(r)$ indicates the aggregated membership function.

The mapping from two input parameters (probability and severity) to one output parameter (risk) provides a basis from which the relationship between probability, severity, and risk can be illustrated by a three-dimensional plot (fuzzy risk surface). The risk surface (Fig. 4) was illustrated based on input parameters and different regions of risk (Markowski & Mannan, 2008).

3. Results

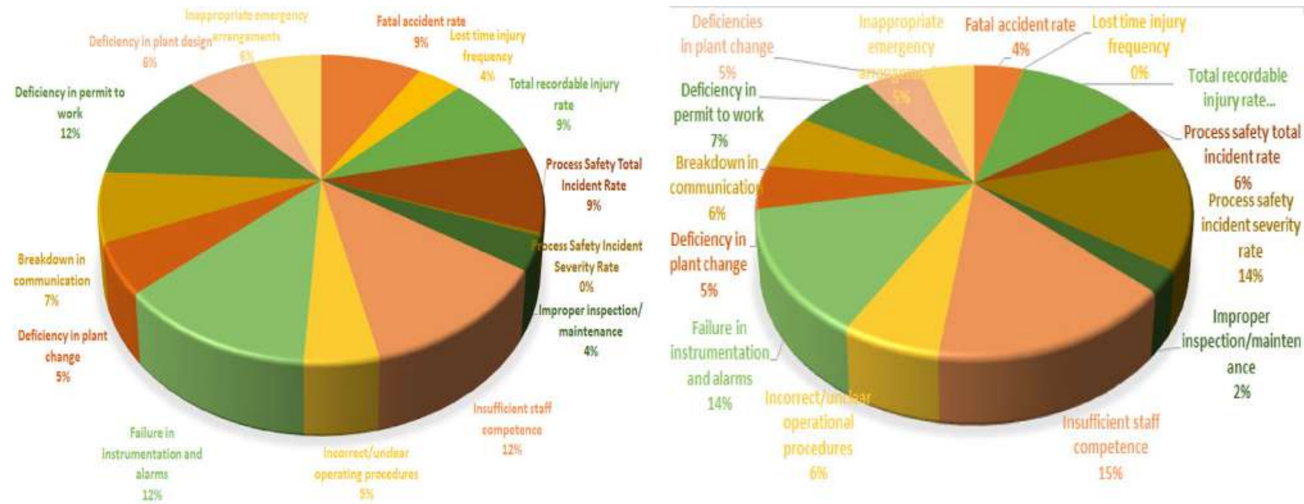
3.1. Lagging indicators

Lagging indicators that are based on incidents and events were defined based on the HSE guide, the CCPS recommendation, and the IOGP guideline. For each lagging indicator, the perceived importance, the actionability, and the score of the indicator from Iranian and Western experts' viewpoints were determined.

As can be seen from Fig. 5, based on the results obtained using FBWM, two important lagging indicators that were consistent between Iranian and Western experts were the failure in instrumentation and alarms and insufficient staff competence. Notably, Western experts identified PSISR as an important lagging indicator, whereas Iranian experts considered this to be the least important lagging indicator. All experts agreed that LTIF and the number of incidents or unexpected disruptions of process due to improper inspection/maintenance were the least important lagging indicators. Experts from the West also considered FAR to be less important.

In addition, deficiency in the permit to work and LTIF were some of the more important actionable lagging indicators in both contexts. LTIF was considered to be less important but actionable in both study contexts (Table 7). Experts from Iran also identified the number of times processes do not proceed as planned due to incorrect/unclear operational procedures and the number of unexpected disruption of process due to failure in instrumentation and alarms as the other important actionable lagging indicators. Those experts from the West noted FAR and inappropriate emergency arrangements as the other important actionable lagging indicators (Table 7).

In terms of the safety scores of lagging indicators, LTIF and PSTIR had low compliance with safety standards in the West and



(a) Iran (b) Western countries

Fig. 5. Perceived relative importance of the lagging indicators for process industries in Iran and Western countries.

Table 7
Actionability and safety scores for lagging indicators in Iran and the West.

Lagging indicator	Iran		Western countries	
	Actionability	Score of lagging indicator	Actionability	Score of lagging indicator
Fatal accident rate (FAR)	2.384	0.037	3.861	0.023
Lost time injury frequency (LTIF)	2.909	0.012	3.667	0.0002
Total recordable injury rate (TRIR)	2.586	0.030	3.472	0.049
Process safety total incident rate (PSTIR)	2.036	0.019	3.125	0.021
Process safety incident severity rate (PSISR)	1.552	0.0004	2.875	0.050
Number of incidents or unexpected disruption of process due to:				
Deficiencies in plant change	2.111	0.020	3.153	0.024
Incorrect/unclear operational procedures	2.788	0.023	3.208	0.030
Improper inspection/maintenance	2.596	0.035	3.639	0.013
Inappropriate emergency arrangements	2.545	0.031	3.681	0.027
Insufficient staff competence	2.313	0.045	3.139	0.064
Breakdown in communication	1.929	0.025	3.139	0.027
Failure in instrumentation and alarms	2.737	0.069	3.361	0.072
Deficiency in plant design	2.510	0.023	3.028	0.020
Deficiency in permit to work	3.152	0.079	3.486	0.040

Iran, respectively. Also, deficiencies in plant change and plant design were similarly rated by both sets of experts as lagging indicators that had low compliance, suggesting that they are universally important contributory factors in process safety incidents. For any lagging indicator, low safety scores indicate that incidents and disruptions associated with that indicator are more likely. Western experts gave improper inspection/maintenance a low score suggesting its importance as a cause of incidents in process industries (Table 7).

3.2. Leading indicators

The local and global weights of each leading indicator from Iranian and Western experts' viewpoints are shown in Table 8. In the current study, a hierarchical structure was developed to determine the weight of each leading indicator. Three criteria consisting of organizational, human, and technical were considered for leading indicators, each of which included sub-criteria. The optimal fuzzy weight was obtained for each leading indicator in the defined criteria.

In both contexts, organizational and human criteria had higher weights than technical criteria. Western experts weighted organizational and human criteria equally (weight = 0.400), whereas Iranian experts considered organizational criteria (weight = 0.424) as the most important criterion. Experts from Iran identified emergency arrangements and management of change (plant change) as the most important sub-criteria of the organizational criterion, while those in the West noted operational procedures and action items follow-up as the most important sub-criteria of the organizational criterion (Table 8).

The global weights of the sub-criteria were used to compare the actual weights of all sub-criteria. For this purpose, the relative weights were multiplied by the weights of the main criteria (Rezaei et al., 2015). In both settings, the appropriate process safety training and competency was the most important leading indicator. This was followed by instrumentation and alarms. Experts in Iran also viewed permit to work as an important leading indicator, while experts in the West focused on fatigue risk management. Overall the least important indicator appeared to be inspection/-maintenance. Some other indicators showed large variation between settings, for example plant design and action items

Table 8
Results of FBWM-weights of criteria and sub-criteria related to leading indicators.

Criteria	Iran	Western countries	Sub-criteria	Iran			Western countries		
	Criteria weights	Criteria weights		Sub-criteria weights	Global weights	Rank	Sub-criteria weights	Global weights	Rank
Organizational	0.424	0.400	Mechanical integrity	0.091	0.028	13	0.143	0.038	8
			Action items follow-up	0.111	0.034	12	0.179	0.048	6
			Management of change (plant change)	0.176	0.053	6	0.143	0.038	9
			Safety culture	0.158	0.048	8	0.143	0.038	10
			Operating & maintenance procedures (operational procedures)	0.148	0.045	9	0.179	0.048	7
			Inspection/maintenance	0.116	0.035	11	0.071	0.019	13
			Emergency arrangements	0.199	0.060	5	0.143	0.038	11
Human	0.294	0.400	Process safety training and competency	0.403	0.085	1	0.445	0.119	1
			Fatigue risk management	0.359	0.075	4	0.364	0.097	2
			Communication	0.238	0.050	7	0.182	0.049	5
Technical	0.282	0.200	Instrumentation and alarms	0.400	0.081	3	0.400	0.053	3
			Plant design	0.200	0.040	10	0.400	0.053	4
			Permit to work	0.400	0.081	2	0.200	0.027	12

Table 9
Actionability, safety scores, and risk values for leading indicators in Iran and the West.

Leading indicators	Iran			Western countries		
	Actionability	Score of leading indicator	Risk value	Actionability	Score of leading indicator	Risk value
Percentage of successful process implementation due to the appropriate:						
Mechanical integrity	1.879	0.018	2.960	3.056	0.084	3.700
Action items follow-up	1.626	0.052	2.930	3.056	0.087	3.170
Management of change (plant change)	2.273	0.082	2.930	3.194	0.070	3.190
Safety culture	1.586	0.028	2.880	2.083	0.056	3.040
Operating & maintenance procedures (operational procedures)	3.150	0.077	2.870	3.028	0.089	3.140
Inspection/maintenance	2.045	0.049	3.040	3.278	0.040	3.300
Emergency arrangements	2.636	0.115	3.110	3.444	0.088	3.480
Process safety training and competency	3.242	0.151	3.010	2.889	0.187	2.930
Fatigue risk management	2.056	0.030	2.860	2.083	0.132	3.840
Communication	2.141	0.037	2.900	2.556	0.076	2.930
Instrumentation and alarms	2.297	0.060	2.990	3.000	0.115	3.540
Plant design	2.364	0.058	3.030	2.500	0.092	3.320
Permit to work	2.893	0.156	3.100	3.139	0.059	3.270

follow-up were rated highly by experts in the West but not in Iran. Conversely, permit to work and emergency arrangements were rated highly by experts in Iran but not in the West.

Furthermore, as shown in Table 9, scores for actionability of leading indicators generally were greater in the reports of Western experts than those from Iran. Western experts identified emergency arrangements, inspection/maintenance, and management of change (plant change) as the three most actionable leading indicators, whereas Iranian experts considered process safety training and competency, operating and maintenance procedures (operational procedures), and permit to work as the three most actionable leading indicators. The least actionable indicator in both settings was safety culture because it is difficult to manage and manipulate.

In terms of the safety scores, while the leading indicator with the greatest weight by both sets of experts was process safety training and competence, the other most highly ranked indicators differed. These were permit to work and emergency arrangements for Iranian experts and fatigue risk management and instrumentation and alarm for Western experts. In terms of safety score, fatigue risk management obtained a relatively lower weight than other indicators in Iran compared with its relative score in the West, and permit to work obtained relatively lower weight in the West than other indicators compared with the situation in Iran (Table 9).

The perceived risk values for leading indicators were different between experts from Iran and the West (Table 9). In Iran, experts considered the three greatest risks associated with emergency arrangements, permit to work, and inspection/maintenance. In contrast, Western experts rated fatigue risk management, mechanical integrity, and instrumentation and alarms as the three greatest risks. Notably, the risk level related to fatigue risk management was perceived highest by the Western experts but lowest by those from Iran. Safety-related communication was not rated as a high risk in either setting, suggesting that this is well covered in practice.

4. Discussion

Safety indicators in process facilities are used as a predictive signal for major accidents. These indicators report the performance of the installation reflecting the effectiveness of the safety management system and differences in risk levels. Process safety indicators have been developed in different industries and at different time periods based on safety level and company goals (Swuste et al., 2016). In addition, the application of process safety indicators differs between countries, so a comparison of process safety indicators may show similarities or differences between developing and developed countries and thus may help to enhance the safety performance in process facilities of both sets of countries (Besserman & Mentzer, 2017; Swuste et al., 2016).

This study showed some similarities and some clear differences in the lagging indicators believed to be the more important ones by experts in Iran and the West. Failure in instrumentation and alarms and insufficient staff competence were important in both settings. Deficiencies in permit to work processes were considered important in Iran, whereas PSISR was considered to be important in the West. Failure in complying with permit to work processes is identified as a reason for some accidents such as HSE, 2005. Establishing an appropriate and effective permit to work system in process industries can help prevent and reduce process accidents (HSE, 2005; Jahangiri et al., 2016). In addition, process industries in Iran need to attend to the severity of process incidents (Soltanzadeh et al., 2019). The contributory effects of failure of work permit procedures in accidents, the importance of instrumentation and alarm systems in the safety analysis and in mitigating an abnormal state and major-accident conditions, and the effective role of training and competence on major accidents are reported in other studies (Do Koo et al., 2019; Hemmatian et al., 2014; Keown, 1989; Kim et al., 2019). The greater importance attached to PSISR in the West than in Iran perhaps suggests that there is a need for developing countries to attend to some specific process safety indicators and rate-based process safety metrics (such as PSISR) for measuring process safety performance and improving safety (CCPS, 2011).

With regards to leading indicators, this study shows that some leading indicators such as process safety training and competency, instrumentation and alarms, and fatigue risk management are important in both Iran and the West. The importance and the current status of process safety training and competency in the process industries is clearly critical and is considered an essential leading indicator (Sultana et al., 2019). Both operator fatigue and failed and insufficient instrumentation can lead to major accidents in the process industries (Knegtering & Pasman, 2009), so the proper functioning of instrumentation and alarms and the proper management of fatigue risk are considered important indicators for executing the processes safely and preventing process safety incidents. Experts in this study confirmed this. An important difference between the data obtained from experts in developing and developed countries was related to plant design. Plant design (compliance of safety critical items of plant with current design standards or codes) was identified as another important leading indicator in the West, whereas it had lower importance in Iran. Ensuring safety critical items of plant or equipment are compliant with the relevant standard is essential for the continued delivery of safe outcomes (HSE, 2006).

Perceptions of risk for leading indicators, as indicated by fuzzy risk assessments, were higher in the opinion of Western experts compared to those in Iran. This may be a function of the relatively greater age and experience of the respondents from the West compared with those from Iran. Experience of decision-making in critical operational situations could influence the expert's subjective judgments (Aven & Krohn, 2014). Past experience and the experience of negative safety outcomes can also influence the level of perceived risk and people's perception of hazards (Keller et al., 2006). In addition, the difference may also be a function of cultural background. Perception and evaluation processes are different between different societies having different cultural values and risk components, and this can affect individual's perception of risks (De Camprieu et al., 2007).

The higher values of perceived risks by experts in Western countries may lead to greater efforts to improve process safety, enhance compliance with safety rules and procedures, and may create a greater desire for participation in process safety-related issues. Fuzzy risk assessments for leading indicators revealed that emergency arrangements and permit to work were perceived to be the greatest risk by experts from Iran, whereas experts in the West considered fatigue risk management and mechanical integrity to be the greatest risks. Higher risk perceptions can result in more protective behavior (Xia et al., 2017).

In addition, FBWM used in the current work, as a recently developed method, gives managers and process safety practitioners in both developing and developed countries the opportunity to establish effective strategies for enhancing process safety by identifying the most influential factors and indicating where attention and effort should be placed. In comparison to existing MCDM approaches, FBWM needs less data and a full pairwise comparison matrix is not needed. The structured pairwise comparison system in the Best Worst Method produces more consistent results (Guo & Zhao, 2017; Salimi & Rezaei, 2018).

This study has a number of limitations. Only the experts' opinions and judgments about process safety indicators were considered and the actual data from specific facilities were not taken into account. Future work could compare site-specific information and process safety indicators in actual facilities from both developing and developed countries to show differences and similarities in the application of process safety indicators in actual facilities. Furthermore, the data in relation to developing countries were gathered only from Iran. Therefore, the representativeness of data is insufficient and the generalizability of the conclusions to other developing countries may be limited. Further studies in other developing countries can increase the generalizability of the results (e.g., future research might compare Southeast Asia or South America where risk perceptions differ).

4.1. Practical implications

The results suggest that some lagging indicators such as the number of incidents or unexpected disruption of process due to insufficient staff competence and failure in instrumentation and alarms are important from the perspectives of process safety experts of both developing and developed countries. So, continued attention needs to be given to these lagging indicators to prevent future incidents and adverse events.

In terms of leading indicators, the study has yielded some interesting results. Important leading indicators common to both contexts were safety training and competency, and instrumentation and alarms. Attention should continue to be given to these indicators irrespective of location. Experts in the two settings also identified other important leading indicators, but these differed. Experts in the West identified fatigue risk management, while those in Iran noted permit to work. One explanation for this difference might be in the evolution of indicators of process safety. As some indicators, evidently more proximal to the specific task or process, are routinely taken care of, others might become more salient. In this way permit to work precede fatigue risk management in the evolution of leading safety indicators in process industries.

Assigning different weights to different process safety indicators helps to identify the most important process safety indicators and to define a small and effective number of indicators for process facilities in both developing and developed countries. This gives opportunities for managers and safety professionals in process industries to have a good view of effective indicators and allows them to focus on more important ones (Salimi & Rezaei, 2018). Fuzzy Best Worst Method as the methodology used in the current study can help determine the weight and importance of process safety indicators. Identifying the most important process safety indicators is essential for organizations in developed and developing countries to define effective indicators to improve process safety performance, create a safer process industry, and prevent losses and process safety incidents.

5. Conclusion

Besserman and Mentzer (2017) pointed out that developing and developed countries occupy different stages in the application of process safety indicators and have areas of improvement in process safety that could help to enhance the safety performance in process facilities globally. In process industries, for improvement of process safety performance, the challenge is to define a small and effective number of process safety indicators (lagging and leading indicators). Developing a framework that differentiates the importance of process safety indicators based on the opinions of safety professionals helps to identify the most important process safety indicators. This can also be used to highlight the difference in perception between developing and developed regions and provides a basis to define an effective number of process safety indicators based on their importance (weight). This can lead to safety improvements in process facilities globally. FBWM was used to identify universally important lagging and leading indicators. In both settings, these are the number of times processes do not proceed as planned due to insufficient staff competence and failure in instrumentation and alarms (lagging indicators), and the percentage of successful process implementation due to appropriate process safety training and competency and instrumentation and alarms (leading indicators). This method has also shown differences in opinion between experts in Iran and the West. In terms of leading indicators, the most obvious of these are the percentage of successful process implementation due to plant design, action items follow-up, permit to work, and emergency arrangements. We suggest that these differences may be due to the experience and cultural background of the respondents, but also to the level of maturity/stage of evolution of the process industries in these countries, respectively.

Acknowledgments

The authors are very grateful to all experts who participated in this study.

Declaration of Competing Interest

The authors declare that they have no competing interests.

References

Amin, M. T., Khan, F., & Amyotte, P. (2019). A bibliometric review of process safety and risk analysis. *Process Safety and Environmental Protection, 126*, 366–381.
ANSI/API. (2010). *ANSI/API RP-754 Process Safety Performance Indicators for the Refining & Petrochemical Industries*. American Petroleum Institute.
ANSI/API. (2016). *ANSI/API RP-754 Process Safety Performance Indicators for the Refining & Petrochemical Industries*. American Petroleum Institute.
Antucheviciene, J., Kala, Z., Marzouk, M., & Vaidogas, E. R. (2015). Solving civil engineering problems by means of fuzzy and stochastic MCDM methods: current state and future research. *Mathematical Problems in Engineering, 2015*, 16. Article ID 362579.
Askarian, A., Jafari, M. J., Omidi, L., Miri Lavasani, M. R., Taghavi, L., & Ashori, A. (2018). Hazard identification and risk assessment in two gas refinery units. *Health Scope, 7*(1).
Aven, T., & Krohn, B. S. (2014). A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliability Engineering & System Safety, 121*, 1–10.
Aven, T., Renn, O., & Rosa, E. A. (2011). On the ontological status of the concept of risk. *Safety Science, 49*(8–9), 1074–1079.
Besserman, J., & Mentzer, R. A. (2017). Review of global process safety regulations: United States, European Union, United Kingdom, China, India. *Journal of Loss Prevention in the Process Industries, 50*, 165–183.
CCPS (2008). *Process safety leading and lagging metrics*. New York: Center for Chemical Process Safety.
CCPS (2011). *Process Safety Leading and Lagging Metrics*. New York: Center for Chemical Process Safety.
De Camprieu, R., Desbiens, J., & Feixue, Y. (2007). 'Cultural'differences in project risk perception: An empirical comparison of China and Canada. *International Journal of Project Management, 25*(7), 683–693.
Do Koo, Y., An, Y. J., Kim, C.-H., & Na, M. G. (2019). Nuclear reactor vessel water level prediction during severe accidents using deep neural networks. *Nuclear Engineering and Technology, 51*(3), 723–730.
Gul, M., & Guneri, A. F. (2016). A fuzzy multi criteria risk assessment based on decision matrix technique: A case study for aluminum industry. *Journal of Loss Prevention in the Process Industries, 40*, 89–100.
Guo, S., & Zhao, H. (2017). Fuzzy best-worst multi-criteria decision-making method and its applications. *Knowledge-Based Systems, 121*, 23–31.
Hafezalkotob, A., & Hafezalkotob, A. (2017). A novel approach for combination of individual and group decisions based on fuzzy best-worst method. *Applied Soft Computing, 59*, 316–325.
Hemmatian, B., Abdolhamidzadeh, B., Darbra, R., & Casal, J. (2014). The significance of domino effect in chemical accidents. *Journal of Loss Prevention in the Process Industries, 29*, 30–38.
HSE. (2005). *Guidance on permit-to-work systems systems: A guide for the petroleum, chemical and allied industries*. Health and Safety Executive.
HSE. (2006). *Developing process safety indicators, a step-by-step guide for chemical and major hazard industries, HSG254, ISBN 978 0 7176 6180 0*. http://www.hse.gov.uk/pubns/books/hsg254.htm.
HSE. (2008). *Key programme 3-Asset integrity programme*. Health and Safety Executive, United Kingdom.
IOGP. (2016a). *Process Safety-Leading Key Performance Indicators, Supplement to Report 456. Report No.556. International Oil and Gas Producers, UK*.
IOGP. (2016b). *Safety performance indicators: 2015 data*.
IOGP. (2019). *Safety performance indicators, 2019 data*.
Jahangiri, M., Hoboubi, N., Rostamabadi, A., Keshavarzi, S., & Hosseini, A. A. (2016). Human error analysis in a permit to work system: A case study in a chemical plant. *Safety and Health at Work, 7*(1), 6–11.
Keller, C., Siegrist, M., & Gutscher, H. (2006). The role of the affect and availability heuristics in risk communication. *Risk Analysis, 26*(3), 631–639.
Keown, C. F. (1989). Risk perceptions of Hong Kongese vs. Americans. *Risk Analysis, 9*(3), 401–405.
Khan, F., Abunada, H., John, D., & Benmosbah, T. (2010). Development of risk-based process safety indicators. *Process Safety Progress, 29*(2), 133–143.
Kim, T. K., Park, J. K., Lee, B. H., & Seong, S. H. (2019). Deep-learning-based alarm system for accident diagnosis and reactor state classification with probability value. *Annals of Nuclear Energy, 133*, 723–731.
Knegtering, B., & Pasman, H. (2009). Safety of the process industries in the 21st century: A changing need of process safety management for a changing industry. *Journal of Loss Prevention in the Process Industries, 22*(2), 162–168.
Louvar, J. (2010). Guidance for safety performance indicators. *Process Safety Progress, 29*(4), 387–388.
Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-machine Studies, 7*(1), 1–13.
Manzey, D., & Marold, J. (2009). Occupational accidents and safety: The challenge of globalization. *Safety Science, 47*(6), 723–793.
Markowski, A. S., & Mannan, M. S. (2008). Fuzzy risk matrix. *Journal of Hazardous Materials, 159*(1), 152–157.
Markowski, A. S., & Mannan, M. S. (2009). Fuzzy logic for piping risk assessment (pfLOPA). *Journal of Loss Prevention in the Process Industries, 22*(6), 921–927.
Moradi Hanifi, S., Omidi, L., & Moradi, G. (2019). Risk calculation and consequences simulation of natural gas leakage accident using ALOHA software. *Health and Safety at Work, 9*(1), 13–20.
OECD. (2008a). *Guidance on Developing Safety Performance Indicators Related to Chemical Accident Prevention, Preparedness and Response. OECD Environment, Health and Safety Publications, Series on Chemical Accidents No. 18 for Public Authorities and Communities/Public, Environment Directorate, Paris*.
OECD. (2008b). *Guidance on Developing Safety Performance Indicators Related to Chemical Accident Prevention, Preparedness and Response. OECD Environment, Health and Safety Publications, Series on Chemical Accidents No. 198 Guidance for Industry, Environment Directorate, Paris*.
Øien, K., Utne, I. B., & Herrera, I. A. (2011). Building safety indicators: Part 1–theoretical foundation. *Safety Science, 49*(2), 148–161.
Omidi, L., Salehi, V., Zakerian, S., & Nasl Saraji, J. (2021). Assessing the influence of safety climate-related factors on safety performance using an Integrated Entropy-TOPSIS Approach. *Journal of Industrial and Production Engineering*, 1–10.
Omidi, L., Zakerian, S. A., Nasl Saraji, J., Hadavandi, E., & Yekaninejad, M. S. (2018). Prioritization of Human Factors Variables in the Management of Major Accident Hazards in Process Industries Using Fuzzy AHP Approach. *Health Scope, 7*(4).
Omidi, L., Zakerian, S. A., Saraji, J. N., Hadavandi, E., & Yekaninejad, M. S. (2018). Safety performance assessment among control room operators based on feature extraction and genetic fuzzy system in the process industry. *Process Safety and Environmental Protection, 116*, 590–602.
OSHA (1992). *Process safety management of highly hazardous chemicals*. USA: Occupational Safety and Health Administration.
Pasman, H., & Rogers, W. (2014). How can we use the information provided by process safety performance indicators? Possibilities and limitations. *Journal of Loss Prevention in the Process Industries, 30*, 197–206.
Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega, 53*, 49–57.
Rezaei, J., Hemmes, A., & Tavasszy, L. (2017). Multi-criteria decision-making for complex bundling configurations in surface transportation of air freight. *Journal of Air Transport Management, 61*, 95–105.
Rezaei, J., Wang, J., & Tavasszy, L. (2015). Linking supplier development to supplier segmentation using Best Worst Method. *Expert Systems with Applications, 42*(23), 9152–9164.

Salimi, N., & Rezaei, J. (2018). Evaluating firms' R&D performance using best worst method. *Evaluation and Program Planning, 66*, 147–155.
Sinelnikov, S., Inouye, J., & Kerper, S. (2015). Using leading indicators to measure occupational health and safety performance. *Safety Science, 72*, 240–248.
Soltanzadeh, A., Heidari, H., Mohammadi, H., Mohammadbeigi, A., Sarsangi, V., & Darakhshan Jazari, M. (2019). Comprehensive causal analysis of occupational accidents' severity in the chemical industries; A field study based on feature selection and multiple linear regression techniques. *Journal of Health and safety at Work, 9*(4), 298–310.
Stough, J. (2012). *From Research to Practice: A Story of an Actionable Safety Leading Indicator Index*. International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production.
Stough, J. (2012). *Using Leading Indicators to Continuously Improve OHS & Sustainability Performance*.
Sultana, S., Andersen, B. S., & Haugen, S. (2019). Identifying safety indicators for safety performance measurement using a system engineering approach. *Process Safety and Environmental Protection, 128*, 107–120.
Swuste, P., Theunissen, J., Schmitz, P., Reniers, G., & Blokland, P. (2016). Process safety indicators, a review of literature. *Journal of Loss Prevention in the Process Industries, 40*, 162–173.
Tang, D. K. H., Dawal, S. Z. M., & Olugu, E. U. (2018a). Actual safety performance of the Malaysian offshore oil platforms: Correlations between the leading and lagging indicators. *Journal of Safety Research, 66*, 9–19.
Tang, D. K. H., Dawal, S. Z. M., & Olugu, E. U. (2018b). Integrating fuzzy expert system and scoring system for safety performance evaluation of offshore oil and gas platforms in Malaysia. *Journal of Loss Prevention in the Process Industries, 56*, 32–45.
Wang, T.-C., & Lee, H.-D. (2009). Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. *Expert Systems with Applications, 36*(5), 8980–8985.
Xia, N., Wang, X., Griffin, M. A., Wu, C., & Liu, B. (2017). Do we see how they perceive risk? An integrated analysis of risk perception and its effect on workplace safety behavior. *Accident Analysis & Prevention, 106*, 234–242.
Yen, J., & Langari, R. (1999). *Fuzzy logic: intelligence, control, and information* (Vol. 1) Saddle River, NJ: Prentice Hall Upper.
Zhen, X., Vinnem, J. E., & Næss, S. (2019). Building safety in the offshore petroleum industry: Development of risk-based major hazard risk indicators at a national level. *Process Safety and Environmental Protection, 128*, 295–306.

Leila Omidi is an assistant professor at the Department of Occupational Health Engineering, Tehran University of Medical Sciences, Iran. Her research focuses on process safety, safety behaviour, and human factors influencing safety.

Khadijeh Mostafaee Dolatabad is an assistant professor in the Faculty of Management and Economics, Tarbiat Modares University, Iran. Her research focuses on management science, decision making, and modeling operational risk.

Colin Pilbeam (PhD) is a Reader in Safety Leadership, based in the Cranfield Safety and Accident Investigation Centre at Cranfield University. His research focuses on how organizational structures, processes, and practices influence safe working and organizational safety performance.

Journal of Safety Research 84 (2023) 306–315

Contents lists available at ScienceDirect

Journal of Safety Research

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

Driver, roadway, and weather factors on severity of lane departure crashes in Maine

Alainie Sawtelle [a], Mohammadali Shirazi [a,*], Per Erik Garder [a], Jonathan Rubin [b]

[a] Department of Civil and Environmental Engineering, University of Maine, Orono, ME 04469, United States
[b] School of Economics, University of Maine, Orono, ME 04469, United States

ARTICLE INFO

Article history:
Received 24 January 2022
Received in revised form 3 September 2022
Accepted 9 November 2022
Available online 18 November 2022

Keywords:
Crash data
Crash severity
Lane departure
Statistical analysis
Multinomial logit

ABSTRACT

Introduction: In Maine, lane departure crashes account for over 70% of roadway fatalities. The majority of roadways in Maine are rural. Moreover, Maine has aging infrastructure, houses the oldest population in the United States, and experiences the third coldest weather in the United States. *Methods:* This study analyzes the impact of roadway, driver, and weather factors on the severity of single-vehicle lane departure crashes occurring from 2017 to 2019 on rural roadways in Maine. Rather than using police reported weather, weather station data were utilized. Four facility types: Interstates, minor arterials, major collectors, and minor collectors were considered for analysis. The Multinomial Logistic Regression model was used for the analysis. The property damage only (PDO) outcome was considered as the reference (or base) category. *Results:* The modeling results show that the odds of a crash leading to major injury or fatality (KA outcome) increases by 330%, 150%, 243%, and 266% for older drivers (65 or above) compared to young drivers (29 or less) on Interstates, minor arterials, major collectors, and minor collectors, respectively. During the winter period (October to April), the odds of KA severity outcome (with respect to the PDO) decreases by 65%, 65%, 65%, and 48% on Interstates, minor arterials, major collectors, and minor collectors, respectively, presumably due to reduced speeds during winter weather events. *Conclusion:* In Maine, factors such as older drivers, operating under the influence, speeding, precipitation, and not wearing a seatbelt showed higher odds of leading to injury. *Practical Applications*: This study provides safety analysts and practitioners in Maine a comprehensive study of factors that influence the severity of crashes in Maine at different facilities to improve maintenance strategies, enhance safety using proper safety countermeasures, or increase awareness across the state.

1. Introduction

Compared to other New England states, Maine has the highest roadway fatality rate (Bouchard et al., 2020). Lane departure (including run-off-road and head-on crashes) account for *more than 70%* of roadway fatalities in Maine. Maine is a rural state, with over 80% of the roadways being in rural areas. Maine has aging infrastructures, the oldest population in the United States, and experiences significant number of extreme weather events during the long winter season (often spanning from November to April). Maine is unique in many ways and therefore an interesting case study to better understand the impact of aging infrastructure, older population, and extreme weather conditions on severity of lane departure crashes.

In terms of infrastructure, the ASCE 2020 Annual Infrastructure Report Card gave Maine a C- grade (Bouchard et al., 2020). The report also gave roadways in Maine a D grade. The Annual Report suggests that the Maine highway system managed by the state has an annual gap in necessary funds of $135 million to make necessary roadway upgrades on aging infrastructure, proper maintenance, renovations, and improving safety. Maine also houses the oldest population in the United States (Himes & Kilduff, 2019). The population has been showing an aging trend since the 1990 census, where the median age was 33.9 years-old, and the U.S. median was 32.9 years-old (Meyer, 2001). In 2020, the median age was 45.0 in Maine, whereas the median age was 38.2 in the United States.

The state experiences lengthy winter seasons and around six months of winter precipitation, freezing temperatures, and several extreme storm events. In fact, the state is *the third coldest state in the U.S.* (World Population Review, 2021). The total precipitation and snowfall totals vary by location in the state. From 2017 to

* Corresponding author.
E-mail addresses: alainie.sawtelle@maine.edu (A. Sawtelle), shirazi@maine.edu (M. Shirazi), garder@maine.edu (P.E. Garder), rubinj@maine.edu (J. Rubin).

https://doi.org/10.1016/j.jsr.2022.11.006
0022-4375/

2019 (duration used in this study), coastal Maine received an average of 51.6 inches of precipitation and 101.9 inches of snowfall. Northern Maine received an average of 41.9 inches of precipitation and 138.0 inches of snowfall in the same period. Despite its relatively small size, due to the vast differences in terrain from the coastline to the western mountain region, weather conditions and temperatures vary substantially throughout the state.

Several studies have explored the impact of different demographic, weather, and roadway factors on crash severity. For instance, drivers older than 65 were found to have a 64–105% higher chance of involvement in a severe or fatal crash in New Mexico and California (Wu, Zhang, Zhu, Liu & Tarefder, 2016; Kim, Ulfarsson, Kim, & Shankar, 2013). The likelihood of being in a severe crash was found to be 38–43% lower for drivers younger than 25 in South Central States (Li et al., 2019). Female and older male drivers experience increased likelihood of severe injury in Indiana (Morgan & Mannering, 2011). Kim et al. (2013) found that male drivers are 107% more likely to be in a fatal crash; however, Li et al. (2019) found that male drivers are 6–17% less likely to be in fatal crashes. When not wearing a seatbelt, studies showed that the crash severity increases by 265–318% (Abdel-Aty, 2003; Li et al., 2019). Fatalities decrease by 60% when wearing a seatbelt (Kim et al., 2013). Speeding crashes are 105% more likely to result in fatalities (Kim et al., 2013). Operating under the influence increases the likelihood of severe and fatal crashes by 73–502% (Kim et al., 2013; Li et al., 2019).

Researchers also found an average increase of 9% in fatality rate during adverse weather conditions (Qiu & Nixon, 2008). Rain conditions decrease the crash severity in England (Edwards, 1998). Variables such as grade, curve, impaired driving, multiple lanes, and not using a seatbelt increases probability of crashes being severe in rain conditions (Li et al., 2019). Snow days in the contiguous 48 states decrease fatalities by 16% (Eisenberg & Warner, 2005). When road conditions were wet, the probability of severe crashes decreased by 40% (Li et al., 2019). A minimum visibility decrease of one unit leads to a 1% increase in the probability of non-injury crashes on freeways in China (Zhang, Wen, Yamamoto, & Zeng, 2021). It was found that as wind speed increased by one unit, there was a 0.9% decrease in severe and fatal crashes (Zhang et al., 2021). It was found that there is a 70% higher chance of serious injury on roadways that are dry and pavement temperature is above freezing in Iowa (Shaheed, Gkritza, Carriquiry, & Hallmark, 2016). When visibility was within six miles and surface condition was not dry, the probability of a severe crash decrease by 45% (Shaheed et al., 2016). Theofilatos (2017) concluded no correlation between adverse weather and severe crashes on urban arterials in Athens. Roadway grade was found to increase severe injury during rain by 50% (Li et al., 2019). A 1% increase in grade was found to increase severe and fatal crashes by 2.86% (Zhang et al., 2021). The crash severity also increases by a range of 20–80% on curves (Li et al., 2019).

Limited research has been done to explore contributing factors on lane departure crashes considering the combination of driver, roadway, and daily weather (rather than weather cited in crash reports) in areas similar to the state of Maine. As noted earlier, Maine is unique in many ways due to factors such as aging infrastructure, housing the oldest population in the United States, and experiencing adverse weather conditions; it is hypothesized that the combination of discussed factors contributes to the severity of lane departure crashes, and a disproportionate number of lane departure fatalities in Maine. This study uses the Multinomial Logistic Regression model to understand the impact of various roadway, driver, and weather factors on the severity of single-vehicle lane departure crashes that occurred in the 3-year period from 2017 to 2019. The analysis is divided based on four different facility type including: (1) principal arterials – Interstates (referred to as Interstates in this paper), (2) minor arterials, (3) major collectors, and (4) minor collectors. The results of this study provide a better understanding of contributing factors on severity of lane departure crashes on different roadway facilities leading to improved management, maintenance, and safety.

2. Data description

Maine is the only state in New England that is part of the Highway Safety Information System (HSIS). The state actively collects an abundance of reliable and useful data on highway safety, including roadway, crash, and traffic data. Such reliable data on rural roads are crucial for robust analysis. We gathered crash data and contributing factors collected in Maine and created a uniform dataset for each facility type. As discussed, four rural roadway facility types were considered for the analysis: Interstates, minor arterials, major collectors, and minor collectors. The range of speed limits for these facilities are 50–75 mph for Interstates, 25–55 mph for minor arterials, 25–55 mph for major collectors, and 25–50 mph for minor collectors. A total of 11,409 single-vehicle lane departure crashes were reported from 2017 to 2019 in Maine. The total crashes for Interstates, minor arterials, major collectors, and minor collectors are 2,190, 1,994, 4,940, and 2,285, respectively. It is important that these facilities are analyzed separately due to the design, safety conditions, and differences in maintenance strategies (as described above). Four injury severity categories were considered for analysis. Fatal-incapacitating injury crashes (KA), non-incapacitating injury (B), possible injury (C), and property damage only (PDO).

The contributing factors were classified in four major subcategories. First, the driver factors. This subcategory includes variables such as driver age and sex, as well as behavioral factors such as speeding, operating under the influence (OUI), and seatbelt usage. Over 15 driver variables were considered, and eventually-seven variables were included in the analysis. The second subcategory included crash variables, such as time of day, crash type, day of the week, or vehicle type. In total 20 variables in this category were considered and eventually-four variables were included in the analysis. The third subcategory included roadway characteristics, such as curve presence, posted speed limit, lane width, and more. Over 12 variables were considered, and eventually-three variables were included in the analysis. The fourth subcategory included weather variables, a total of seven weather variables were considered and eventually-four variables were selected in the analysis.

The weather data were extracted from the National Oceanic and Atmospheric Administration (NOAA) for the day of crash from 16 weather stations (NOAA National Centers for Environmental Information). As noted in previous studies, the number of weather stations are limited (Zhao, Wang, Liu, & Jackson, 2019). To allocate the weather variables to each crash record, Thiessen Polygons were created around the 16 weather stations using ArcGIS Pro (Environmental Systems Research Institute). Thiessen Polygons are polygons created around a point (in this case a weather station) so that each point within the polygon is closest to the respective weather station. Therefore, it was assumed that the weather station inside each polygon represents the weather in that area. The map of the polygons is presented in Fig. 1 Sawtelle et al., 2022. The southern and coastal region in the state have higher population, higher density of roadway network, and consequently more crashes; therefore, more weather stations were available and used. However, the northern, and western areas in the state do not have much population, roadway system, and consequently crashes (these regions are mainly forests and woods.) Hence, fewer weather stations exist in these areas and used in analysis.

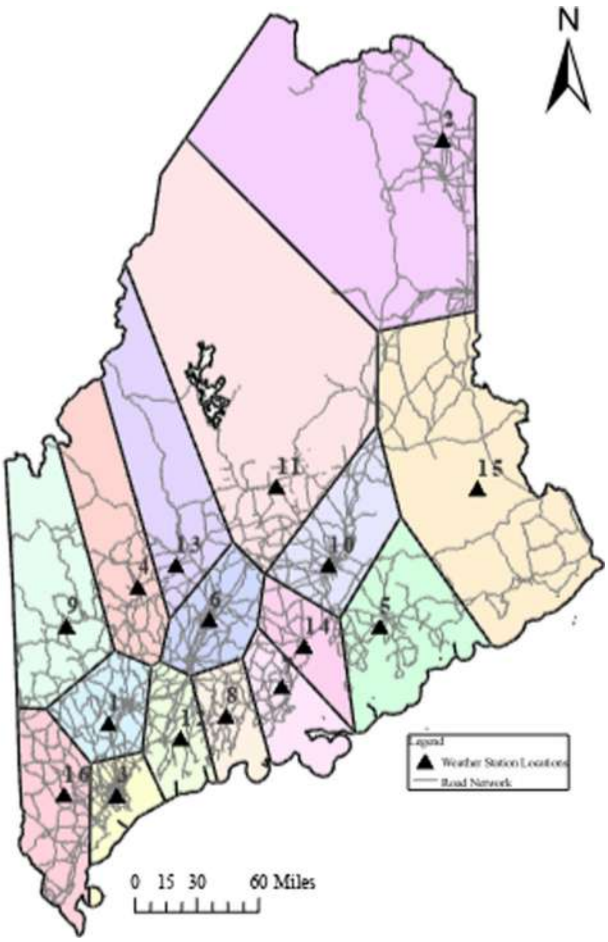


Fig. 1. Thiessen polygons and weather station locations.

As noted previously, many variables or combination of variables were considered, but not included in the final analysis (due to exploring correlation, significant test, and statistical fit). These variables include (but are not limited to) shoulder width, shoulder pavement type, lighting condition, the presence of rumble strips, vehicle type, hitting a fixed object, freezing temperatures, wind, and more. The categorical variables were also created based on extensive preliminary analyses. For example, for the driver age variable, it was found that designating "young" to drivers under the age of 30, "middle" to drivers from 30 and 64, and older to drivers of 65-years or above is the best representation of age category for this study. As another example, the variable "time of day" was divided into peak and off-peak time after extensive investigations. The peak time is between 6:00 AM-10:00 AM and 3:00PM-7:00PM Monday-Friday; the off-peak is otherwise. The speed limit variable differentiates between roadways with posted speed limits above 45mph on all facilities besides Interstates. The time between dawn and dusk was considered as the nighttime variable. The seasonal period variable represents the winter period from November to April and the non-winter period from May to October. The season variable accounts for factors during the winter that were not considered in the models. In this study, the surface conditions are considered as not dry if an officer noted the surface as wet, snow, slush, etc. and dry otherwise. This variable is not the same as weather variables as the surface condition may or may not necessarily be dry after storms. The variable snow day was used to describe if the area where a crash occurred experienced at least one inch of snow accumulation on the day of the crash. The variable precipitation describes if there was any precipitation accumulation on the day the crash occurred. Tables 1–4 show the summary of data used for the analysis for Interstates, minor arterials, major collectors, and minor collectors, respectively.

3. Methodology

Crash severity is identified as one of the following five categories: property damage only (PDO), possible injury (C-Injury), non-incapacitating injury (B-Injury), incapacitating injury (A-Injury), and fatal (K) crash. For the analysis, K and A crash outcomes were combined. To model crash severity, a Multinomial Logistics (MNL) model was used (Hilbe, 2011; Shankar & Mannering, 1996; Washington, Karlaftis, & Mannering, 2020; Shirazi, Geedipally, & Lord, 2017; Geedipally, Gates, Stapleton, Ingle, & Avelar, 2019; Zhao et al., 2021).

Similar to some of the previous studies (see, Geedipally at al., 2019), the MNL model was found to be a more appropriate model compared to the mixed logit for the data in hand. When using the MNL model, one category is designated as the reference category, and all other categories are compared to the reference; in this study, the PDO severity outcome was considered as the reference category. The probability of the i-th observation experiencing the j-th output injury is defined as follows:

$$p_{ij} = \frac{e^{U_{ij}}}{1 + \sum_j e^{U_{ij}}}$$

where, p_{ij} is the probability of the occurrence of crash severity "j" for observation "i", and U_{ij} is the deterministic part of the crash type likelihood. A linear function is used to link the crash severity with the various contributing factors as follows:

$$U_{ij} = \beta_{0j} + \sum_k \beta_{kj} X_{ik}$$

where β_{0j} is the constant term for j-th category, X_{ik} is the k-th variable for the i-th observation and β_{kj} is the coefficient for the k-th variable j-th crash type. The coefficients are estimated using the maximum likelihood approach. To interpret the results, the Odds Ratio (OR) were also estimated (Rahman, Sun, Das & Khanal, 2021; Holdridge, Shankar & Ulfrsson, 2005) and reported in the results section.

4. Results and discussion

A multinomial logit model was estimated for each facility type. As noted before, the PDO severity outcome was used as the reference (or base) category in each model. Therefore, the modeling results and the corresponding odds ratios discussed in this section are with respect to the PDO crash outcome. Tables 5–8 show the modeling results (e.g., the estimated coefficient of significant variables), and the corresponding odds ratios for Interstate, minor arterials, major collectors, and minor collectors, respectively. The tables also include the Akaike Information Criterion (AIC), Log-Likelihood, and McFadden's R^2 to analyze the goodness of fit (GOF).

4.1. Interstates

Table 5 shows the modeling results for rural Interstate roadways in Maine. As discussed, the driver age variable was classified into three groups (young, middle, and older). The young-driver category, indicating drivers with an age of 29 or less, was used as the reference (or base) group. The results show a positive correlation between the age of middle and older drivers and the Level B and Level KA severity outcomes. Given a crash, the odds of Level B and Level KA severity outcomes estimated with respect to the PDO increases by 36% and 79%, respectively, for middle aged dri-

Table 1
Count and frequency of variables for the Interstate facility.

Variables		PDO		C		B		KA	
		Count	Ratio	Count	Ratio	Count	Ratio	Count	Ratio
Driver Age	Young	679	31.0%	138	6.3%	103	4.7%	23	1.1%
	Middle	735	33.6%	153	7.0%	148	6.8%	41	1.9%
	Older	100	4.6%	28	1.3%	27	1.2%	15	0.7%
Male Driver Indicator	Male	1,024	46.8%	183	8.4%	176	8.0%	58	2.6%
	Not Male	490	22.4%	136	6.2%	102	4.7%	21	1.0%
Driver License	Suspended	27	1.2%	10	0.5%	12	0.5%	7	0.3%
	Active	1,487	67.9%	309	14.1%	266	12.1%	72	3.3%
Sobriety	OUI	43	2.0%	8	0.4%	17	0.8%	15	0.7%
	Not OUI	1,471	67.2%	311	14.2%	261	11.9%	64	2.9%
Distractions	Distracted	74	3.4%	24	1.1%	17	0.8%	8	0.4%
	Not Distracted	1,440	65.8%	295	13.5%	261	11.9%	71	3.2%
Driver Speed	Speeding	13	0.6%	3	0.1%	4	0.2%	3	0.1%
	Not Speeding	1,501	68.5%	316	14.4%	274	12.5%	76	3.5%
Seatbelt	Not Wearing	18	0.8%	21	1.0%	30	1.4%	22	1.0%
	Wearing	1,496	68.3%	298	13.6%	248	11.3%	57	2.6%
Crash Type	Rollover	23	1.1%	8	0.4%	15	0.7%	3	0.1%
	Not Rollover	1,491	68.1%	331	15.1%	263	12.0%	76	3.5%
Time of Day	Peak	648	29.6%	163	7.4%	109	5.0%	34	1.6%
	Not Peak	866	39.5%	156	7.1%	169	7.7%	45	2.1%
Night-time	Night	696	31.8%	127	5.80%	117	5.3%	33	1.5%
	Not Night	818	37.4%	192	8.77%	161	7.4%	46	2.1%
Curve	Present	323	14.7%	72	3.3%	63	2.9%	12	0.5%
	Not Present	1,191	54.4%	247	11.3%	215	9.8%	67	3.1%
Grade	Not Level	346	15.8%	97	4.4%	69	3.2%	18	0.8%
	Level	1,168	53.3%	222	10.1%	209	9.5%	61	2.8%
Season	Winter	1,103	50.4%	211	9.6%	166	7.6%	29	1.3%
	Not Winter	411	18.8%	108	4.9%	112	5.1%	50	2.3%
Surface Condition	Not Dry	1,084	49.5%	212	9.7%	163	7.4%	23	1.1%
	Dry	430	19.6%	107	4.9%	115	5.3%	56	2.6%
Snow	> 1 inch	182	8.3%	40	1.8%	20	0.9%	1	0.0%
	< 1 inch	1,332	60.8%	279	12.7%	258	11.8%	78	3.6%
Precipitation	Present	488	22.3%	105	4.8%	82	3.7%	18	0.8%
	Not Present	1,026	46.8%	214	9.8%	196	8.9%	61	2.8%

vers when compared with young drivers. The results show that the odds of Level B and Level KA severity outcomes estimated with respect to the PDO increases, respectively, by 70% and about 330% for older drivers when compared with young drivers. The modeling results show that the odds of a crash leading to a Level C or Level B severity outcome is, respectively, 38% and 29% smaller for male drivers. The results indicate that the odds of Level B and Level KA severity outcomes is 105% and 154% higher for drivers with a suspended driver license; these results are expected due to the risky behavior of these drivers. Speeding (driving above speed limit) often contributes to more severe crashes. The modeling results show that speeding increases the odds of Level KA severity outcome by 238%. The modeling results indicate that the odds of Level C severity outcome increases by 60% when the driver is distracted.

The modeling results shows a significant association between the severity of crashes and use of seatbelt. Given a crash, the odds of Level C severity outcome increases by over 5.5 times, Level B outcome by over 9.7 times, and Level KA outcome by over 27.4 times when a seat belt is not used. The odds of Level B and Level KA severity outcomes estimated with respect to the PDO increases by 330% and 229% respectively, when the vehicle rolls over. The modeling results show that crashes that occur during the peak hours have higher odds of resulting in Level C severity outcomes (about 38% more). The interaction of nighttime and operating under the influence was a significant variable for Level KA severity outcome. The odds of a crash resulting in a Level KA severity outcome is 232% higher when a driver is operating under the influence in the nighttime (between dawn and dusk). The odds of resulting in Level C injury outcome increases by 59% when the roadway is not level, likely due to reduced visibility.

Given a crash, the odds of Level C, Level B, and Level KA severity outcomes (with respect to the PDO), respectively, decreases by 27%, 38%, and 65% during the winter period (November-April). These results are expected because in the winter Interstates experience over 2.5 times more PDO crashes. Despite the significant increase in PDO crashes, the number of severe crashes remains more or less the same. In other words, although the inclement weather causes more PDO crashes, it does not increase the severity of crashes, due to presumably more cautious driving behavior under bad weather conditions. Given a crash, the odds of Level B and Level KA severity outcomes estimated with respect to the PDO decreases by 27% and 67%, respectively, when the surface is not dry. This observation is likely due to cautious driving behavior.

4.2. Minor arterials

Table 6 shows the modeling results for rural minor arterial roadways. The modeling results show that, given a crash, the odds of Level B and Level KA severity outcomes is, respectively, 140% and 150% higher for older drivers when compared with young drivers. In addition, given a crash, the odds of Level C and Level B crash outcomes are about 30% smaller for male drivers compared to female drivers. As discussed, drivers with suspended licenses are expected to be involved in more severe crashes due to their risky behavior. This observation was reflected in modeling results for minor arterials as well. The odds of Level C, Level B, and Level KA severity outcomes estimated with respect to the PDO, respectively, increases by 64%, 170%, and 287% for drivers with suspended license. The modeling results also show that the odds of Level C severity outcome increases by 42% when the driver is under the influence. Not wearing a seatbelt has the largest impact on

Table 2
Count and frequency of variables for the minor arterial facility.

Variables		PDO		C		B		KA	
		Count	Ratio	Count	Ratio	Count	Ratio	Count	Ratio
Driver Age	Young	524	26.3%	164	8.2%	77	3.9%	24	1.2%
	Middle	652	32.7%	209	10.5%	88	4.4%	43	2.2%
	Older	124	6.2%	42	2.1%	38	1.9%	9	0.5%
Male Driver Indicator	Male	851	42.7%	250	12.5%	124	6.2%	53	2.7%
	Not Male	449	22.5%	165	8.3%	79	4.0%	23	1.2%
Driver License	Suspended	38	1.9%	20	1.0%	15	0.8%	8	0.4%
	Active	1,262	63.3%	395	19.8%	188	9.4%	68	3.4%
Sobriety	OUI	98	4.9%	55	2.8%	24	1.2%	20	1.0%
	Not OUI	1,202	60.3%	360	18.1%	179	9.0%	56	2.8%
Distractions	Distracted	130	6.5%	45	2.3%	24	1.2%	7	0.4%
	Not Distracted	1,170	58.7%	370	18.6%	179	9.0%	69	3.5%
Driver Speed	Speeding	20	1.0%	6	0.3%	1	0.1%	6	0.3%
	Not Speeding	1,280	64.2%	409	20.5%	202	10.1%	70	3.5%
Seatbelt	Not Wearing	49	2.5%	48	2.4%	35	1.8%	42	2.1%
	Wearing	1,251	62.7%	367	18.4%	168	8.4%	34	1.7%
Crash Type	Rollover	32	1.6%	19	1.0%	10	0.5%	5	0.3%
	Not Rollover	1,268	63.6%	396	19.9%	193	9.7%	71	3.6%
Time of Day	Peak	580	29.1%	162	8.1%	90	4.5%	36	1.8%
	Not Peak	720	36.1%	253	12.7%	113	5.7%	40	2.0%
Night-time	Night	581	29.14%	179	8.98%	84	4.21%	35	1.76%
	Not Night	719	36.06%	236	11.84%	119	5.97%	41	2.06%
Speed Limit	> 45mph	1,099	55.1%	365	18.3%	165	8.3%	64	3.2%
	< 45mph	201	10.1%	50	2.5%	38	1.9%	12	0.6%
Curve	Present	608	30.5%	192	9.6%	110	5.5%	39	2.0%
	Not Present	693	34.8%	223	11.2%	93	4.7%	37	1.9%
Grade	Not Level	469	23.5%	141	7.1%	76	3.8%	19	1.0%
	Level	831	41.7%	274	13.7%	127	6.4%	57	2.9%
Season	Winter	953	47.8%	232	11.6%	94	4.7%	26	1.3%
	Not Winter	347	17.4%	183	9.2%	109	5.5%	50	2.5%
Surface Condition	Not Dry	773	38.8%	184	9.2%	72	3.6%	15	0.8%
	Dry	527	26.4%	231	11.6%	131	6.6%	61	3.1%
Snow	> 1 inch	131	6.6%	14	0.7%	6	0.3%	2	0.1%
	< 1 inch	1,169	58.6%	401	20.1%	197	9.9%	74	3.7%
Precipitation	Present	348	17.5%	88	4.4%	41	2.1%	13	0.7%
	Not Present	952	47.7%	327	16.4%	162	8.1%	63	3.2%

severity of crashes for minor arterials as well. Failing to wear a seatbelt increases the odds of Level C, Level B, or Level KA severity outcomes by 1.9-, 3.8-, and 23.1-times, respectively. Crash severity increases when a rollover crash occurs. Given a crash, vehicle rollover increases the odds of Level C, Level B, and Level KA severity outcomes (with respect to the PDO) by 139%, 169%, and 273%, respectively. For road segments with a posted speed limit of greater than 45mph, the odds of a crash resulting in Level C severity outcome increases by 46%. When a crash occurs on a curved segment, the odds of Level B severity outcome increases by 29%.

For minor arterials, the PDO crashes increase during the winter period by about 2.7-times; however, severe crashes (KA, B, and C outcomes) do not increase in proportion to PDOs. This observation was reflected in modeling results as well. During the winter period, the odds of Level C, Level B, and Level KA severity outcomes (with respect to the PDO) decreases by 45%, 54%, and 65%, respectively. On roadways with surface conditions that are described as "not dry," the odds of Level B and Level KA severity outcomes decreases by 31% and 63%, respectively. For minor arterials, the odds of Level C and Level B severity outcomes decreases by 27% and 49%, respectively, during the days with at least one inch of snowfall. These results are expected because, often, during snow days more PDO (due to inclement weather) but less severe (due to cautious driving behavior) crashes are expected.

4.3. Major collectors

Table 7 shows the modeling results for rural major collector roadways. Given a crash, for middle-aged drivers, the modeling results show increased odds of 47% in Level KA severity outcome compared to younger drivers. Likewise, for older drivers, the odds of Level C, Level B, and Level KA crash outcomes increases by 91%, 39%, and 243%, respectively, compared to young drivers. The results show that, given a crash, the odds of Level C and Level KA severity outcomes is, respectively, 38% and 29% smaller for male drivers compared to female drivers. When drivers are under the influence of drugs or alcohol, it is expected that they are involved in more severe crashes due to more reckless or aggressive driving behavior. The estimated model shows the same expectation. When operating under the influence, the odds of Level C, Level B, and Level KA severity outcomes (with respect to the PDO) increases by 45%, 74%, and 131%, respectively. In addition, the odds of crashes resulting in Level C and Level KA severity outcomes increases by 100% and 390%, respectively, when it is both nighttime and the driver is speeding.

Like Interstates and minor arterials, there is a significant association between injury/fatality outcomes (KA, B, and C outcomes) and not wearing a seatbelt. When a seatbelt is not used, the odds of Level C, Level B, and Level KA severity outcomes estimated with respect to the PDO increases by 1.8-, 3.6-, and 22.0-times, respectively. The vehicle rollover increases the odds of Level C, Level B, and Level KA severity outcomes by 140%, 183%, and 289%, respectively. The odds of Level B severity outcome decreases by 18% during the peak hour, likely because of congestion and speed reduction during peak hours. The odds of a crash leading to Level C, Level B, and Level KA crash severity outcomes increases by 23%, 23%, and 125%, respectively, on roads with speed limit of 45mph or above. This observation is expected, as the vehicle speed

Table 3
Count and frequency of variables for the major collector facility.

Variables		PDO		C		B		KA	
		Count	Ratio	Count	Ratio	Count	Ratio	Count	Ratio
Driver Age	Young	1,469	29.7%	436	8.8%	237	4.8%	78	1.6%
	Middle	1,448	29.3%	461	9.3%	247	5.0%	108	2.2%
	Older	241	4.9%	131	2.7%	51	1.0%	33	0.7%
Male Driver Indicator	Male	1,994	40.4%	572	11.6%	361	7.3%	151	3.1%
	Not Male	1,164	23.6%	456	9.2%	174	3.5%	68	1.4%
Driver License	Suspended	108	2.2%	46	0.9%	36	0.7%	20	0.4%
	Active	3,050	61.7%	982	19.9%	499	10.1%	199	4.0%
Sobriety	OUI	210	4.3%	117	2.4%	95	1.9%	62	1.3%
	Not OUI	2,948	59.7%	911	18.4%	440	8.9%	157	3.2%
Distractions	Distracted	244	4.9%	116	2.3%	58	1.2%	16	0.3%
	Not Distracted	2,914	59.0%	912	18.5%	477	9.7%	203	4.1%
Driver Speed	Speeding	56	1.1%	31	0.6%	20	0.4%	32	0.6%
	Not Speeding	3,102	62.8%	997	20.2%	515	10.4%	187	3.8%
Seatbelt	Not Wearing	132	2.7%	117	2.4%	107	2.2%	123	2.5%
	Wearing	3,026	61.3%	911	18.4%	428	8.7%	96	1.9%
Crash Type	Rollover	78	1.6%	55	1.1%	32	0.6%	14	0.3%
	Not Rollover	3,080	62.3%	973	19.7%	503	10.2%	205	4.1%
Time of Day	Peak	1,454	29.4%	427	8.6%	197	4.0%	79	1.6%
	Not Peak	1,704	34.5%	601	12.2%	338	6.8%	140	2.8%
Night-time	Night	1378	27.89%	447	9.05%	248	5.02%	84	1.70%
	Not Night	1780	36.03%	581	11.76%	287	5.81%	135	2.73%
Speed Limit	> 45mph	2,486	50.3%	834	16.9%	429	8.7%	189	3.8%
	< 45mph	672	13.6%	194	3.9%	106	2.1%	30	0.6%
Curve	Present	1,635	33.1%	520	10.5%	306	6.2%	132	2.7%
	Not Present	1,523	30.8%	508	10.3%	229	4.6%	87	1.8%
Grade	Not Level	1,315	26.6%	418	8.5%	226	4.6%	86	1.7%
	Level	1,843	37.3%	610	12.3%	309	6.3%	133	2.7%
Season	Winter	2,339	47.3%	594	12.0%	293	5.9%	77	1.6%
	Not Winter	819	16.6%	424	8.6%	242	4.9%	142	2.9%
Surface Condition	Not Dry	2,067	41.8%	532	10.8%	221	4.5%	64	1.3%
	Dry	1,091	22.1%	496	10.0%	314	6.4%	155	3.1%
Snow	> 1 inch	332	6.7%	63	1.3%	16	0.3%	2	0.0%
	< 1 inch	2,826	57.2%	965	19.5%	519	10.5%	217	4.4%
Precipitation	Present	2,339	47.3%	594	12.0%	293	5.9%	77	1.6%
	Not Present	819	16.6%	424	8.6%	242	4.9%	142	2.9%

is a major contributing factor to severity of crashes. When crashes occur on curves, the odds of Level B or Level KA severity outcomes increases by 23% and 37%, respectively.

During the winter period, major collectors experience 2.9-times more PDO crashes than the non-winter period. However, the severe crash outcomes do not increase in proportion to the PDOs. The odds of Level C, Level B, and Level KA severity outcomes estimated with respect to the PDO decreases by 43%, 29%, and 65%, respectively, during the winter period. The odds of Level B severity outcome decreases by 38% when the surface is not dry. The severity of crashes decreases on days with at least one inch of snow accumulation as well. During inclement weather, especially winter conditions, drivers slow down due to slippery conditions and lower visibility; therefore, the negative correlation with severe crashes is expected. During snow days with more than 1 inch of snow, the odds of Level C, Level B, and Level KA severity outcomes estimated with respect to the PDO decreases by 20%, 58%, and 78%, respectively. Precipitation increases the odds of level B-level crash severities by 21% compared to days without precipitation.

4.4. Minor collectors

Table 8 shows the modeling results for rural minor collector roadways. Given a crash, the results show increased odds of 58% in Level KA severity outcomes for middle-aged drivers compared to young drivers. Likewise, the odds of level B and level KA crash severity outcomes is, respectively, 68% and 266% higher for older drivers compared to the younger drivers. The results show that, given a crash, the odds of Level C and Level B severity outcomes decreases by 48% and 22%, respectively, for male drivers compared to female drivers. The "speeding" variable was found to be significant for Level C, Level B, and Level KA severity outcomes for minor arterials. These results are expected as speeding may result in losing control of the vehicle; higher speeds also result in more severe impact. The modeling results show that the odds of Level C, Level B, and Level KA severity outcomes increases by 58%, 123%, and 148%, respectively, when drivers are speeding (drive above speed limit).

Like previous facilities, not wearing a seat belt is the most influential factor in severity of crashes. The odds of a crash leading to Level C, Level B, and Level KA severity outcomes (with respect to the PDO) increases by 3.1-, 4-, and 13.3-times when the seatbelt is not used. The odds of Level C severity outcome estimated with respect to the PDO increases by 78% when the vehicle rolls over. The modeling results show that, given a crash, the odds of Level B and Level KA severity outcomes increases by 164% and 162%, respectively, when it is nighttime, and the driver operates under the influence. The results show that the odds of Level B and Level KA severity outcomes increases by 46% and 153%, respectively, when the speed limit is 45mph or greater. The odds of a crash leading to a Level KA severity outcome increases by 88% on curved segments. Likewise, the odds a crash leading to a Level C severity outcome increases by 28% when the roadway segment is not level.

During the winter period, minor collectors experience 3.1-times more PDO crashes than the non-winter season. However, the number of severe crashes remains relatively same. For minor collectors, the modeling results indicate that during the winter period, the odds of Level B and Level KA severity outcomes estimated with respect to the PDO decreases by 44% and 48%, respectively. Likewise, the odds of Level C, Level B, and Level KA severity outcomes is decreased by 32%, 38%, and 46%, respectively, when the surface is

Table 4
Count and frequency of variables for the minor collector facility.

Variables		PDO		C		B		KA	
		Count	Ratio	Count	Ratio	Count	Ratio	Count	Ratio
Driver Age	Young	762	33.3%	210	9.2%	109	4.8%	35	1.5%
	Middle	662	29.0%	188	8.2%	102	4.5%	42	1.8%
	Older	111	4.9%	32	1.4%	21	0.9%	11	0.5%
Male Driver Indicator	Male	949	41.5%	208	9.1%	141	6.2%	57	2.5%
	Not Male	586	25.6%	222	9.7%	91	4.0%	31	1.4%
Driver License	Suspended	48	2.1%	22	1.0%	11	0.5%	8	0.4%
	Active	1,487	65.1%	408	17.9%	221	9.7%	80	3.5%
Sobriety	OUI	84	3.7%	41	1.8%	45	2.0%	21	0.9%
	Not OUI	1,451	63.5%	389	17.0%	187	8.2%	67	2.9%
Distractions	Distracted	121	5.3%	40	1.8%	29	1.3%	7	0.3%
	Not Distracted	1,414	61.9%	390	17.1%	203	8.9%	81	3.5%
Driver Speed	Speeding	41	1.8%	25	1.1%	21	0.9%	11	0.5%
	Not Speeding	1,494	65.4%	405	17.7%	211	9.2%	77	3.4%
Seatbelt	Not Wearing	52	2.3%	59	2.6%	43	1.9%	37	1.6%
	Wearing	1,483	64.9%	371	16.2%	189	8.3%	51	2.2%
Crash Type	Rollover	45	2.0%	24	1.1%	12	0.5%	6	0.3%
	Not Rollover	1,490	65.2%	406	17.8%	220	9.6%	82	3.6%
Time of Day	Peak	705	30.9%	198	8.7%	103	4.5%	34	1.5%
	Not Peak	830	36.3%	232	10.2%	129	5.6%	54	2.4%
Nighttime	Night	631	27.61%	183	8.01%	88	3.85%	35	1.53%
	Not Night	904	39.56%	247	10.81%	144	6.30%	53	2.32%
Speed Limit	> 45mph	1,069	46.8%	313	13.7%	174	7.6%	72	3.2%
	< 45mph	466	20.4%	117	5.1%	58	2.5%	16	0.7%
Curve	Present	870	38.1%	248	10.9%	128	5.6%	64	2.8%
	Not Present	665	29.1%	182	8.0%	104	4.6%	24	1.1%
Grade	Not Level	673	29.5%	209	9.1%	98	4.3%	37	1.6%
	Level	862	37.7%	221	9.7%	134	5.9%	51	2.2%
Season	Winter	1,161	50.8%	291	12.7%	127	5.6%	35	1.5%
	Not Winter	374	16.4%	139	6.1%	105	4.6%	53	2.3%
Surface Condition	Not Dry	1,049	45.9%	239	10.5%	108	4.7%	27	1.2%
	Dry	486	21.3%	191	8.4%	124	5.4%	61	2.7%
Snow	> 1 inch	174	7.6%	23	1.0%	21	0.9%	1	0.0%
	< 1 inch	1,361	59.6%	407	17.8%	211	9.2%	87	3.8%
Precipitation	Present	407	17.8%	81	3.5%	52	2.3%	12	0.5%
	Not Present	1,128	49.4%	349	15.3%	180	7.9%	76	3.3%

Table 5
Modeling results for Interstate.

Variables		Estimate (Std.)			Odds Ratio		
		C	B	KA	C	B	KA
Intercept		−1.333 (0.194)	−1.441 (0.188)	−2.924 (0.368)	-	-	-
Driver Age	Middle	–[b]	0.308 (0.144)	0.583 (0.288)	-	1.361	1.791
	Older	-	0.528 (0.251)	1.458 (0.386)	-	1.696	4.296
Male Driver Indicator	Male	−0.479 (0.130)	−0.344 (0.142)	-	0.620	0.709	-
Driver License	Suspended	-	0.719 (0.376)[a]	0.931 (0.524) [a]	-	2.050	2.536
Driver Speed	Speeding	-	-	1.218 (0.720) [a]	-	-	3.380
Distractions	Distracted	0.471 (0.257) [a]	-	-	1.602	-	-
Seatbelt	Not Wearing	1.874 (0.334)	2.369 (0.313)	3.346 (0.381)	6.514	10.691	28.381
Crash Type	Rollover	-	1.459 (0.347)	1.192 (0.677) [a]	-	4.303	3.293
Time of Day	Peak	0.319 (0.127)	-	-	1.376	-	-
Nighttime and OUI	Yes	-	-	1.199 (0.451)	-	-	3.317
Grade	Not Level	0.464 (0.139)	-	-	1.590	-	-
Season	Winter	−0.310 (0.145)	−0.467 (0.151)	−1.062 (0.276)	0.733	0.627	0.346
Surface Condition	Not Dry	-	−0.317 (0.154)	−1.100 (0.292)	-	0.728	0.333
AIC		3,810					
Log-Likelihood		−1,863.17					
McFadden's R[b]		0.073					

[a] Variable statistically significant at 90% otherwise significant at 95%.
[b] The empty cells show that the variable is not statistically significant to the respective model or not applicable.

not dry (likely due to more cautious behavior of drivers). On snow days with at least one inch of snow, the odds of Level C and Level KA severity outcomes estimated with respect to the PDO decreases by 33% and 71%, respectively.

5. Summary and conclusions

In Maine, lane departure crashes are the leading cause of crash fatalities. A majority of these crashes occur on rural roadways.

Table 6
Modeling results for minor arterials.

Variables		Estimate (Std.)			Odds Ratio		
		C	B	KA	C	B	KA
Intercept		−0.923 (0.220)	−1.310 (0.273)	−3.232 (0.483)	-	-	-
Driver Age	Older	–[b]	0.875 (0.235)	0.918 (0.440)	-	2.398	2.504
Male Driver Indicator	Male	−0.344 (0.120)	−0.360 (0.163)	-	0.709	0.698	-
Driver License	Suspended	0.493 (0.290)[a]	0.994 (0.332)	1.354 (0.478)	1.637	2.702	3.871
Sobriety	OUI	0.351 (0.192)	-	-	1.420	-	-
Seatbelt	Not Wearing	1.066 (0.221)	1.561 (0.250)	3.183 (0.296)	2.905	4.764	24.107
Crash Type	Rollover	0.870 (0.307)	0.988 (0.394)	1.316 (0.568)	2.388	2.685	3.728
Speed Limit	≥ 45mph	0.376 (0.175)	-	-	1.456	-	-
Curve	Present	-	0.255 (0.158)[a]	-	-	1.291	-
Season	Winter	−0.591 (0.138)	−0.784 (0.182)	−1.039 (0.298)	0.554	0.456	0.354
Surface Condition	Not Dry	-	−0.373 (0.190)	−0.996 (0.357)	-	0.689	0.369
Snow	≥ 1 inch of snow	−0.310 (0.187)[a]	−0.679 (0.316)	-	0.733	0.507	-
AIC		3,565					
Log-Likelihood		−1,743.62					
McFadden's R[b]		0.092					

[a] Variable statistically significant at 90% otherwise significant at 95%.
[b] The empty cells show that the variable is not statistically significant to the respective model or not applicable.

Table 7
Modeling results for major collectors.

Variables		Estimate (Std.)			Odds Ratio		
		C	B	KA	C	B	KA
Intercept		−0.694 (0.133)	−1.748 (0.177)	−3.691 (0.314)	-	-	-
Driver Age	Middle	–[b]	-	0.385 (0.171)	-	-	1.470
	Older	0.645 (0.126)	0.327 (0.175)[a]	1.232 (0.250)	1.906	1.386	3.428
Male Driver Indicator	Male	−0.472 (0.076)	-	−0.346 (0.170)	0.624	-	0.707
Sobriety	OUI	0.374 (0.131)	0.558 (0.147)	0.838 (0.203)	1.454	1.744	2.312
Nighttime and speeding	Yes	0.694 (0.322)	-	1.589 (0.405)	2.003	-	4.897
Seatbelt	Not Wearing	1.023 (0.138)	1.517 (0.147)	3.137 (0.178)	2.781	4.558	23.039
Crash Type	Rollover	0.875 (0.184)	1.040 (0.223)	1.358 (0.331)	2.398	2.830	3.886
Time of Day	Peak	-	−0.195 (0.102)[a]	-	-	0.823	-
Speed Limit	≥ 45mph	0.203 (0.094)	0.208 (0.122)[a]	0.812 (0.222)	1.225	1.231	2.253
Curve	Present	-	0.205 (0.098)	0.317 (0.159)	-	1.228	1.372
Season	Winter	−0.565 (0.094)	−0.340 (0.116)	−1.041 (0.184)	0.568	0.712	0.353
Surface Condition	Not Dry	-	−0.483 (0.122)	-	-	0.617	-
Snow	≥ 1 inch of snow	−0.230 (0.116)	−0.868 (0.183)	−1.511 (0.483)	0.795	0.420	0.221
Precipitation	Yes	-	0.193 (0.106)[a]	-	-	1.212	-
AIC		8,956					
Log-Likelihood		−4,432.92					
McFadden's R[b]		0.095					

[a] Variable statistically significant at 90% otherwise significant at 95%.
[b] The empty cells show that the variable is not statistically significant to the respective model or not applicable.

Table 8
Modeling results for minor collectors.

Variables		Estimate (Std.)			Odds Ratio		
		C	B	KA	C	B	KA
Intercept		−1.026 (0.190)	−1.720 (0.247)	−3.917 (0.458)	-	-	-
Driver Age	Middle	–[b]	-	0.458 (0.256)[a]	-	-	1.581
	Older	-	0.517 (0.268)[a]	1.298 (0.397)	-	1.677	3.661
Male Driver Indicator	Male	−0.655 (0.114)	−0.251 (0.151)[a]	-	0.520	0.778	-
Drive Speed	Speeding	0.455 (0.277)[a]	0.802 (0.300)	0.907 (0.409)	1.576	2.231	2.476
Seatbelt	Not Wearing	1.423 (0.206)	1.618 (0.230)	2.659 (0.276)	4.149	5.043	14.276
Crash Type	Rollover	0.576 (0.270)	-	-	1.779	-	-
Nighttime and OUI	Yes	-	0.971 (0.259)	0.962 (0.360)	-	2.641	2.616
Speed Limit	≥ 45mph	-	0.380 (0.170)	0.930 (0.301)	-	1.462	2.534
Curve	Present	-	-	0.634 (0.262)	-	-	1.884
Grade	Not Level	0.243 (0.114)	-	-	1.275	-	-
Season	Winter	-	−0.573 (0.182)	−0.663 (0.272)	-	0.564	0.516
Surface Condition	Not Dry	−0.490 (0.142)	−0.393 (0.188)	−0.623 (0.292)	0.675	0.613	0.536
Snow	≥ 1inch of snow	−0.400 (0.161)	-	−1.245 (0.621)	0.671		0.288
AIC		4,012					
Log-Likelihood		−1,964.144					
McFadden's R[b]		0.085					

[a] Variable statistically significant at 90% otherwise significant at 95%.
[b] The empty cells show that the variable is not statistically significant to the respective model or not applicable.

Maine is unique in many ways, such as aging infrastructure and population, a challenging climate, and diverse terrain. This study used Multinomial Logit Regression model to estimate severity outcome models for four facility types (Interstates, minor arterials, major collectors, and minor collectors) to analyze the impact of roadway, driver, and weather factors on severity of crashes. The PDO category was used as the reference (or base) category. Therefore, odds ratio results are with respect the PDO outcome. Older drivers (aging 65 and older) variable was significant for all analyzed facilities. Crashes that involved older drivers showed increased odds of Level KA severity outcome by 330%, 150%, 243%, and 266% on Interstate, minor arterials, major collectors, and minor collectors, respectively, compared to younger drivers. Failure to use a seatbelt was the most influential variable causing severe crashes. When the seatbelt is not used, the odds of Level KA severity outcome estimated with respect to the PDO increases by 27.4-, 23.1-, 23.0-, and 13.3- times higher on Interstate, minor arterials, major collectors, and minor collectors, respectively. As discussed during the winter period, there are significantly more PDO crashes for each facility type (due to inclement weather). However, the severity of crashes does not necessarily increase in proportion to PDOs. During the winter period, the results show that the odds of crashes resulting in Level KA severity outcome estimated with respect to the PDO decreases by 65%, 65%, 65%, and 48% for Interstate, minor arterial, major collectors, and minor collector facilities, respectively. The crash data were also mapped to daily weather data obtained from weather stations to use various weather variables in the model. The modeling results show that crashes that occur on snow days have decreased odds of resulting in Level KA severity outcome by 78% and 71% on major and minor collectors, respectively. When the surface is not dry, the odds of Level KA severity outcome decreases by 67%, 63%, and 46% on Interstates, minor arterials, and minor collectors, respectively. Inclement weather or bad surface conditions result in more PDO but less severe crash outcomes since drivers are more cautious, use lower speeds, and are more aware in these conditions.

6. Practical applications

In Maine, lane departure crashes account for over 70% of roadway fatalities. The state of Maine experiences adverse winter weather conditions, experiences the third coldest temperatures in the United States, has varying geography, houses the oldest population in the United States, and has old roadway infrastructure. These factors are also relatable to several other rural states (e.g., Vermont, New Hampshire) and rural areas of more urban states in the northeast where limited research considering crash severity has been done due to gaps in reliable data collection. In fact, Maine is the only state in New England that is part of the Highway Safety Information System (HSIS). The state actively collects an abundance of reliable and useful data on highway safety, including roadway, crash, and traffic data. Such reliable data on rural roads are crucial for robust analysis. This study developed statistical models to analyze severity of lane departure crashes in Maine, considering various driver, roadway, and weather factors. The findings of this study provide insights for safety analysts, practitioners, and agencies in Maine (as well as other states in the northeast or Atlantic regions in Canada) to better understand the factors impacting lane departure crash severities at four rural facility types (i.e., minor collectors, major collectors, minor arterials, principal arterials-Interstates) in order to allocate necessary funds to develop countermeasures or improve safety across the state.

Declaration of Competing Interest

The authors declare that they have no known competing financial or conflicting interests.

Acknowledgments

This research was funded by the Maine Department of Transportation (Maine DOT) and the Region 1 UTC program [Transportation Infrastructure Durability Center (TIDC)]. Authors would like to thank Mr. Robert Skehan, Mr. Dale Peabody, Mr. Brian Burne and Mr. Edward Beckwith from Maine Department of Transportation for their support throughout the research. The opinions expressed by the authors in this research do not necessarily reflect those of the Maine DOT or TIDC UTC program.

References

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research, 34*(5), 597–603.

Bouchard, D., Bizier, E., Kuchinski, J., Hayes, H., Merfield, P., Frazier, L., & Northgraves, J. (2020). *Report Card for Maine's Infrastructure*. ASCE.

Edwards, J. B. (1998). The relationship between road accident severity and recorded weather. *Journal of Safety Research, 29*(4), 249–262.

Eisenberg, D., & Warner, K. E. (2005). Effects of snowfalls on motor vehicle collisions, injuries, and fatalities. *American Journal of Public Health, 95*(1), 120–124.

Environmental Systems Research Institute (ESRI), (2020).ArcGIS Pro(Version 2.5). Esri Inc. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview.

Geedipally, S. R., Gates, T. J., Stapleton, S., Ingle, A., & Avelar, R. E. (2019). Examining the safety performance and injury severity characteristics of rural county roadways. *Transportation Research Record, 2673*(10), 405–415.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

Himes, C., & Kilduff, L. Which U.S. States Have the Oldest Populations? *Resource Library*. (2019). http://www.prb.org/resources/which-us-states-are-the-oldest/.

Holdridge, J. M., Shankar, V. N., & Ulfarsson, G. F. (2005). The crash severity impacts of fixed roadside objects. *Journal of Safety Research, 36*(2), 139–147.

Kim, J. K., Ulfarsson, G. F., Kim, S., & Shankar, V. N. (2013). Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis & Prevention, 50*, 1073–1081.

Li, Z., Ci, Y., Chen, C., Zhang, G., Wu, Q., Qian, Z. S., & Ma, D. T. (2019). Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. *Accident Analysis & Prevention, 124*, 219–229.

Meyer, J. (2001). Age:2000 Census 2000 Brief. *U.S. Census Bureau.* Retrieved From https://www.census.gov/prod/2001pubs/c2kbr01-12.pdf.

Morgan, A., & Mannering, F. L. (2011). The effects of road-surface conditions, age, and gender on driver-injury severities. *Accident Analysis & Prevention, 43*(5), 1852–1863.

NOAA National Centers for Environmental Information, State of the Climate: National Climate Report for April 2019, published online May 2019, retrieved on June 14, 2021 from https://www.ncdc.noaa.gov/sotc/national/201904.

Qiu, L., & Nixon, W. A. (2008). Effects of adverse weather on traffic crashes: Systematic review and meta-analysis. *Transportation Research Record, 2055*(1), 139–146.

Rahman, M. A., Sun, X., Das, S., & Khanal, S. (2021). Exploring the influential factors of roadway departure crashes on rural two-lane highways with logit model and association rules mining. *International Journal of Transportation Science and Technology, 10*(2), 167–183.

Shankar, V., & Mannering, F. (1996). An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *Journal of Safety Research, 27*(3), 183–194.

Shaheed, M. S., Gkritza, K., Carriquiry, A. L., & Hallmark, S. L. (2016). Analysis of occupant injury severity in winter weather crashes: A fully Bayesian multivariate approach. *Analytic Methods in Accident Research, 11*, 33–47.

Shirazi, M., Geedipally, S. R., & Lord, D. (2017). A Monte-Carlo simulation analysis for evaluating the severity distribution functions (SDFs) calibration methodology and determining the minimum sample-size requirements. *Accident Analysis & Prevention, 98*, 303–311.

Sawtelle, A., Shirazi, M., Garder, P. E., & Rubin, J. (2022). Exploring the impact of seasonal weather factors on frequency of lane-departure crashes in Maine. *Journal of Transportation Safety & Security*, 1–22.

Theofilatos, A. (2017). Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. *Journal of Safety Research, 61*, 9–21.

World Population Review. (2021). Coldest States 2021. Coldest States 2021. https://worldpopulationreview.com/state-rankings/coldest-states.

Washington, S., Karlaftis, M., Mannering, F., & Anastasopoulos, P. (2020). *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC.

Wu, Q., Zhang, G., Zhu, X., Liu, X. C., & Tarefder, R. (2016). Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways. *Accident Analysis & Prevention, 94*, 35–45.
Zhang, X., Wen, H., Yamamoto, T., & Zeng, Q. (2021). Investigating hazardous factors affecting freeway crash injury severity incorporating real-time weather data: Using a Bayesian multinomial logit model with conditional autoregressive priors. *Journal of Safety Research, 76*, 248–255.
Zhao, X., Ding, Y., Yao, Y., Zhang, Y., Bi, C., & Su, Y. (2021). A multinomial logit model: Safety risk analysis of interchange area based on aggregate driving behaviour data. *Journal of Safety Research, 12*.
Zhao, S., Wang, K., Liu, C., & Jackson, E. (2019). Investigating the effects of monthly weather variations on Connecticut freeway crashes from 2011 to 2015. *Journal of Safety Research, 71*, 153–162.

Alainie Sawtelle currently is a Graduate Research Assistant at the University of Maine, pursuing her Master of Science degree in Civil Engineering with focus in Transportation Engineering. Alainie graduated with a bachelor's degree in Civil Engineering from the University of Maine in 2020. Alainie's research includes analyzing frequency and severity of lane-departure crashes in Maine. Alainie plans to enter the industry as a consulting engineer upon completion of her degree.

Mohammadali Shirazi, Ph.D. is an Assistant Professor in Civil and Environmental Engineering at University of Maine. Dr. Shirazi's research is geared towards the exploration of state-of-the-art statistical models to analyze traffic safety. Dr. Shirazi received his Ph.D. in Civil Engineering from Texas A&M University in 2018. After his Ph.D. degree, he worked as a research fellow at University of Michigan. Since joining the University of Maine in September 2019, Dr. Shirazi has led several research projects as a PI or Co-PI.

Per Erik Garder, Ph.D. is a Professor of Civil Engineering at the University of Maine, USA. Dr. Gårder received his Ph.D. from Lund University in Sweden. From 1983 to 1992, Dr. Garder was a faculty member of the Royal Institute of Technology (KTH) in Stockholm, Sweden. His research interest is focused on evaluating facilities with emphasis on traffic safety. He is an Associate Editor of Accident Analysis and Prevention. He was also the Editor of the section of Transportation Safety and Security in Elsevier's 2021 International Encyclopedia of Transportation, First Edition.

Jonathan Rubin, Ph.D. is the Director of the Margaret Chase Smith Policy Center and Professor of Economics at the University of Maine. Rubin received his Ph.D. in Agricultural Economics from the University of California, Davis. Dr. Rubin specializes in the economics of energy, light-duty transportation, greenhouse gas emissions and alternative fuels. His research investigates the environmental impacts of connected autonomous vehicles, low carbon transportation fuels and biofuel pathways.

Journal of Safety Research 84 (2023) 1–6

Contents lists available at ScienceDirect

Journal of Safety Research

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

Drugged driving among U.S. adolescents, 2016–2019, USA

Robert Andrew Yockey [a,*], Cristina S. Barroso [b]

[a] Department of Biostatistics and Epidemiology, University of North Texas Health Science Center, Fort Worth, TX 76107, United States
[b] College of Nursing, University of Tennessee, Knoxville, TN 37996, United States

ARTICLE INFO

Article history:
Received 6 January 2022
Received in revised form 21 July 2022
Accepted 17 October 2022
Available online 26 October 2022

Keywords:
Drugged driving
Adolescence
Health behavior

ABSTRACT

Introduction: Drugged driving, the operation of a vehicle under the influence of any illegal drugs and alcohol, is a growing problem, but remains understudied among adolescents. The purpose of this article is to estimate past-year driving under the influence of alcohol, marijuana, and other drugs among a large sample of U.S. adolescents and potential associations (e.g., age, race, metropolitan status, sex). *Design:* A cross-sectional secondary data analysis of the 2016–2019 National Survey on Drug Use and Health among 17,520 adolescents ages 16–17-years old was conducted. Weighted logistic regression models were built to determine potential associations to drugged driving. *Results:* An estimated 2.00% of adolescents drove under the influence of alcohol in the past year, 5.65% drove under the influence of marijuana in the past year, and an estimated 0.48% drove under the influence of other drugs other than marijuana in the past year. Differences were based on race, past-year drug use, and county status. *Conclusions:* Drugged driving is a growing problem among adolescents and interventions are greatly needed to mitigate these behaviors among youth.

1. Introduction

Nearly 258,000 adolescents in the United States were treated in the emergency department from injuries resulting in automobile accidents in 2019 (Centers for Disease Control and Prevention (CDC), 2021), placing significant burdens on healthcare facilities and structures. Globally, more than 1.2 million people are killed due to road traffic injuries and nearly 50 million are injured, increasing hospitalizations and treatment costs (World Health Organization, 2009). Furthermore, road traffic injuries are a significant indicator of morbidity and mortality among younger populations. In 2019, an estimated 2,400 adolescents in the United States died due to vehicle car collisions, making it the leading cause of death among adolescents (Blum & Qureshi, 2011).

Use of alcohol and other drugs (e.g., marijuana, prescription opioids) are significant risk factors for these crashes (Brookoff, Cook, Williams, & Mann, 1994). Known as "drugged driving," several consequences can result from drugged driving including impairment in driving performance, perceptions of distance, increased risk for crashing/injury, and inaccurate time perceptions (Brookoff et al., 1994; Compton, 2017; Sewell, Poling, & Sofuoglu, 2009). This is of concern, considering that in 2015 more people were killed from drugged driving in the United States than drunk driving (43% vs 37%, respectively). (Hedlund, 2017).

Risk factors for drugged driving are complex, but evidence points toward a multi-level framework (e.g., cultural, psychosocial) to explain why individuals engage in these risky behaviors. For example, one study found that sensation seeking, negative emotional driving, and impulsivity were significant associations to driving under the influence of cannabis (Richer & Bergeron, 2009). Another study of individuals in Spain found that greater family problems/disruptions, prior drug use, and identifying as male were strong predictors of driving under the influence (Tomas Dols et al., 2010).

There are limited data on drugged driving among adolescents from national databases (DuPont, Logan, & Shea, 2011; O'Malley and Johnston, 2007, 2013; Terry-McElrath, O'Malley, & Johnston, 2014). Moreover, there is a paucity of recent literature on drugged driving among adolescents. It is important to provide the most recent estimates for harm reduction and behavioral initiatives (e.g., interventions, education classes). Further, to inform prevention efforts and strategies to bolster primary care, a call for more studies on drugged driving among adolescents has been noted (Knight et al., 2018). To our knowledge, this is one of the first studies to examine alcohol use, marijuana use, and 'other drug' use while driving among a large, nationally representative sample of youth using recent national data. The National Institute on Drug Abuse (DuPont et al., 2011) has commissioned a white paper and

* Corresponding author.
E-mail address: Robert.yockey@unthsc.edu (R. Andrew Yockey).

https://doi.org/10.1016/j.jsr.2022.10.002
0022-4375/

has warranted more research into drugged driving by using data from national sources/databases, specifically the National Survey on Drug Use and Health, to track the prevalence and correlates of drugged driving. Using four years of nationally representative data, the purpose of the present study was to examine the prevalence of drugged driving (use of alcohol, marijuana, and illicit drugs such as methamphetamine and cocaine) among adolescents (ages 16–17) and potential associations (e.g., race, metropolitan status, depression status, biological sex, past-year substance use).

2. Methods

Pooled data from the 2016–2019 National Survey on Drug Use and Health (NSDUH) were analyzed. Briefly, the NSDUH is a cross-sectional, annual, nationally representative survey conducted in the United States to assess substance use, behavioral health utilization, and mental health prevalence among individuals 12 years or older. The NSDUH utilizes a complex sampling design to ensure adequate representation and probability selection of individuals. Other details of the NSDUH are detailed elsewhere (Substance Abuse and Mental Health Services Administration, 2015). The present analysis was restricted to 16–17 year olds, as most U.S. states grant driver licenses at age 16 and this age group will likely be licensed. Further, the analysis was a complete case analysis, given the small proportion of missingness and the nature of categorical coding implemented to our variables (Allison, 2005; Audigier, Husson, & Josse, 2017).

3. Outcomes

3.1. Driving under the influence of alcohol or marijuana

Past-year driving under the influence of alcohol was assessed by the question: "Did you drive a car or vehicle while under the influence of alcohol in the past-12 months?" Past-year driving under the influence of marijuana was assessed by the question: "Did you drive a car or vehicle while under the influence of marijuana in the past-12 months?" Responses were binary in nature (1 = "Yes," 0 = "No").

3.2. Illicit drugged driving

Several questions were asked to the participants of whether they operated a vehicle while under the influence of cocaine, inhalants, methamphetamine, heroin, or hallucinogens, and the NSDUH survey combined this into an "illicit" drugged driving variable (e.g., "Did you drive under the influence of illicit drugs in the past year"?). Answer responses were binary in nature (1 = "Yes," 0 = "No"). Previous research has shown these drugs to be used comorbidly with alcohol or marijuana while driving (Robertson, Mainegra Hing, Pashley, Brown, & Vanlaar, 2017).

4. Depression

Self-reported past year major depressive episode was assessed, since previous research (Karjalainen, Lintonen, Joukamaa, & Lillsunde, 2013) has found an association between mental disorders and drugged driving. Response options were (1 = "Yes," 0 = "No").

4.1. Demographics

Participants' self-reported sex (male, female), race (Non-Hispanic White, Non-Hispanic African American, Hispanic, and Other) and county (large metro, small metro, and non-metro) were used. Here, 'Other' is a combination of Pacific Islander, Native American, and mixed race.

5. Analysis

The use of multiple-imputed variables, when available, provided by NSDUH were utilized to limit the amount of missing data. Frequencies and descriptive statistics with appropriate 95% confidence intervals were calculated to capture the sample characteristics. Bivariate associations were made with Rao-Scott chi-square. We calculated adjusted prevalence ratios (aPRs) (Cummings, 2009) with a multivariable generalized linear model using Poisson and log link for each covariate. All analyses were conducted in Stata 17.0 (College Station, TX) and used the 'svy' commands, were two-tailed, weighted to be representative of the U.S. population, and were designed-based (Heeringa, West, & Berglund, 2017). We included survey year to control for the random intercept of the variable. Analyses took place in Stata (version 17.0). A University Institutional Review Board deemed the present analyses not to be human subjects research and was therefore exempt from review.

6. Results

6.1. Sample characteristics and drugged driving prevalence

The analytic sample consisted of 17,520 adolescents aged 16–17-years old. The sample consisted of nearly equal percentages of boys and girls (50.7 vs 49.4%, respectively). Within the past year, an estimated 2.00% of adolescents drove under the influence of alcohol, 5.65% drove under the influence of marijuana, and 0.48% drove under the influence of other drugs other than marijuana.

6.2. Driving under the influence of alcohol

Compared to 2016, adolescents in 2019 were less likely to drive under the influence of alcohol in the past year (aPR: 0.63, 95% CI 0.42, 0.93). Adolescents living in non-metro areas were 1.45 times (95% CI 1.06, 1.97) more likely to drive under the influence of alcohol, compared to adolescents living in metro areas. Adolescents who used marijuana (aPR: 8.55, 95% CI 5.97, 12.2) or illicit drugs other than marijuana (aPR: 2.81, 95% CI 2.11, 3.74) in the past year were more likely to drive under the influence of alcohol in the past year. No other differences were found.

6.3. Driving under the influence of marijuana

Compared to males, females were less likely to drive under the influence of marijuana (aPR: 0.82, 95% CI 0.71, 0.96). Adolescents who reported a major depressive episode in the past year were 1.21 times (95% CI 1.03, 1.43) more likely to drive under the influence of marijuana in the past year. Adolescents who reported past year use of alcohol (aPR: 15.3, 95% CI 1.3, 20.8) or use of illicit drugs other than marijuana (aPR: 3.75, 95% CI 3.22, 4.37) were more likely to drive under the influence of marijuana within the past year. No other differences (e.g., race, county status) were found.

6.4. Driving under the influence of illicit drugs other than marijuana

The only significant predictors of driving under the influence of illicit drugs other than marijuana were past year use of alcohol and marijuana. Specifically, adolescents who used alcohol in the past year were 4.93 times more likely [95% CI 1.62, 15.0] to drive under the influence of illicit drugs other than marijuana within the past year. Further, adolescents who used marijuana in the past year

Table 1
Demographic Characteristics.

Variable	Full Sample (n = 17,520)	No Driving Under the Influence of Alcohol (n = 17,113)	Driving Under the Influence of Alcohol (n = 407)	No Driving Under the Influence of Marijuana (n = 16,442)	Driving Under the Influence of Marijuana (n = 1,078)	No Driving Under the Influence of Illicit Drugs (n = 17,433)	Driving Under the Influence of Illicit Drugs (n = 87)
Survey Year							
2016	24.8 (24.0, 25.6)	97.9 (97.4, 98.4)	**2.06 (1.60, 2.65)***	94.4 (93.5, 95.2)	5.60 (4.82, 6.51)	99.6 (99.3, 99.8)	0.42 (0.24, 0.72)
2017	25.5 (24.7, 26.3)	97.8 (97.2, 98.3)	2.19 (1.73, 2.77)	94.2 (93.4, 95.0)	5.77 (4.99, 6.65)	99.4 (99.0, 99.8)	0.60 (0.36, 0.99)
2018	25.5 (24.7, 26.4)	97.6 (96.9, 98.2)	2.36 (1.81, 3.08)	94.2 (93.2, 95.1)	5.79 (4.92, 6.81)	99.6 (99.3, 99.8)	0.38 (0.21, 0.69)
2019	24.3 (23.4, 25.1)	98.6 (98.1, 99.0)	1.37 (1.00, 1.87)	94.6 (93.6, 95.4)	5.41 (4.59, 6.37)	99.5 (98.9, 99.7)	0.53 (0.27, 1.02)
Sex							
Male	50.7 (49.7, 51.6)	98.1 (97.7, 98.5)	1.88 (1.55, 2.28)	94.1 (93.4, 94.7)	5.91 (5.31, 6.58)	99.5 (99.3, 99.7)	0.47 (0.31, 0.73)
Female	49.4 (48.4, 50.3)	97.9 (97.5, 98.2)	2.13 (1.78, 2.55)	94.6 (94.0, 95.2)	5.38 (4.80, 6.01)	99.5 (99.3, 99.7)	0.49 (0.33, 0.72)
Race							
Non-Hispanic White	53.0 (52.1, 54.0)	97.5 (97.1, 97.9)	**2.50 (2.14, 2.93)*****	93.1 (92.5, 93.7)	**6.87 (6.26, 7.53)*****	99.5 (99.2, 99.6)	0.54 (0.38, 0.78)
Non-Hispanic African American	13.6 (12.9, 14.2)	99.2 (98.6, 99.6)	0.78 (0.44, 1.37)	96.8 (95.8, 97.5)	3.25 (2.48, 4.24)	99.8 (99.2, 99.9)	0.24 (0.07, 0.78)
Hispanic	23.8 (22.9, 24.7)	98.4 (97.8, 98.8)	1.63 (1.18, 2.25)	95.4 (94.3, 96.2)	4.65 (3.80, 5.67)	99.6 (99.1, 99.8)	0.41 (0.19, 0.86)
Multi-Racial/ Other	9.63 (9.08, 10.2)	98.1 (97.0, 98.8)	1.89 (1.20, 2.96)	95.2 (93.9, 96.3)	4.77 (3.69, 6.15)	99.4 (98.7, 99.7)	0.65 (0.31, 1.36)
County Status							
Large metro	56.9 (56.0, 57.9)	98.3 (97.9, 98.6)	**1.71 (1.39, 2.11)***	94.8 (94.1, 95.3)	**5.25 (4.69, 5.88)****	99.6 (99.4, 99.7)	0.41 (0.26, 0.64)
Small metro	29.2 (28.3, 30.0)	97.7 (97.1, 98.1)	2.34 (1.87, 2.91)	93.4 (92.5, 94.2)	6.64 (5.85, 7.52)	99.5 (99.1, 99.7)	0.55 (0.33, 0.90)
Non-metro	13.9 (13.3, 14.5)	97.5 (96.9, 98.0)	2.50 (1.99, 3.13)	94.8 (93.8, 95.7)	5.19 (4.35, 6.18)	99.4 (98.9, 99.6)	0.63 (0.36, 1.10)
Past-Year Major Depressive Episode							
No	81.5 (80.8, 82.3)	98.4 (98.1, 98.6)	**1.65 (1.40, 1.94)*****	95.2 (94.7, 95.6)	**4.83 (4.40, 5.31)*****	99.6 (99.5, 99.8)	**0.37 (0.25, 0.54)*****
Yes	18.5 (17.8, 19.3)	96.4 (95.5, 97.2)	3.57 (2.85, 4.47)	90.8 (89.4, 91.9)	9.24 (8.06, 10.6)	99.0 (98.5, 99.4)	0.98 (0.62, 1.54)
Past-Year Drug Use							
Alcohol	38.4 (37.5, 39.3)	–	–	86.3 (85.2, 87.3)	**13.7 (12.7, 14.8)*****	98.8 (98.4, 99.1)	**1.16 (0.86, 1.57)*****
Marijuana	23.6 (22.8, 24.4)	93.2 (92.2, 94.2)	**6.77 (5.85, 7.81)*****	–	–	98.1 (97.5, 98.6)	**1.87 (1.38, 2.52)*****
Illicit drugs other than marijuana	10.5 (9.89, 11.1)	90.9 (89.2, 92.4)	**9.08 (7.57, 10.9)*****	73.8 (71.2, 76.2)	**26.2 (23.8, 28.9)*****	–	–

CI: confidence interval.
*Behaviors are not mutually exclusive – e.g., can drink drive and drive under influence of marijuana.

Table 2
Adjusted Prevalence Ratios and 95% CI with Drugged Driving.

Variable	Driving Under the Influence of Alcohol (aPR)	95% CI	Driving Under the Influence of Marijuana (aPR)	95% CI	Driving Under the Influence of Illicit Drugs (aPR)	95% CI
Survey Year						
2016	1.00	Ref.	1.00	Ref.	1.00	Ref.
2017	1.01	[0.72, 1.41]	0.98	[0.81, 1.19]	1.32	[0.63, 2.74]
2018	1.13	[0.79, 1.61]	1.07	[0.87, 1.32]	0.82	[0.37, 1.84]
2019	0.63	[0.42, 0.93]*	0.97	[0.79, 1.18]	1.10	[0.47, 2.56]
Sex						
Male	1.00	Ref.	1.00	Ref.	1.00	Ref.
Female	1.09	[0.83, 1.43]	0.82	[0.71, 0.96]**	0.84	[0.46, 1.51]
Race						
Non-Hispanic White	1.00	Ref.	1.00	Ref.	1.00	Ref.
Non-Hispanic African American	0.41	[0.23, 0.74]	0.87	[0.67, 1.14]	0.73	[0.21, 2.54]
Hispanic	0.79	[0.55, 1.12]	0.80	[0.75, 1.23]	0.97	[0.42, 2.23]
Multi-Racial/ Other	0.94	[0.60, 1.48]	0.96	[0.75, 1.23]	1.76	[0.78, 3.94]
County Status						
Large metro	1.00	Ref.	1.00	Ref.	1.00	Ref.
Small metro	1.21	[0.90, 1.62]	1.17	[1.00, 1.37]	1.28	[0.65, 2.52]
Non-metro	1.45	[1.06, 1.97]*	0.97	[0.79, 1.18]	1.70	[0.81, 3.59]
Past-Year Major Depressive Episode						
No	1.00	Ref.	1.00	Ref.	1.00	Ref.
Yes	1.28	[0.96, 1.73]	1.21	[1.03, 1.43]*	1.77	[0.97, 3.23]
Past-Year Drug Use						
Alcohol	–	–	15.3	[11.3, 20.8]***	4.93	[1.62, 15.0]***
Marijuana	8.55	[5.97, 12.2]***	–	–	16.4	[5.18, 51.9]***
Illicit drugs other than marijuana	2.81	[2.11, 3.74]***	3.75	[3.22, 4.37]***	–	

***p <.0001, **p <.001, *p <.05.
PR: prevalence ratio.

were 16.4 times more likely [95% CI 5.18, 51.9] to drive under the influence of illicit drugs other than marijuana in the past year. No other significant differences were found (see Table 1 and Table 2).

7. Discussion

7.1. Principal findings

Differences in drugged driving were found based on age, race, and metro status. An estimated 2.0% of adolescents drove under the influence of alcohol in the past year, and greater than 1 in 30 adolescents drove under the influence of marijuana and other drugs (e.g., non-medical prescription drugs, illegal drugs). Moreover, racial minorities were less likely to report drugged driving. Further, adolescents who self-reported major depression were more likely to engage in driving under the influence of cannabis, while adolescents who also used other drugs were more likely to engage in drugged driving.

7.2. Findings in context

Our findings are consistent with previous literature that racial minorities are at a lower risk of drunk driving (Yockey, Vidourek, & King, 2020), compared to White adolescents (O'Malley & Johnston, 2013). Moreover, for driving under the influence of marijuana, we found the opposite from previous literature (O'Malley & Johnston, 2013) mainly that racial minorities were less likely to report driving under the influence of marijuana and other drugs. Although not explicated, our findings may provide the necessary frameworks for an intersectionality approach (Gattamorta, Salerno, & Castro, 2019)(e.g., overlapping identities) regarding self-reported drugged driving, given that intersectionality may inform screening procedures for at-risk youth and bolster prevention services such as education classes and community resources[13] aimed at curbing drugged driving.

The use of other drugs in the past year had strong associations toward driving under the influence. This may be because attitudes toward substances such as marijuana are changing, with previous research highlighting several views among adolescents indicating that driving under the influence may lead to less social consequences (McCarthy, Lynch, & Pederson, 2007) and impairment in driving behaviors (Swift, Jones, & Donnelly, 2010). A developmental approach to risky behaviour (Arnett, 1992) in adolescence may prove useful in answering these trends – namely, that if left unchecked, risky behaviors culminate in adolescence and increase in severity and form over time, with a focus on sensation seeking, peer influence, and adolescent egocentrism (i.e., inability to distinguish between their perception of what others think about them and what people think). In the context of driving and the use of drugs, our findings suggest that a sizeable percentage of U.S. adolescents engage in these behaviors. Moreover, adolescents are more likely to engage in other reckless driving behaviors behind the wheel such as speeding, not wearing seatbelts, and infrequent use of seatbelts, even when not under the influence (Jonah, 1986). The use of several strategies (e.g., media campaigns, state sanctions) may deter current and future drugged driving behaviors at an early age (Razaghizad et al., 2021).

Although only significant for driving under the influence of marijuana, the finding that reporting a major depressive episode points to multifaceted approaches when treating driving under the influence. This finding is concerning, given the increased crash risk while driving under the influence of marijuana (Hill et al., 2017). Further, it is well established that poor mental health has a strong association with risky behaviors (Johnson & Taliaferro, 2012). Associated consequences of poor mental health (e.g., poor sleep) may worsen driving performance and, simultaneously, lead to impaired decision making, (Rubinsztein, Michael, Underwood, Tempest, & Sahakian, 2006) which may increase proneness to engaging in these types of behaviors.

Our findings also highlight county differences in drunk driving, but not driving under the influence of marijuana or other drugs. Specifically, unadjusted results showed adolescents who lived in small or non-metro counties were more likely to report driving under the influence of alcohol; adjusted results showed that adolescents who lived in small counties were more likely to report driving under the influence of alcohol. The place and environment where adolescents live are a strong influence on future psychosocial health behaviors and trajectories. Rural environments may predispose adolescents towards engaging in unhealthy behaviors (e.g., drugged driving) and may shape behaviors, attitudes, and beliefs about these behaviors (Veitch, 2009). The specific determinants of counties and their influence of drugged driving among adolescents remains a fruitful area for further research for harm reduction efforts and behavioral health initiatives.

8. Limitations

The strong national, probability-based sampling in the NSDUH and large number of participants in this study must be balanced against its limitations. Data were self-reported; thus, under/over enhancement of answers may be present. We did not assess past-month drugged driving; more studies are needed to determine recentness of drugged driving. Data are cross-sectional in nature; therefore, causality cannot be determined. Future research is also warranted on trends in drugged driving to examine an intersectionality approach towards elucidating within-racial differences. Wide confidence intervals within our estimates were present, therefore, caution is warranted when interpreting results.

9. Conclusions

Drugged driving among adolescents is an understudied, but a growing problem that warrants further attention. Although cross-sectional data were assessed, differences were found based on race/ethnicity, age, and locality. These differences are critical and should be considered for interventions (e.g., educational classes) when examining these relationships.

References

Allison, P. D. (2005). Imputation of categorical variables with PROC MI. *SUGI 30 Proceedings, 113*(30), 1–14.

Arnett, J. (1992). Reckless behavior in adolescence: A developmental perspective. *Developmental Review, 12*(4), 339–373. https://doi.org/10.1016/0273-2297(92)90013-R.

Audigier, V., Husson, F., & Josse, J. (2017). MIMCA: Multiple imputation for categorical variables with multiple correspondence analysis. *Statistics and Computing, 27*(2), 501–518.

Blum, R. W. & Qureshi F. (2011). Morbidity and mortality among adolescents and young adults in the United States. Baltimore, MD.

Brookoff, D., Cook, C. S., Williams, C., & Mann, C. S. (1994). Testing reckless drivers for cocaine and marijuana. *New England Journal of Medicine, 331*(8), 518–522. https://doi.org/10.1056/NEJM199408253310807. PMID: 8041419.

Centers for Disease Control and Prevention (CDC), (2021). National Center for Injury Preventionand Control (NCIPC). WISQARS (Web-based Injury Statistics Query and Reporting System). U.S. Department of Health and Human Services; July 2020. Available at: http://www.cdc.gov/injury/wisqars/index.html. Accessed 13 September 2021.

Compton, R. P. (2017). *Marijuana-impaired driving-a report to congress* (No. DOT HS 812 440). United States. National Highway Traffic Safety Administration.

Cummings, P. (2009). The relative merits of risk ratios and odds ratios. *Archives of Pediatrics and Adolescent Medicine, 163*(5), 438–445. https://doi.org/10.1001/archpediatrics.2009.31. PMID: 19414690.

DuPont, R. L., Logan, B. K. & Shea, C. L., et al. (2011). Drugged driving research: A white paper. National Institute on Drug Abuse, MD. Retrieved from https://www.ojp.gov/ncjrs/virtual library/abstracts/drugged-driving-research-white-paper.

Gattamorta, K. A., Salerno, J. P., & Castro, A. J. (2019). Intersectionality and health behaviors among US high school students: Examining race/ethnicity, sexual identity, and sex. *Journal of School Health, 89*(10), 800–808. https://doi.org/10.1111/josh.12817. Epub 2019 Jul 28 PMID: 31353476.

Hedlund, J. (2017) Governors Highway Safety Association. Drug-impaired driving: a guide for states. Retrieved from https://www.ghsa.org/sites/default/files/2017 04/GHSA_DruggedDriving2017_FINAL.pdf.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2017). *Applied survey data analysis*. Chapman and Hall/CRC.

Hill, L. L., Lauzon, V. L., Winbrock, E. L., Li, G., Chihuri, S., & Lee, K. C. (2017). Depression, antidepressants and driving safety. *Injury Epidemiology, 4*(1), 10. https://doi.org/10.1186/s40621-017-0107-x. Epub 2017 Apr 3. PMID: 28367591; PMCID: PMC5376538.

Johnson, K. E., & Taliaferro, L. A. (2012). Health behaviors and mental health of students attending alternative high schools: A review of the research literature. *J Spec Pediatr Nurs., 17*(2), 79–97. https://doi.org/10.1111/j.1744-6155.2011.00311.x. Epub 2011 Nov 1 PMID: 22463469.

Jonah, B. A. (1986). Accident risk and risk-taking behaviour among young drivers. *Accident Analysis and Prevention, 18*(4), 255–271. https://doi.org/10.1016/0001-4575(86)90041-2. PMID: 3741578.

Karjalainen, K., Lintonen, T., Joukamaa, M., & Lillsunde, P. (2013). Mental disorders associated with driving under the influence of alcohol and/or drugs: A register-based study. *European Addiction Research, 19*(3), 113–120. https://doi.org/10.1159/000342569. Epub 2012 Nov 6 PMID: 23147745.

Knight, J. R., Csemy, L., Sherritt, L., Starostova, O., Van Hook, S., Bacic, J., et al. (2018). Screening and brief advice to reduce adolescents' risk of riding With substance-using drivers. *Journal of Studies on Alcohol and Drugs, 79*(4), 611–616. https://doi.org/10.15288/jsad.2018.79.611. PMID: 30079877; PMCID: PMC6090102.

McCarthy, D. M., Lynch, A. M., & Pederson, S. L. (2007). Driving after use of alcohol and marijuana in college students. *Psychology of Addictive Behaviors, 21*(3), 425–430. https://doi.org/10.1037/0893 164X.21.3.425. PMID: 17874895.

O'Malley, P. M., & Johnston, L. D. (2007). Drugs and driving by American high school seniors, 2001 2006. *Journal of Studies on Alcohol and Drugs, 68*(6), 834–842. https://doi.org/10.15288/jsad.2007.68.834. PMID: 17960301.

O'Malley, P. M., & Johnston, L. D. (2013). Driving after drug or alcohol use by US high school seniors, 2001–2011. *American Journal of Public Health, 103*(11), 2027–2034. https://doi.org/10.2105/AJPH.2013.301246. Epub 2013 Sep 12. PMID: 24028266; PMCID: PMC3828684.

Razaghizad, A., Windle, S. B., Gore, G., Benedetti, A., Ells, C., Grad, R., et al. (2021). Interventions to prevent drugged driving: A systematic review. *American Journal of Preventive Medicine, 61*(2), 267–280. https://doi.org/10.1016/j.amepre.2021.03.012. Epub 2021 Jun 5 PMID: 34099354.

Richer, I., & Bergeron, J. (2009). Driving under the influence of cannabis: Links with dangerous driving, psychological predictors, and accident involvement. *Accident Analysis and Prevention, 41*(2), 299–307. https://doi.org/10.1016/j.aap.2008.12.004. Epub 2009 Jan 17 PMID: 19245889.

Robertson, R. D., Mainegra Hing, M., Pashley, C. R., Brown, S. W., & Vanlaar, W. G. M. (2017). Prevalence and trends of drugged driving in Canada. *Accident Analysis and Prevention, 99*(Pt A), 236–241. https://doi.org/10.1016/j.aap.2016.12.008. Epub 2016 Dec 13 PMID: 27984814.

Rubinsztein, J. S., Michael, A., Underwood, B. R., Tempest, M., & Sahakian, B. J. (2006). Impaired cognition and decision-making in bipolar depression but no 'affective bias' evident. *Psychological Medicine, 36*(5), 629–639. https://doi.org/10.1017/S0033291705006689. Epub 2006 Jan 19 PMID:16420729.

Sewell, R. A., Poling, J., & Sofuoglu, M. (2009). The effect of cannabis compared with alcohol on driving. *American Journal on Addictions, 18*(3), 185–193. https://doi.org/10.1080/10550490902786934. PMID: 19340636; PMCID: PMC2722956.

Substance Abuse and Mental Health Services Administration. (2015). National Survey on Drug Use and Health: 2014 and 2015 Redesign Changes [Internet]. Rockville (MD): *Substance Abuse and Mental Health Services Administration (US); 2015 Aug. 2, Overview of NSDUH Sampling and Data Collection Procedures.* Available from: https://www.ncbi.nlm.nih.gov/books/NBK519780/.

Swift, W., Jones, C., & Donnelly, N. (2010). Cannabis use while driving: A descriptive study of Australian cannabis users. *Drugs: Education, Prevention and Policy., 17*(5), 573–586. https://doi.org/10.3109/09687630903264286.

Terry-McElrath, Y. M., O'Malley, P. M., & Johnston, L. D. (2014). Alcohol and marijuana use patterns associated with unsafe driving among U.S. high school seniors: High use frequency, concurrent use, and simultaneous use. *Journal of Studies on Alcohol and Drugs, 75*(3), 378–389. https://doi.org/10.15288/jsad.2014.75.378. PMID: 24766749; PMCID: PMC4002852.

Tomas Dols, S., Alvarez González, F. J., Llorens Aleixandre, N., Vidal-Infer, A., Torrijo Rodrigo, M. J., & Valderrama-Zurián, J. C. (2010). Predictors of driving after alcohol and drug use among adolescents in Valencia (Spain). *Accident Analysis and Prevention, 42*(6), 2024–2029. https://doi.org/10.1016/j.aap.2010.06.013. Epub 2010 Jul 22 PMID: 20728658.

Veitch, C. (2009). Impact of rurality on environmental determinants and hazards. *Australian Journal of Rural Health, 17*(1), 16–20. https://doi.org/10.1111/j.1440-1584.2008.01031.x.

World Health Organization (2009). *Global status report on road safety: Time for action*. World Health Organization.

Yockey, A., Vidourek, R., & King, K. (2020). Drugged driving among US adults: Results from the 2016 2018 national survey on drug use and health. *J Safety Res., 75*, 8–13. https://doi.org/10.1016/j.jsr.2020.10.006. Epub 2020 Nov 15 PMID: 33334496.

R. Andrew Yockey, PhD, is an Assistant Professor of Biostatistics and Epidemiology at the University of North Texas Health Science Center. His research interests include quantitative methods, psychiatric epidemiology, substance use prevention, and survey research.

Cristina Barroso, DrPH, is an Associate Professor in the College of Nursing at the University of Knoxville. Her research interests include adolescent health, physical activity, substance use prevention, and health disparities.

JOURNAL OF SAFETY RESEARCH

Journal of Safety Research 84 (2023) 330–346

Contents lists available at ScienceDirect

Journal of Safety Research

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

Effectiveness of interventions for mobile phone distracted pedestrians: A systematic review

Md Eaysir Arafat [a,*], Grégoire S. Larue [a,b], Sepehr Ghasemi Dehkordi [a,c]

[a] Queensland University of Technology (QUT), Centre for Accident Research and Road Safety-Queensland (CARRS-Q), Brisbane, Australia
[b] University of the Sunshine coast (UniSC), Road Safety Research Collaboration, Sippy Downs, Australia
[c] Australian Road Research Board (ARRB), Brisbane, Australia

ARTICLE INFO

Article history:
Received 14 March 2022
Received in revised form 11 September 2022
Accepted 14 November 2022
Available online 23 November 2022

Keywords:
Mobile phone
Distraction
Vulnerable road users
Countermeasures
Behavior

ABSTRACT

Introduction: Mobile phones are used universally due to their versatility and easy-to-use features; this includes when users are walking and when crossing streets. At intersections, using a mobile phone is a secondary task that can distract from the primary task of scanning the road environment and ensuring it is safe to traverse. Such a distraction has been shown to increase risky pedestrian behavior compared to non-distracted behavior. Developing an intervention to make distracted pedestrians aware of imminent danger is a promising approach to refocus pedestrians on their primary task and avoid incidents. Interventions have already been developed in different parts of the world, such as in-ground flashing lights, painted crosswalks, and mobile phone app-based warning systems. *Method:* A systematic review of 42 articles was performed to determine the effectiveness of such interventions. This review found that three types of interventions are currently developed, with differing evaluations. Interventions based on infrastructure tend to be evaluated based on behavioral change. Mobile phone-based apps tend to be evaluated on their ability to detect obstacles. Legislative changes and education campaigns are not currently evaluated. Further, technological development often occurs independently of pedestrians' needs, reducing the likely safety benefits of such interventions. The interventions related to infrastructure mainly focus on warning pedestrians without considering pedestrian mobile phone use, potentially leading to numerous irrelevant warnings and reduced user acceptance. The lack of a comprehensive and systematic approach to evaluating these interventions is also an issue requiring consideration. *Practical Applications:* This review demonstrates that despite significant recent progress surrounding pedestrian distraction, more work is required to identify the most effective interventions to implement. Future studies with a well-designed experimental framework are necessary to compare the different approaches, and warning messages, and ensure the best guidance for road safety agencies.

1. Introduction

More than one-fifth of fatally injured road users do not travel by car, motorcycle, or cycle; they are pedestrians (WHO, 2013). Pedestrians are the most vulnerable road users of the transportation system (Zegeer & Bushell, 2012). They are at a higher risk than drivers and passengers at the time of collision due to their slow movement, vulnerability, and higher impact after the crash (Moudon et al., 2011). Narváez et al. (2019) found that one in five of 1,536 surveyed pedestrians were involved in at least one incident as a pedestrian in the previous five years, and 21% of these incidents resulted in pedestrian injuries. Approximately half of these injuries were severe. In fact, pedestrian fatalities due to a crash in the United States increased 3.9% in comparison with 2019, the highest number since 1989 according to the U.S. Department of Transportation's National Highway Traffic Safety Administration (NHTSA) recently published annual traffic crash data (Transportation, 2022).

Distraction is one of the contributing factors to pedestrian injuries (Pešić et al., 2016; Thompson et al., 2013). Distraction interferes with the decision-making process in a critical situation, and pedestrians fail to notice important visual and auditory information (Lennon et al., 2017) while walking or crossing streets. The most common distractive activities that pedestrians engage in are: using mobile phones, eating, smoking, group talking, drinking, and carrying bags (Hamann et al., 2017; Shaaban et al., 2018; Thompson et al., 2013). Advances in technology see the number of mobile phone users increasing steadily, and statistics show that in 2016 users numbered 4.70 billion and in 2020, 5.22 billion

* Corresponding author.
E-mail address: mdeaysir.arafat@hdr.qut.edu.au (M.E. Arafat).

https://doi.org/10.1016/j.jsr.2022.11.008
0022-4375/

(Chang, 2022). Smartphone ownership varies from country to country, with research indicating 9 out of 10 people use a mobile phone in most developed countries. The ratio varies to 4 out of 10 in developing countries, especially in India where only 24% reported having a smartphone (Silver, 2019). In the United States, 96% of young adults aged 18–29 have a smartphone with only 61% of people over 65 having one. Though there is a technology disparity between different communities and age groups (Center, 2021), studies found mobile phone distraction while walking is increasing significantly (Byington & Schwebel, 2013; Chen et al., 2018; Jain & Gruteser, 2019). And mobile phone use at pedestrian crossings is a growing concern because it contributes to pedestrian-related crashes. An observational study conducted in Melbourne found that 20% of pedestrians used mobile phones when crossing roads (Osborne et al., 2020). Researchers found that more than one quarter of pedestrians are distracted while crossing roads (Basch et al., 2014; Solah et al., 2016; Thompson et al., 2013), whereas other studies indicated the amount was one-third compared to total pedestrians (Basch et al., 2015; Horberry et al., 2019; Scopatz & Zhou, 2016). Some researchers found almost half the pedestrians are distracted while crossing roads (Thompson et al., 2013), and an experiment conducted by Syazwan et al. (2017) in Malaysia found 84.8% of people using a mobile phone while crossing roads, followed by drinking, eating, and reading. In the United States in 2012, 1,500 pedestrians nationwide treated in the hospital were distracted by mobile phone conversations when crossing roads. This number was more than double the figures recorded in 2005 (Scopatz & Zhou, 2016).

The use of mobile phones causes at least three types of distraction for their users: visual distraction (e.g., texting, browsing, gaming, and reading articles), auditory distraction (e.g., listening to music and talking), and cognitive distraction that can be caused by a combination of visual and auditory distraction. Several studies determined the effect of these three types of phone distraction at road intersections (Alejalil & Davoodi, 2017; Courtemanche et al., 2019; Simmons et al., 2020), and it was Jiang et al. (2018) and Thompson et al. (2013) who found that texting was the most significant form of distraction at signalized intersections, followed by talking and listening to music. Visual distractions such as texting or browsing were found to cause a delay for pedestrians to initiate their crossing and a reduced visual scan of the surrounding environment (Tapiro et al., 2018). Texting at the time of crossing intersections reduces the walking speed of pedestrians, increasing their crossing time by 18% (Thompson et al., 2013), and diverts a large proportion of their visual attention from the road to their phone screen (Russo et al., 2018). Pedestrians who use their mobile phone for talking purposes are less likely to show careful behavior than pedestrians listening to music, however, those who listen to music are more likely to look straight ahead, failing to look left and right (Aghabayk et al., 2021). Talking over the phone causes a reduction in attention to the surroundings, and pedestrians only notice vehicles when they are close to them (Davis & Barton, 2017).

Different intervention techniques have been developed in different parts of the world to ensure the safety of mobile phone distracted pedestrians, and these can be categorized as infrastructural, technological, legislative, and public awareness. The legislative intervention includes rules, regulations, fines, and warnings (Osborne et al., 2020). For example, the Utah Transit Authority (UTA) implemented a fine of $100 for distracted walking in 2012, though it failed to be applied statewide (Davidson, 2012).

Public awareness related interventions, on the other hand, include education, posters, and campaigns (Osborne et al., 2020). Safety warning cards were also handed out in New York as part of a safety campaign in 2013 (Seeing Eye People, 2013), and the Ontario police department distributed pamphlets for pedestrians warning against distraction (Law, 2013). In Australia, the Pedestrian Council of Australia conducted a similar safety campaign to find that only a small proportion of road users consider distracted walking as an issue (PCA, 2012), indicating a limited impact on pedestrian behavior (Mwakalonge et al., 2015).

Infrastructure-based interventions are incorporated with road infrastructure, including inbuilt technology such as road marking and flashing lights to be used to warn phone distracted pedestrians. Other types of infrastructure-related interventions enforce pedestrians to use a particular type of intervention, such as pedestrian road separation and safety barriers (Osborne et al., 2020). The infrastructure-based interventions have been trialed and examined separately in different studies worldwide with several countries, for example, trialing solutions based on LED lights at ground level. One intersection in Melbourne was illuminated with flashing lights at ground level, and synchronized with the traffic lights to increase the awareness of distracted pedestrians of their surroundings (Wong, 2017); another study was conducted by Larue and Watling (2021) at a rail crossing with flashing LEDs to ensure the safety of phone distracted pedestrians. Other infrastructure-based interventions include warning signs, reminders at the intersection such as "Heads Up Phones Down" (Barin et al., 2018), and safety road marking (Osborne et al., 2020).

Technology based interventions are mostly related to road user devices. Phone-based technologies have been developed utilizing phone sensors such as a phone camera, accelerometer, gyroscope, magnetometer, and proximity sensor for obstacle detection and warning (Won et al., 2020; Zhuang & Fang, 2020), or obstacle detection with the help of external sensors and warnings using phone apps. Phone apps based on interconnecting with the surrounding infrastructure either utilize vehicle-to-pedestrian (V2P) communication or infrastructure to pedestrian (I2P) communication (Lewandowski et al., 2013; Rahimian et al., 2018) in accordance with critical location identification using GPS or Bluetooth beacon technology. Warning techniques on mobile phone devices themselves include a pop-up window, screen border color change, screen color transparency, audio warning, vibration, and a picture replicating a traffic light on the phone screen (Holländer et al., 2020; Kim et al., 2018; Kim et al., 2015). External road user devices have also been tested including caps, sunglasses (Gruenefeld et al., 2018), and helmets (Marsalia et al., 2016), for obstacle detection and warning.

Given the breadth of interventions currently being developed for distracted pedestrians, it is essential to understand how effective such interventions are and provide guidance on the most effective approach to reduce risky behaviors from phone-distracted pedestrians at road intersections. In this study, a systematic review was conducted of the literature to understand the effectiveness of these interventions and identify current research gaps.

2. Methods

A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) based systematic review (Moher et al., 2009) was conducted to identify studies related to the interventions for phone distracted pedestrians. The PRISMA-based flow diagram is presented in Fig. 1.

2.1. Data sources

The databases used for the systematic review were PubMed, Scopus, Embase, and Web of Science. These databases were selected to provide a wide variety of data from different disciplines related to the topic of interest. English language articles with full text were considered for the review, and the time frame for selected articles was the last 10 years (2012- December 2021).

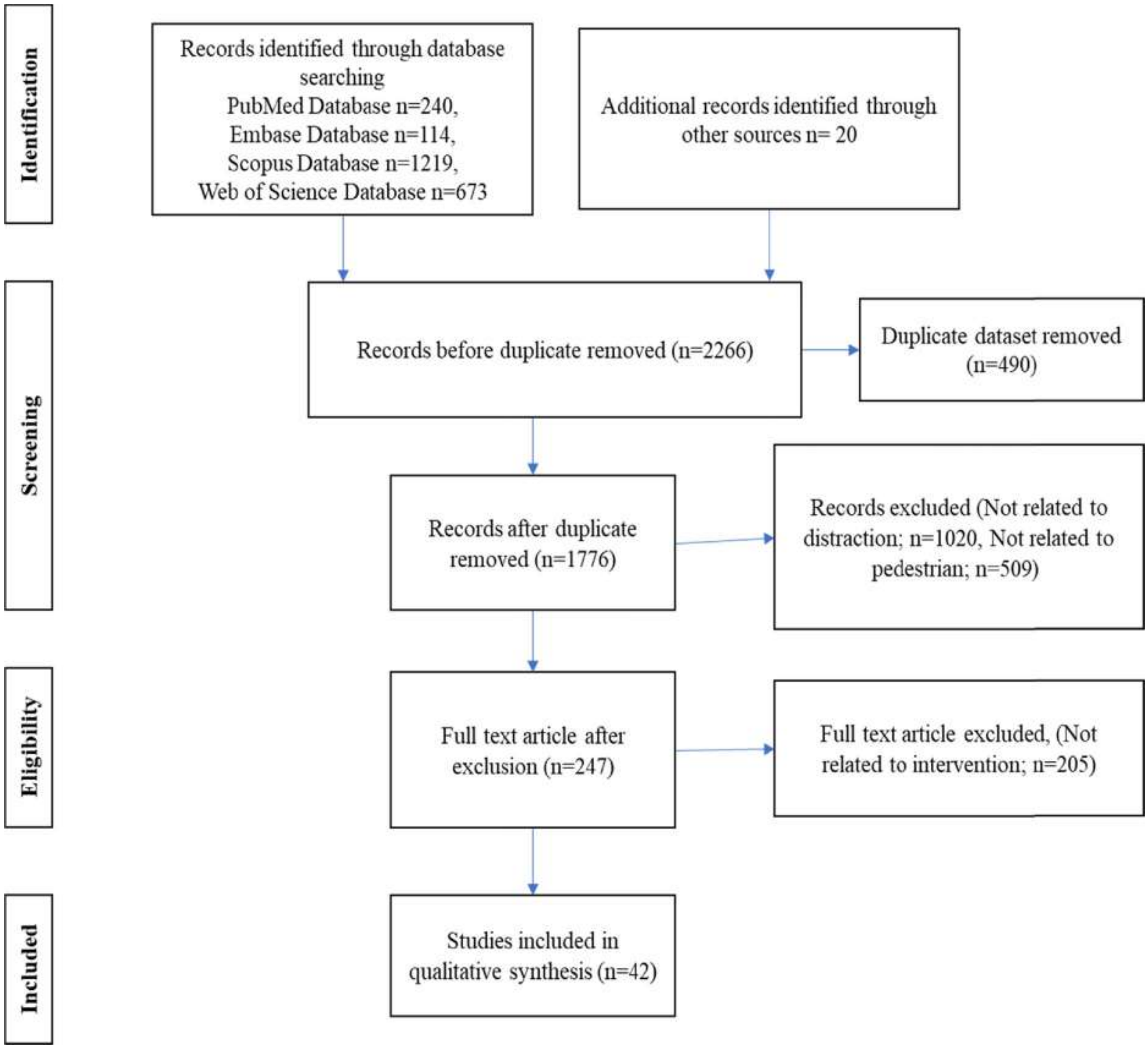


Fig. 1. PRISMA flow diagram identifies the retrieval process for studies that evaluate interventions for distracted pedestrians.

2.2. Study selection

Only articles published in English were considered for this review. Furthermore, the publication date was required to be within the last 10 years (2012–2021). The period was selected based on the emerging influence of mobile phones and their evolving effects on road users.

The database search results, including the title, abstracts, and full text articles were exported to the bibliography software EndNote. Duplicate articles were removed before further screening. The screening and selection were done independently by two authors and then validated by a third author. The final selection of articles was achieved through discussion and consensus between all three authors.

2.3. Eligibility criteria

Articles that only considered mobile phone distraction without considering the effects of any interventions were excluded from the review. Studies not related to pedestrians' (driver, cyclists, biker, rideshare, and scooter) injury model, navigation, and localization were also excluded. Moreover, meta-analyses, literature reviews, author opinions, and articles without full text were excluded (Bruyneel & Duclos, 2020).

In stage one, the search technique started using the filters of the English language and a 10 year period. The search technique comprised three groups of keywords. The first group was related to the type of road user and used the keywords 'pedestrian' and 'vulnerable road user.' The second group related to distraction and included 'phone' [and synonyms], 'personal electronic device,' 'mobile device,' 'hand-held device,' and 'smartphone.' The last group related to interventions and included the following keywords 'intervention,' 'countermeasure,' 'warning alert,' 'prevention,' 'safety,' ' mitigation,' 'education,' and 'media campaign.'

The articles from each group were then combined for further analysis. The final search results were transferred to the EndNote software with the abstract for additional screening. Duplicate articles were removed, and further screening was conducted manually. The abstracts and full text articles were screened to identify the articles not related to pedestrians, distraction, warning, or interventions, or focusing only on navigation, which were excluded.

2.4. Data extraction

The data extraction was performed by two authors and further validated by the third author. The information extracted from the selected studies included: location; experiment area; obstacle detection technology; type of distraction; warning type; distraction identification technique; and participant numbers. The articles related to cyclists, two-wheelers, rideshare and drivers, unrelated to safety or distraction, focusing solely on modeling injury, navigation, or localization, were excluded from the review. A total of 247 articles containing pedestrian, mobile phone distraction, and synonyms were found. Then, full text articles were reviewed for further screening, and 205 articles were removed as they did not include an intervention. Forty-two articles satisfied all the inclusion criteria of pedestrian, distraction, and intervention, and were included in the review.

3. Results

Articles related to the legislation category did not satisfy the inclusion criteria of this review, and just one study related to public awareness was found. Therefore, these two categories are combined for discussion in the remainder of the article. Consequently, the interventions related to mobile phone distracted pedestrians are divided into three categories presented in Fig. 2: infrastructural (related to infrastructure); technological (i.e., road user devices); and legislation and public awareness.

The characteristics of the studies that were selected in this review are summarized in Fig. 3. Fig. 3 provides an overview of where the study was conducted, the number and types of participants included, and the type of distraction and warning.

The interventions pertaining to the infrastructure included in-ground flashing lights, painted crosswalks, and audio signals. Warning methods focused on road user devices categorized into two subcategories: phone app (e.g., audio signal, pop-up window, vibration, screen transparency, and screen lock) and external devices (e.g., helmet, smart cap, and sunglasses).

The intervention effectiveness evaluations identified in the literature focused on three distinctive dimensions: behavioral evaluations, technological evaluations, and user perceptions. Behavioral evaluation focused on the influence of the intervention on pedestrian road crossing behavior such as considering behavior change using the Pedestrian Behavior Scale (PBS) (Larue & Watling, 2021), mobile phone use tendency before and after intervention (Barin et al., 2018), reaction time after warning (Kim et al., 2015), head turns before starting to cross (Chen et al., 2012), and choice of proper safety gap between traffic while crossing (Rahimian et al., 2018). Technological evaluations focused on the technological side of the interventions, such as accuracy of obstacle detection, energy efficiency in terms of phone battery life, the latency of obstacles and distracted pedestrian detection, and timeliness of the warning message. User perceptions gauged user preferences through surveys and questionnaires, as technology acceptance is critical to ensure predicted benefits occur in practice (Kim et al., 2018). These three metrics provide an overall assessment of the efficacy of interventions to reduce risks related to pedestrian phone distraction and are therefore used as the basis for evaluating the impact of these different interventions.

Table 1 displays the 42 studies included in the final analysis. Of these 42 studies, 7 were infrastructure-based interventions and highlighted in blue (two papers by Larue et al. (2021) and Larue and Watling (2021) and were combined because they used the same sample. Thirty-two articles were related to road user devices and presented in variations of yellow. Nineteen studies (highlighted in dark yellow) were related to phone apps using phone sensors to detect danger and provide a warning message. Four studies (highlighted in light yellow) considered external sensors such as ultrasonic and infrared sensors attached with the phone for danger detection and phone apps for warning. Two studies (highlighted in the color cream) utilized beacon technology for positioning the distracted pedestrian and provided a warning on the mobile phone. Three studies (highlighted in beige), and two papers by Rahimian et al. (2018) and Rahimian et al. (2016), were combined because they used the same sample used V2P technology for danger detection and provided a warning on the mobile phone. Four studies (highlighted in green) used external road user devices for danger detection and warning, and one study (high-

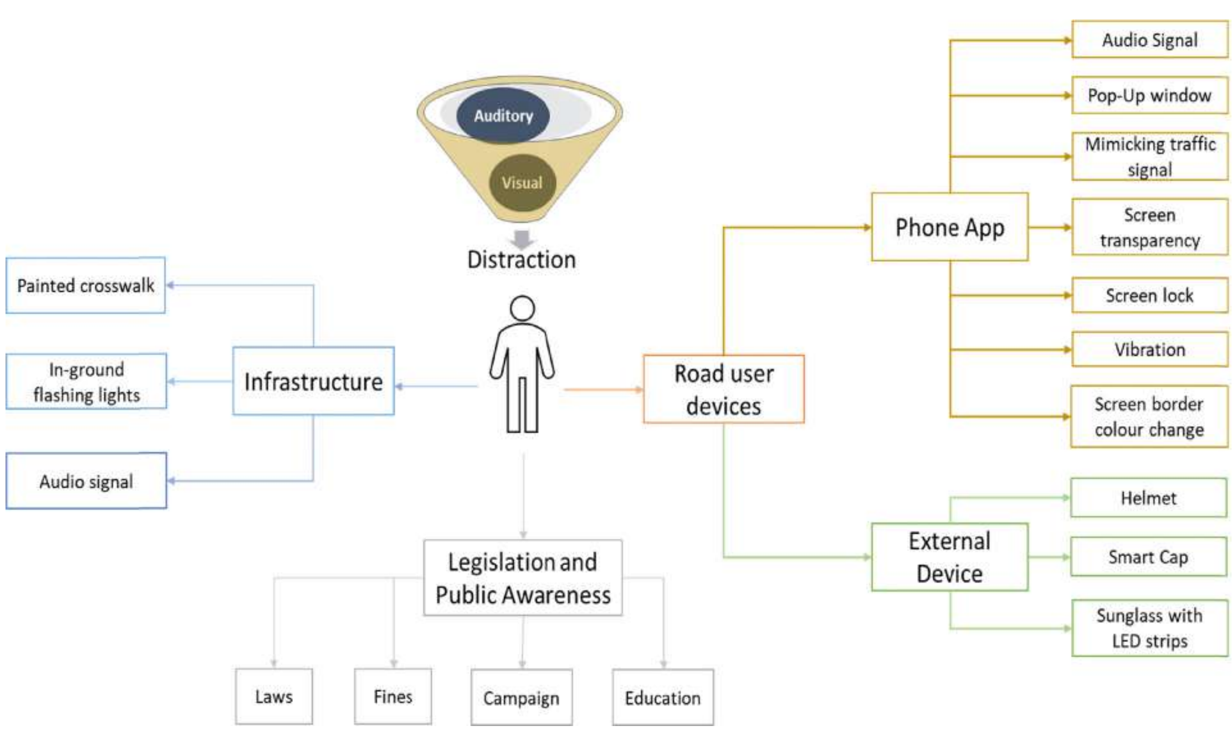


Fig. 2. Key interventions for each category (interventions with blue, yellow, and green borders were considered in different studies, whereas interventions with gray borders at the bottom did not evaluate in any studies except one study related to the campaign).

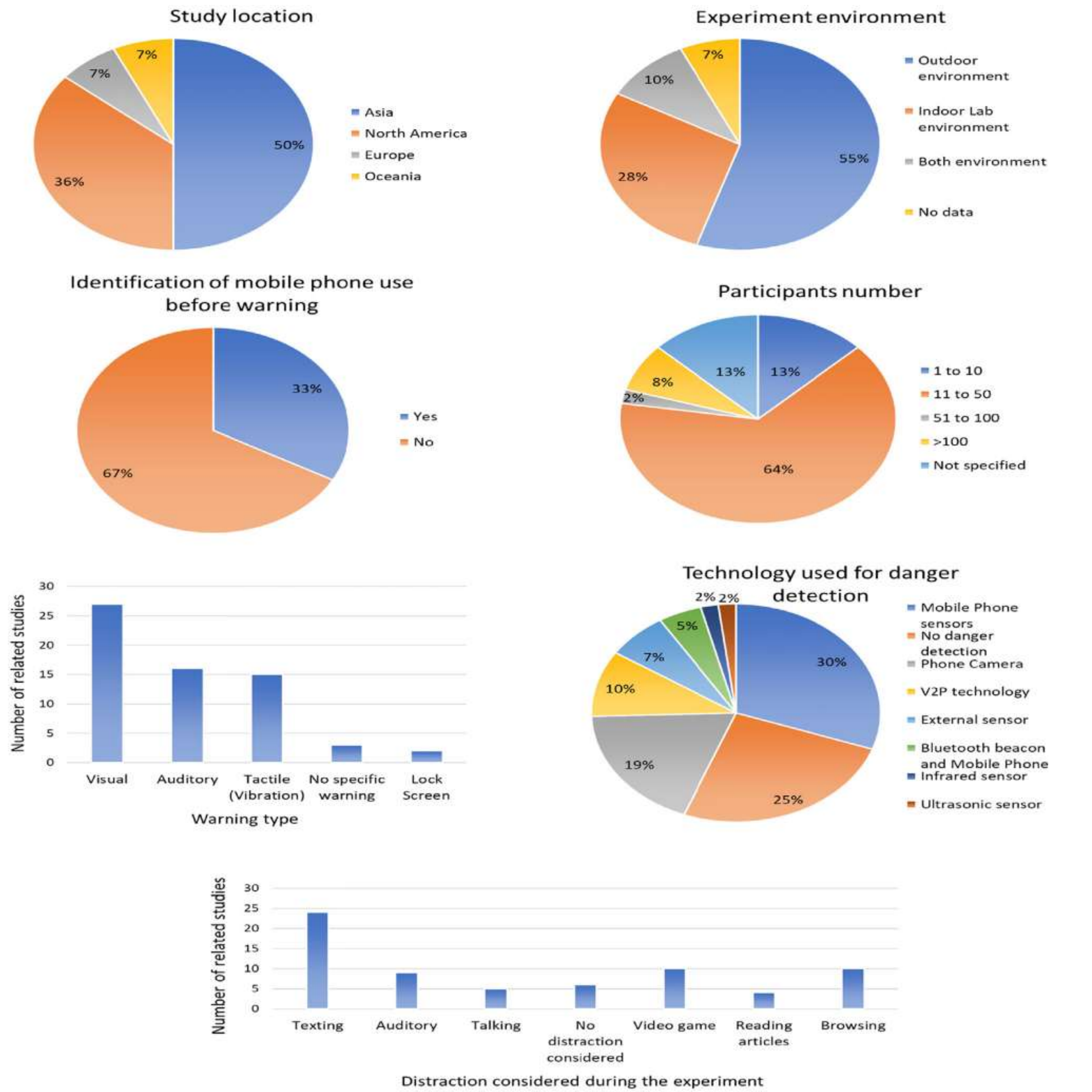


Fig. 3. The pie charts indicate the percentage of studies with location of experiment, experiment environment, participants number, identification of mobile phone use, and technology for danger detection. The bar graphs show (considering duplication of studies in multiple categories) distraction types and warning types considered.

lighted in grey) considered public awareness as an intervention. The effectiveness of interventions was categorized as behavior evaluation, technology evaluation, and user acceptance (Table 1).

3.1. Studies' characteristics

Articles considered in this review were primarily experimental (Fig. 4). Almost 50% of studies were conducted in a laboratory or inside a university campus in a controlled environment. Laboratory studies that utilized a virtual environment to simulate the intersections are categorized as a virtual environment in Fig. 4 (Chen et al., 2012; Gruenefeld et al., 2018; Rahimian et al., 2016; Rahimian et al., 2018; Schwebel et al., 2017; Sobhani & Farooq, 2018). In some cases, treadmills were used to offer a walking experience (Gruenefeld et al., 2018; Kim et al., 2021; Marsalia et al., 2016). And 41% (16 articles) experimented in the real environment outside, such as at intersections, rail crossings, and busy roads. A few studies occurred in both the laboratory and outdoor environments (Jain, 2015; Schwebel et al., 2017), while three studies did not identify the experiment locations.

The study designs considered in this review were within-subjects designs, between-subjects designs, observational studies, and technology performance studies. Within-subjects designs considered control conditions (i.e., without distraction or without intervention), with distraction and with intervention, and compared the effect of intervention in a particular group of people

(Goh et al., 2020; Kim et al., 2021; Larue & Watling, 2021). On the other hand, between-subjects designs considered the same condition as the within-subjects design but considered a different group of people for each condition (Rahimian et al., 2018). Moreover, observational studies considered the pre and post effect of their intervention and, in some cases, follow-up after four months (Barin et al., 2018) or five months (Schwebel et al., 2017). Finally, technology performance studies mostly related to the mobile phone app based interventions and focused on an experiment with a particular technological characteristic (e.g., object detection accuracy, battery life etc.) and determining the technological efficiency (Kang & Han, 2020; Zhuang & Fang, 2020).

3.2. Demographics

Among the 42 studies, 14 studies were conducted with undergraduate, postgraduate, or college students (Kang & Han, 2020; Larue & Watling, 2021; Rahimian et al., 2016; Rahimian et al.,

Table 1
Details of the reviewed studies, including demographics, study setting, distraction and warning types, mobile phone use detection and effectiveness of the intervention (See above-mentioned references for further information).

Intervention types	Study	Sample and demographics	Study design	Distraction types[1]	Warning types[2]	Mobile phone use detection	Effectiveness of interventions		
							Behavior evaluation	Technology evaluation	User acceptance
Infrastructure-based	Larue and Watling (2021)	N=34, mean age: 33.6	Within-subjects design	**V, A**	**V, A, B**	No	√	-	√
	Kim et al. (2021)	N= 38, mean age 23.8	Within-subjects design	**V, A**	**V**	No	√	-	-
	Larue et al. (2020)	N=24, mean age 30.4	Within-subjects design	**V**	**V**	No	√	-	-
	Barin et al. (2018)	N=11,533	Observational study (pre-post intervention)	**V, A**	**V**	No	√	-	-
	Sobhani and Farooq (2018)	N= 42, age range 18-45	Between-subjects design	**V**	**V**	No	√	-	-
	Sobhani et al. (2017)	N= 20, age-range 20-45	Within-subjects design	**V**	**V**	Yes	√	-	-

(continued on next page)

Table 1 (*continued*)

	Violano et al. (2015)	N= 1,362	Observational study (pre-post intervention)	**V, A**	**V**	No	√	-	-
Obstacle detection using built-in phone sensors and warning from a mobile app.	Kang and Han (2020)	N= 20, age-range 22-32	Technology performance study	**V**	**V, A, T**	No	-	√	√
	Won et al. (2020)	N= 1	Technology performance study	**V**	**NS**	Yes	-	√	-
	Zhuang and Fang (2020)	N= 20	Technology performance study	**V**	**V, T**	Yes	-	√	-
	Holländer et al. (2020)	N= 24, age range 19-36	Within-subjects design	**V**	**V**	Yes	√	-	√
	Malathy et al. (2019)	N/A	NS	**NS**	**NS**	No	-	√	-
	Jain and Gruteser (2019)	N= 9	Technology performance study	**V**	**NS**	No	-	√	-
	Sun et al. (2019)	N= 100	Technology performance study	**NS**	**T**	No	-	√	-
	Ou et al. (2019)	N= 25, age-range 18-30	Technology performance study	**V**	**V**	No	-	√	-
	Kang et al. (2019)	N= 20, age-range 21-30	Technology performance study	**V**	**V, A, T**	No	-	-	√
	Tung and Shin (2018)	N= 21	Technology performance study	**NS**	**A**	Yes	-	√	√
	Kim et al. (2018)	N= 74, age-range 20-39	User study	**V**	**V, A**	No	-	-	√

Table 1 (*continued*)

	Li et al. (2018)	N= 16	Technology performance study	**V**	**T**	Yes	-	√	-
	Hwang et al. (2016)	N= 5, mean age 25.6	Technology performance study	**NS**	**NS**	Yes	-	√	-
	Wang et al. (2016)	N=51	Technology performance study	**NS**	**NS**	Yes	-	√	-
	Tang et al. (2016)	N= 1	Technology performance study	**NS**	**NS**	No	√	√	-
	Zhou (2015)	N= 20	Within-subjects design	**V**	**V**	Yes	-	√	-
	Foerster et al. (2014)	N= 21, mean age 29.4	Technology performance study	**V**	**A, T**	No	-	√	-
	Chen et al. (2012)	N= 24	Between-subjects design	**V, A**	**NS**	No	√	-	-
	Wang et al. (2012)	N/A	NS	**A**	**A, T**	Yes	-	√	-
Obstacle detection using outside sensors and warning from a mobile app.	Xia et al. (2019)	N/A	NS	**A**	**A, T**	No	-	√	-
	Liu et al. (2016)	N= 14	Technology performance study	**V**	**A**	Yes	-	√	-
	Wen et al. (2015)	N= 47	Technology performance study	**V**	**A**	Yes	-	√	-
	Jain et al. (2015)	N= 21, age-range 20-40	Technology performance study	**NS**	**V**	No	-	√	-
Position detection using an outside sensor (Bluetooth bacon) and warning from a mobile app.	Schwebel et al. (2021)	N= 437, age-range 17 years or more	Cross over design and Observational study (pre-post intervention)	**V, A**	**V, A, B**	Yes	√	-	√
	Goh et al. (2020)	N= 73, age-	Within-subjects design	**V**	**V, A, T**	Yes	√	√	√

(continued on next page)

Table 1 (*continued*)

		range 20-33							
Danger detection using V2P technology and warning from a mobile app.	Rahimian et al. (2018)	N= 48	Between-subjects design	**V**	**A**	No	√	-	-
	(Liu et al., 2015)	N/A	NS	**V, A**	**V, A, B**	Yes	-	√	-
	Kim et al. (2015)	N= 29	User study	**V**	**V**	No	√	-	√
Obstacle detection and warning using external devices.	Gruenefeld et al. (2018)	N= 8, age-range 22-31	Technology performance study	**V**	**V**	No	-	-	√
	Kalra et al. (2017)	N/A	NS	**NS**	**T**	No	-	√	-
	Marsalia et al. (2016)	N= 27, age-range 20-39	Within-subjects design	**V, A**	**T**	No	√	√	-
	Kumar et al. (2015)	N= 10	Technology performance study	**V**	**T**	Yes	-	√	-
Public awareness	Schwebel et al. (2017)	N=219, mean age 22.6	Observational study (pre-post intervention)	**V**	**V**	No	√	-	-

[1]Distraction types considered were categorised as visual, auditory and not specified and indicated as "V", "A", and "NS" respectively.
[2]Warning types are "V" for visual, "A" for auditory, "T" for tactile, and "NS" for not specified (i.e., studies mentioned that an alert was provided without further details), and "B" for both types (i.e., visual and auditory).

2018; Sobhani et al., 2017). Nine articles did not mention participant type or demographic information but indicated the sample size. Five articles did not mention the participants' number and type (Kalra et al., 2017; Liu et al., 2015; Malathy et al., 2019; Wang et al., 2012; Xia et al., 2019).

Eight studies considered only young adults (i.e., 18–30 years). Ten studies did not include age as a participation criterion (Kang & Han, 2020; Larue & Watling, 2021). One study included eight children participants and four young adults (Barin et al., 2018). Twenty-three studies did not report participants' age.

3.3. Distraction types

Different types of distraction were considered in the various studies. Visual distraction was the most prominent (e.g., texting, browsing, playing a video game, and watching a video; Goh et al., 2020; Hwang et al., 2017; Larue & Watling, 2021; Schwebel et al., 2017), followed by auditory distractions (e.g., listening to music and responding to an auditory task; Barin et al., 2018; Larue & Watling, 2021; Larue et al., 2020; Ou et al., 2019), and talking (Barin et al., 2018; Violano et al., 2015). However, stud-

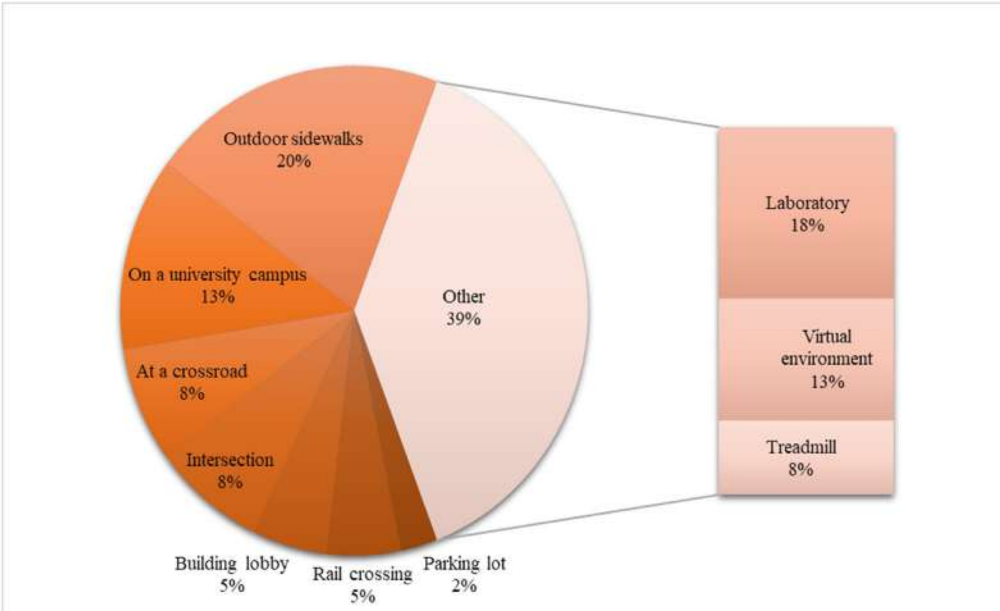


Fig. 4. Percentage of studies conducted in different experimental environments. Note that percentages are not provided as numbers for multiple parts of the figure. Can you fix it. Use the word document attached in the email.

ies focused on the intervention techniques mostly considered visual or auditory distraction when experimenting rather than considering cognitive distraction. For that reason, cognitive distraction was considered out of scope for this review.

3.4. Effectiveness

There are discrepancies in the level of assessment of the different interventions. Infrastructural interventions mainly focus on behavioral evaluation (e.g., Larue et al., 2021; Sobhani & Farooq, 2018), while technology based interventions primarily focus on evaluating the technology itself (e.g., Sun et al., 2019; Won et al., 2020). Fig. 5 presents the evaluation types currently reported in the literature for the different types of interventions. Details for each intervention type are then presented in the following subsections.

3.4.1. Infrastructure-based interventions

The evaluation of infrastructure-based interventions focused on visibility in terms of easily noticed by pedestrians or not, and behavior change. Infrastructure-based interventions can be categorized into two types. The first impacts pedestrian behavior by providing a warning reminder such as in-ground flashing lights or safety road marking. In contrast, the second ensures pedestrian safety through physical structure changes such as road separation or safety barriers. Since no studies used this latter type of intervention, it was omitted from the scope of the study. Larue and Watling (2021) found that in-ground LEDs increased the frequency of checking rail tracks from 70% (without LEDs) to 78% (with LEDs) during auditory distraction, and established that the LED detection rate was 90% for distracted pedestrians, and easily recognizable in a horizontal plane, in-ground, compared to the vertical position of wall-mounted, while walking (Larue et al., 2020). However, Kim et al. (2021) found that the use of LEDs decreased participant detection rates of identifying LEDs when visual cues showed up by 74.1% and increased the reaction time, or response time, after noticing the LEDs from 0.90 s to 1.15 s while experimenting with and without LEDs. Sobhani et al. (2017) indicated that pedestrians took a longer time to choose safer gaps for crossing safely and took more time to check traffic at unsignalised intersections with in-ground LEDs.

Barin et al. (2018) found that painted crosswalks with safety messages initially reduced phone distraction, but the effects were short-term, with behavior returning to normal after four months. Violano et al. (2015) indicated that painted crosswalks with a safety message were not adequate for changing the distracted behavior of the mass community.

3.4.2. Device-based interventions

3.4.2.1. Mobile phone. Phone based interventions focused mainly on technological evaluation of obstacle detection accuracy, battery consumption due to use of the app, and timeliness of warning and obstacle detection, with limited research on behavior change and user acceptance. Fig. 6 indicates the obstacle detection accuracy for the mobile phone app-based interventions and used sensors for detecting the obstacles. Fig. 6 also indicates that using a camera sensor is more prominent for detecting obstacles than other types of sensors, with an average obstacle detection accuracy of more than 80%. Kang et al. (2019) developed a phone camera based obstacle detection technique, identified obstacles using the feature point extraction method, and reported their app could detect static obstacles with the precision of 91%. Jain and Gruteser (2019) also developed a phone camera based obstacle detection method using the material recognition technique called TerraFirma and warned distracted pedestrians when they entered roads from the sidewalks, and found it could detect sidewalks with a false positive rate of 1%. Sun et al. (2019) also developed a phone camera-based sidewalk detection technique capturing the sidewalk images, compared them with the wide-view dataset called PESID and identified the user movement using the phone sensors of a gyroscope, gravity sensor, and GPS to indicate their system could detect sidewalks and static objects with an accuracy of 70%. The app developed by Tung and Shin (2018) worked using phone sensors and followed the module of the motion detector (accelerometer for user movement), acoustic detector (speakers and microphones for measuring the distance between obstacle and user), visual detector (rear camera for obstacle detection) and the fusion algorithm (vibration motor for increasing the efficiency), established these tasks were all interrelated. Malathy et al.

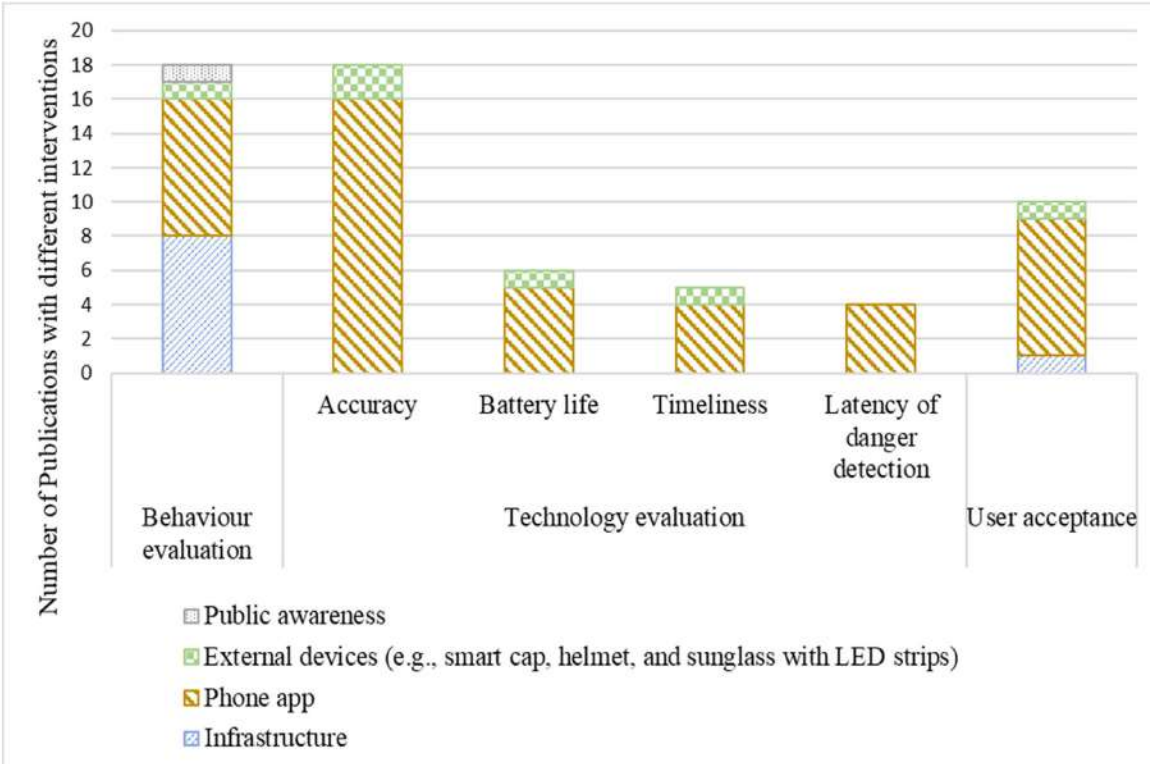


Fig. 5. Effectiveness of interventions with three distinctive dimensions examined in different studies based on intervention types.

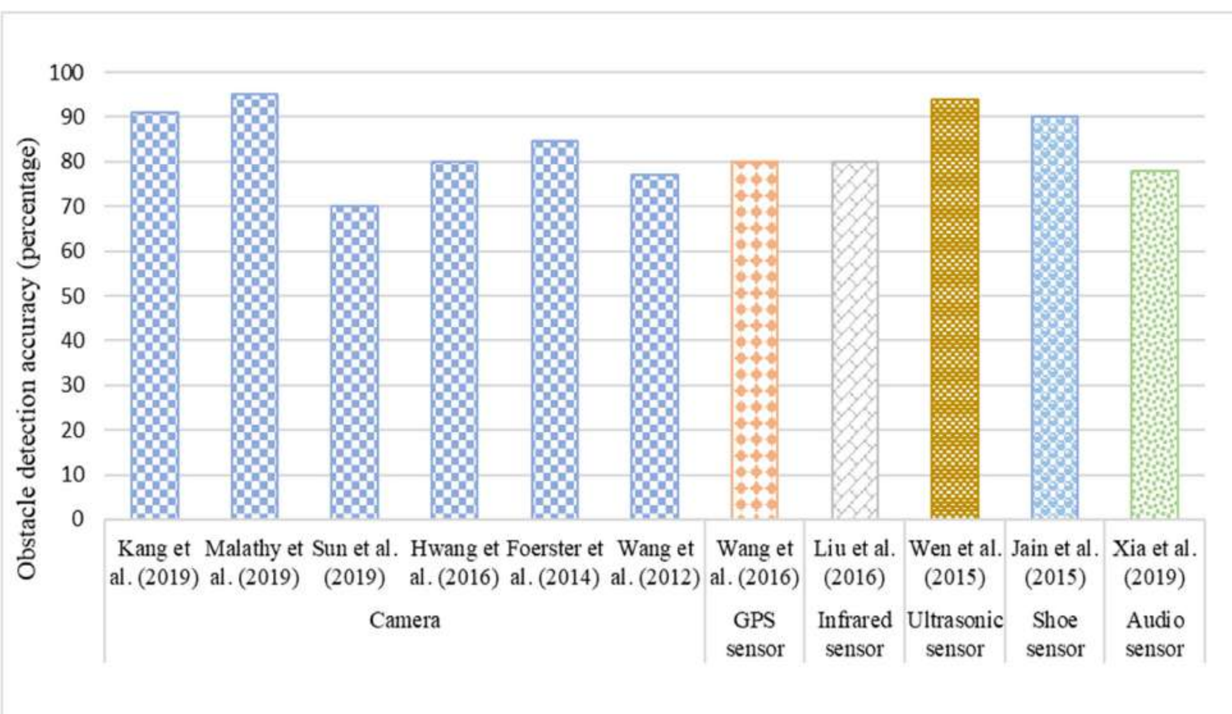


Fig. 6. Obstacle detection accuracy considered in mobile phone app based studies.

(2019) developed an app with an accuracy of 95%, based on the Tung and Shin (2018) framework to identify static obstacles. A mobile phone camera based sidewalk detection technique developed by Tang et al. (2016) captured the image of the sidewalk using a phone camera, classified those images using the KNN and Bayes Classifier, and found that detection accuracy increased with the frequency of the detection algorithm where if the algorithm runs 2 or 3 times, detection accuracy will increase simultaneously. Hwang et al. (2016) also developed a sidewalk detection technique using the rear phone camera, a feature extraction technique for image classification, and context recognition and that technique was used to identify sidewalks, roadways, and intersections and indicated their technique could detect obstacles at 80% accuracy. Foerster et al. (2014) also examined the phone's rear camera for obstacle detection and analyzed the image background (i.e., smartphone-based pixel scanning process) to determine the distance between the obstacle and the user before sending an alert. They found their system could identify static obstacles with an accuracy of 85%. Wang et al. (2012) developed a technique using the phone's rear camera to capture images of the front and rear views of vehicles and classified tying with the built-in dataset using the machine learning approach. The phone app then provided a warning based on vehicle detection and identification of auditory distraction. They indicated their technology could detect the front and rear view of cars with a true positive rate of 77% and 76%, respectively. Wang et al. (2016) considered a GPS sensor for identifying overpass, underpass, and traffic lights using the crowdsensing technique. Chen et al. (2012) developed a web-based vulnerable location identification app that worked with the help of Google Maps, API and Google Web Toolkit SDK to identify the intersection locations and risk factors (i.e., previous accident data) before providing a warning.

In some studies, obstacles were detected using external sensors and warnings were provided to the mobile phone. The external sensors of a headset mounted audio sensor used by Xia et al. (2019) could detect moving cars with 62–78% accuracy. Wen et al. (2015) developed mobile phone app-based intervention used a three-step method: first, identified pedestrians' phone use by checking the phone screen status, whether the phone screen was on or off; the second step was danger detection, conducted using an ultrasonic sensor; and finally, a warning based on user awareness detection, phone holding pattern, and walking speed. Liu et al. (2016) followed this same technique, instead using an infrared sensor rather than the ultrasonic sensor. Results showed an ultrasonic sensor could detect obstacles (i.e., sudden change of ground) with an accuracy of 94%, while an infrared sensor could detect sudden ground change with an accuracy of 80%. Jain et al. (2015) found their technology using some obstacle detection techniques with shoe sensors could detect sidewalks with an accuracy of 90% and a false positive rate of 0.7%.

Some studies used the timeliness of the app warning system to alert distracted pedestrians as a dimension for technological evaluation identifying distracted pedestrians and latency of warning. Schwebel et al. (2021) used Bluetooth beacon technology for pedestrian position detection, and the K-Nearest Neighbour (KNN) algorithm detected pedestrian phone use before delivering a warning. Goh et al. (2020) also developed a situational-awareness-based warning technique using a Bluetooth beacon for identifying the pedestrian location and collected other sensor data (e.g., accelerometer and barometer), and checked the phone status before warning. Won et al. (2020) improved their GPS location identification module using a Hidden Markov model (HMM), identified the pedestrian phone viewing data from the accelerometer reading, and developed communication between phone users and cars using Wi-Fi direct. They found their app could detect distracted pedestrians with an accuracy of 90% and provided a warning with an error rate of 1.6 sec. Zhuang and Fang (2020) also used GPS sensors to identify the vulnerable locations of the pedestrian, accelerometer and gyroscope sensor were used to identify the pedestrians' phone orientation, and proximity sensor data were used to measure the distance between an obstacle and user before warning. They also indicated their app could identify smartphone zombies with more than 90% accuracy. Li et al. (2018) developed a phone camera based facial recognition technology that could identify distracted pedestrians who looked at their phone for more than 6 secs, with a true positive rate of 91%. Tang et al. (2016) developed a system to alert distracted walkers with an average accuracy up to 98%. Zhou (2015) developed a warning technique considering the pedestrians' gait pattern using accelerometer and gyroscope sensors, and a warning was provided based on the phone usage detection and found their intervention technique could warn distracted pedestrians with a false negative rate of less than 3%; and Wen et al. (2015) showed their app could reduce unnecessary warning alarm by 90%.

Phone battery consumption rate was also used in some studies to evaluate phone app effectiveness. The energy efficiency of the app depends on phone sensors (e.g., accelerometer, gyroscope, magnetometer, GPS, and proximity sensor), cameras, microphones, speakers, and system computation (Abdesslem et al., 2009). Won et al. (2020) developed an app with an energy efficiency of 52%, while Zhuang and Fang (2020) claimed their system consumed 1% of battery energy per hour. Tung and Shin (2018) experimented with an app and found it consumed 8% of battery energy per hour. Liu et al. (2016) and Wen et al. (2015) experimented with infrared and ultrasonic sensors, respectively, in conjunction with the mobile phone app and found these systems only consumed 20% of battery energy compared to other technologies based on the mobile phone.

Several warning techniques, such as a pop-up window, screen border color change, screen color transparency, audio warning, vibration, and replicating a traffic light on the phone screen were used in different studies. The suitability and usability of these warning categories depend on participant perceptions. Holländer et al. (2020) found that sidebar warning techniques achieved a success rate of 85% after 144 trials followed by a notification message. Kim et al. (2018) found that 74% of participants considered flashing screen borders unobstructive, followed by vibration (67%). Kim et al. (2015) used Dedicated Short-Range Communication (DSRC) and 4G-LTE for vehicle to pedestrian communication before providing warning to distracted pedestrians and indicated the red warning bar was more effective than the yellow one, and participants preferred 100% or 50% transparent screen with warning messages rather than no transparency.

Alternatively, Kang et al. (2019) found that auditory (alert sound) and tactile (vibration) warnings were the most effective for visual distraction. Schwebel et al. (2021) also showed that using mobile phone based interventions reduced distraction by 64% initially, which increased again during the post-intervention stage without warning. Rahimian et al. (2018) and Rahimian et al. (2016) found that auditory warnings on the phone reduced the time crossing a road, helped to find a suitable gap between traffic to cross a road, and increased the mean gap chosen time in an unsignalised intersection.

3.4.2.2. External device-based interventions. External device based intervention techniques work on using external devices for obstacle detection and warning. Marsalia et al. (2016) developed an intervention technique with a vibrating helmet, with results showing the false alarm rate was 30.3% for providing a warning alarm. Kalra et al. (2017) developed a solar cap that could detect objects effectively and provided a warning, while Kumar et al. (2015) also developed a smart cap that could detect obstacles with a 96% efficiency and motor power consumption while working was 471 mW. Gruenefeld et al. (2018) developed sunglasses with LED strips and found participants preferred moving LEDs compared with static LEDs.

3.4.3. Legislation and public awareness

Countermeasures related to changes in legislation and public awareness campaigns are not mentioned in the literature. Only Schwebel et al. (2017) conducted an experiment considering a public awareness campaign using face-to-face and electronic communication and promoted the slogan "Pocket and Walk It." They also created a virtual reality environment to explain the risk of distracted walking. The post-intervention effect was evaluated after

10 weeks, and six months after the campaign, through a survey. They found it caused some change in behavior intentions in self-reported cases. However, at the community level, it did not cause significant behavior change. None of the studies considers the legislation and public awareness as a mode of intervention, and it is currently not possible to report on their effectiveness.

3.5. Effect size

Effect sizes of well-designed studies were considered using Cohen's d statistic. This section reports effect sizes as reported in the literature, or as estimated from the information reported in the literature. The directionality of the effect is recorded in the sign of Cohen's d, with positive values for positive effects of the intervention, and negative values otherwise.

Larue and Watling (2021) used a within-subjects design to evaluate the effectivity of in-ground flashing lights, audio warnings, and both for mobile phone distracted pedestrians in rail crossing environments using Technology Acceptance Model. They found that in-ground flashing lights have a large effect size (1.22) compared to the audio warning (0.70) and both together (0.22) for actual behavior change (i.e., look for the train before starting crossing when distracted). Kim et al. (2021) also used a within-subjects design. They found that visual distraction reduced the detection of visual cues with a large effect size of negative value 1.43 and increased the reaction time of detecting the visual cues with a large effect of negative 1.28. Larue et al. (2020) conducted a laboratory-based study using a within-subjects design and indicated that the detection of LED lights was higher in walking conditions than in standing conditions when visually distracted, with a large effect size of 1.74. Furthermore, Sobhani and Farooq (2018) conducted a between-subjects design study in a virtual reality environment and found distracted pedestrians with an alert wait more, compared to the distraction-only condition, with a very small effect size (0.004) and decreased the number of head turns with an effect size of negative 0.2. Two of the device-based studies also considered effect size. For example, Rahimian et al. (2018) conducted a between-subjects design in a virtual reality environment considering an unsignalized intersection and found that the waiting time before start crossing increased from the distraction condition with a large effect size of 0.83, as well as the mean gap taken time also increased with an effect size of 0.8. However, road crossing time decreased with a small effect size of 0.3. Chen et al. (2012) also conducted a between-subjects design in a virtual reality environment considering a signalized intersection. They found that waiting time and number of head turns before starting crossing were higher in warning conditions than in the only distraction condition with an effect size of more than one.

4. Discussion

This systematic review aimed to consolidate the literature's current findings regarding the effectiveness of the interventions currently developed, tested, and implemented to mitigate the risks of mobile phone use as a pedestrian. This review found three types of interventions: interventions based on the deployment of new infrastructure; interventions based on technological capabilities of the devices that pedestrians have when they walk; and legislative and awareness campaigns. Overall, the evaluation of such interventions is highly variable between studies, limiting the ability to evaluate the effects of these different interventions on multiple dimensions and limiting the ability to compare such effects. Studies evaluating infrastructure-based interventions focused mainly on behavior change, whereas technology-based interventions focused on technological capability. Legislation and public awareness approaches have not been largely evaluated.

4.1. Limited evaluations of the effectiveness

This review study found that the interventions based on mobile phone apps were evaluated in terms of their ability to detect obstacles. Such research primarily reported detection accuracy, often failing to consider effects on pedestrian behavior. On the other hand, infrastructure-based interventions focused mainly on behavioral change, with limited focus on technology accuracy. Consequently, each intervention is evaluated on a particular dimension, but without a comprehensive evaluation of its effects, or the likelihood that an intervention would be viable as an approach to reduce pedestrian distraction risk.

This review has also identified an extensive variability in assessing intervention effectiveness, even within a particular dimension. Mobile phone app related interventions primarily focused on phone cameras for obstacle detection and found comparatively higher efficiency. The outcomes from these techniques (Sun et al., 2019; Wang et al., 2012) largely depend on the image database and consume more battery energy due to capturing images without considering object type; whereas phone cameras for material recognition (Jain & Gruteser, 2019) and object color and texture identification (Hwang et al., 2016) can be more energy efficient. Despite such promising results, Kim et al. (2018) highlighted that the vision-based approach using a phone camera is challenging for detecting the distance between the obstacle and the user, suggesting that such approaches may face significant challenges before successful deployment in the field. While the V2P technology seems promising, it is also challenging to implement because it largely depends on vehicle properties (such as speed, distance, and communication devices, and technological development) in terms of introducing new sensors or developing autonomous vehicles. The sensor-based technologies can be implemented effectively when we move entirely to the automated vehicle era.

In some cases, the mobile phone app-related technologies used GPS and Bluetooth beacon technology to identify vulnerable locations. While these two approaches can be used for providing the functionalities necessary for a distracted pedestrian intervention, they show different performances in terms of energy efficiency, cost, and localization precision. With only a few milliseconds of transaction latency, Bluetooth beacon technology can identify pedestrian location immediately and precisely with a low power consumption. On the other hand, GPS provides less precision for the localization of the pedestrian and a higher phone battery consumption rate, which suggests that that option may lead to missed or delayed warnings. It is therefore unlikely that a technical solution based on the pedestrian phone would be viable, without the addition of some infrastructure at intersections. Overall, while such technologies look promising, only a few studies exceeded 90% accuracy, which arguably would be the minimum for an intervention to be perceived as useful by pedestrians.

Given that the various interventions were evaluated based on different parameters, it is challenging to compare the effects of different interventions and identify the most promising approach. It is crucial that evaluations can be compared given the large number of interventions currently developed and trialed. This range of interventions covers the technology being used but also the type of message provided to road users. In order to consolidate the research, it is necessary to develop a set of metrics to evaluate the effectiveness of interventions. Such metrics need to be related to the risky behavior needing to be mitigated. However, the parameters indicated in different studies can be considered the foundation for the future modification and development of such interventions for an evaluation taking into account the multiple dimensions required when evaluating an intervention.

Finally, the intervention techniques related to legislation and public awareness campaigns may have been implemented in different parts of the world. However, the effects of implementing such countermeasure techniques were only considered in one study conducted by Schwebel et al. (2017), except that there has been no scientific review regarding the impact of such interventions for mobile phone distracted pedestrians. Such an approach is widely used in road safety (Mwakalonge et al., 2015; Osborne et al., 2020) and will likely result in positive changes, as previously shown by the implication of fines having a positive effect on distracted pedestrians (Mwakalonge et al., 2015). In addition to publicity campaigns (Savolainen et al., 2011), road safety campaigns and social advertisements showed positive effects for phone distracted pedestrians. Wundersitz and Hutchinson (2011) and Schwebel et al. (2017) also found behavior change in self-reported cases after conducting a public awareness campaign, making it critical to evaluate such interventions to understand whether these traditional approaches are also relevant to the emerging issue of pedestrian distraction.

4.2. Limited scope of evaluations

The age group reflected in most articles was the young adult population aged 18–29 years. Very few studies considered other age groups. While mobile phone distraction emerged as an issue with younger adults, such devices are now pervasive and used by all age groups, including when walking as a pedestrian (Hou et al., 2021; Lennon et al., 2017). This suggests a need for research to expand the age of participants included in evaluations of interventions for distracted pedestrians.

Further, the majority of the studies considered a limited number of participants (i.e., 20–40 participants) and, in some cases, experimented with single-digit (i.e., 1 to 10) participants (Gruenefeld et al., 2018; Hwang et al., 2016; Jain & Gruteser, 2019; Tang et al., 2016; Won et al., 2020). Almost half the experiments were conducted on a university campus, either in the laboratory or controlled outdoor area. The participants were primarily undergraduates, postgraduate students, or staff. Overall, such conditions are unlikely to be representative of a real-world scenario, and further research with ecological validity is necessary to confirm whether the positive effects found in the literature are likely to translate into improved behavior and reduced risk in the field.

The distractions considered in all the studies were either visual (texting, browsing, playing a video game) or auditory (listening to music, talking) or, in some cases, both at the same time (Schwebel et al., 2021) and interventions based on the distraction type. The infrastructure-related warning system primarily suggested either visual (flashing light) or auditory warning. In the case of mobile app warning systems, alerts could be visual (pop up window, border flashing), auditory (alert signal, siren) or tactile (vibration). If the distraction occurs due to visual activity, it indicates visual intervention and the same for auditory distraction. Studies mostly considered the effect of auditory warning at the time of visual distraction or vice versa, whereas Liu et al. (2015) and Schwebel et al. (2021) considered both types of distraction simultaneously. Liu et al. (2015) evaluated both types of distraction simultaneously and implemented the visual and auditory warning together to compare the effectiveness of the combined effect, and Schwebel et al. (2021) measured the level of distraction and provided both warnings at the same time if both types of distraction were present. Despite this, none of the studies considered the impact of warnings in cases where both types of distraction occurred.

The cost of implementing these interventions in the real world must be determined. However, no researcher has done so except Xia et al. (2019) who developed a device called PAWS, a wearable headset modified with a BLE transceiver, mics, amplifiers, and regulators to identify the surrounding car horn and tire noises and alert distracted pedestrians. The estimated cost was US$18–20. Although there is limited research about cost, the literature does identify that infrastructure-related interventions (e.g., in-ground flashing lights, road marking) will cost more compared to mobile phone app based interventions because these types of interventions are based on a mobile phone's built-in sensors (e.g., camera, accelerometer, gyroscope and proximity sensor; Won et al., 2020; Zhuang & Fang, 2020). Costs may vary due to use of external sensors, vehicle-to-pedestrian communication devices, and pedestrian-to-infrastructure communication devices. Liu et al. (2016) considered infrared sensors, and Wen et al. (2015) considered ultrasonic sensors for obstacle detection or a Bluetooth beacon for identifying critical locations (Schwebel et al., 2021) before providing a warning to distracted pedestrians. These external sensors will incur extra costs rather than using built-in phone sensors. External devices such as helmets, smart caps, and sunglasses with LED strips will also cost more depending on material type and additional equipment requirements. Therefore, it can be concluded that although none of the studies considered the economic perspective, from a general point of view, it can be predicted that infrastructure-related interventions will cost more to implement compared to phone app and external device-based interventions.

The studies considered for the review were either conducted by engineers, computer scientists, or behavioral scientists. In the majority of the cases, research conducted by engineers or computer scientists focused on the technological development of the intervention techniques, whereas behavior scientists focused on behavior evaluation. However, the common goal is to introduce an intervention technique to ensure distracted pedestrian safety. That goal can be achieved by conducting multidisciplinary research and collaboration between different disciplines, methods, and theories. The collaboration between different groups of people from different fields can make discoveries or find new solutions to the problem at stake (Proctor & Vu, 2019).

4.3. Relevance of the evaluations to distraction risks

This review has identified gaps between the current research on interventions for distracted pedestrians and the effects of distraction as reported in the literature.

Initially, significant technological improvements have occurred in intervention techniques. The research is then incremental from phone sensors to V2P technology and, finally, Bluetooth beacon technology. We are moving towards a new public health challenge that was not an issue two or three decades ago. With the advancement of technology, numerous interventions have been introduced, but sometimes such technology fails to consider the type of risks that need to be targeted. This results in the development of interventions based on technical capability rather than pedestrian needs. For example, obstacle detection, pedestrian phone use status check, and pedestrian movement identification are particularly important to reduce false alarms. However, focusing on pedestrian behavioral change after implementing such an intervention is essential rather than purely focusing on technological development. This is particularly apparent in interventions being developed for detecting obstacles. Indeed, such evaluations often focus on static obstacle detection (Kang & Han, 2020; Kang et al., 2019; Malathy et al., 2019; Sun et al., 2019), sidewalk detection, and sudden ground-level change detection (Ashqar et al., 2019; Goh et al., 2020). Few studies considered the detection of moving objects such as moving cars (Wang et al., 2012; Xia et al., 2019). That also indicates that the majority of the study considers distracted pedestrians on sidewalks, or in off-road environments, rather than on roadways. Even when trying to detect vehicles, such moving vehicle detection approaches focus on detecting the front and back of vehicles. However, it is necessary to acknowledge that a large portion of the risk to a pedestrian is from a vehicle approaching the pedestrian perpendicularly at road intersections. Most of the time, such evaluations focus on detecting the vehicle, failing to consider the vehicle's location, its trajectory compared to the pedestrian, and the likelihood of a crash occurring. Overall, such approaches are found effective in simple conditions compared to what would occur on the road and are therefore prototypes rather than technologies ready for deployment.

Studies related to the interventions considered two types of distraction such as visual and auditory during the experiment. However, it is also necessary to evaluate the effect of the intervention during cognitive distraction activities, which reduce situational awareness behavior (Erkan, 2017), as required intervention techniques may be different. The findings related to the intervention could also be different if the distraction task (e.g., visual or auditory) demands more or less cognitive attention.

The current evaluations, while providing high accuracy levels, focus on metrics that are not the most relevant to a safety–critical situation. Indeed, it is crucial to identify the vehicle's position compared to the pedestrian, their likely collision, and the timing to provide a warning. The obstacle detection techniques mainly focused on the phone camera sensors and other sensors and identified the dangerous locations by using GPS sensors or Bluetooth beacons. However, which technique (i.e., obstacle detection or danger location detection) was more effective was not evaluated in any study and comparison between different phone sensors would also need to be evaluated. The technology-related interventions that depend on vehicle-to-pedestrian communication largely depend on vehicle properties (i.e., speed, distance, and communication devices). It is also unlikely that such communication would encompass all vehicles and pedestrians. Therefore, V2P communication-based interventions largely rely on the level of penetration of the technology, which significantly impacts the performance of the system. For that reason, it may sometimes be preferable to rely on infrastructure-related interventions, though these are not as tailored to the pedestrian specific conditions and have some limitations as well, such as not considering phone distraction before warning and targeting all users of an intersection. All these are necessary to ensure that the technology would be useful and accepted by road users, which is required if such technologies are to result in reduced risks.

Articles evaluating behavioral change found pedestrian behavior change after implementing the interventions. However, interventions do not necessarily focus on the intersections known to result in risks when being distracted by a mobile device or evaluation metrics relevant to the type of impairments observed at such intersections. Such interventions also tend to provide a warning independently of the distraction of the pedestrian, potentially leading to large amounts of warning messages irrelevant to the road user. This likely limits the relevance of the effects reported in the literature, most probably resulting in overestimating the benefits of such interventions. It is also necessary to know why some research only found short term effects (Barin et al., 2018).

Further work is, therefore, necessary to ensure interventions are developed based on known risks for pedestrian distraction rather than on the capabilities of the technology. The evaluation of such technologies (on the technology side) should also focus on metrics relevant to an effective warning message being provided to the road user and should examine the long-term effects of such interventions. Research on the cost-benefit ratio of such interventions is also required to ensure that such intervention approaches are viable. It is also essential to conduct a comparison study between infrastructure-related interventions and road-user-device related interventions considering the randomized design or cross over-design to assess which type of intervention can be utilized in the real world for phone-distracted-pedestrian safety.

4.4. Limitations

Given the broad field, and huge number of advancements in intervention techniques, creating a concise document, synthesising, and synopsis of all data is challenging. To make the process achievable, the considered timeframe was restricted to 10 years. This should be considered as a limitation since studies outside this time frame were omitted. However, most interventions targeting pedestrian distraction are recent, and it is unlikely that this criterion resulted in missing important interventions.

This systematic review only considered articles published in English. This exclusion criterion implies that interventions trialed in non-English speaking countries, or those not published in English, were not identified. Moreover, four databases were used in this review. While these databases are well recognized in this research area, some intervention evaluations may have been present in other databases and may have been missed. However, it is unlikely that major interventions would have been missed, as references cited in the selected articles were examined to ensure no relevant interventions were missed. Finally, only publically available peer-reviewed publications were considered. A number of the trialed interventions were likely evaluated by transport agencies or companies trialing such interventions. However, such information is not available for analysis and cannot be included in this review. Also, the scientific merits of such evaluations could be difficult to assess given the lack of peer-review of such evaluations.

5. Conclusion

Mobile phones can cause a distraction to road users, leading to serious road injuries. This systematic review showed current interventions are primarily focused on infrastructure, road user devices, legislation, and public awareness. The evaluation of these interventions covered three dimensions: effects on behavior; evaluation of the capabilities of the technology; and user acceptance. The infrastructure-related interventions mainly focused on behavior evaluation. These studies found such interventions increase visual screening before crossing and help with choosing a safe gap. Alternatively, road-user devices focused on technology evaluation and, in some cases, user acceptance. Road-user devices currently help detect static and dynamic objects (e.g., cars) with more than 80% efficiency. Some interventions can also detect whether pedestrians are using their mobile phones, resulting in fewer false alarms. User acceptance highlights participants' preference for particular interventions, such as adding flashing borders on mobile phone screens, considered less obstructive than phone vibrations. However, it is clear there is a lack of comprehensive assessments of these interventions, and some interventions have not been extensively evaluated (e.g., legislation and public awareness). This review highlights the gap in knowledge between different intervention techniques for mobile phone distracted pedestrians and the intervention techniques' effectiveness. Such findings make it challenging to compare the safety benefits of each type of intervention. In turn, this leads to limited information currently available to identify the most effective approach to recommend to policymakers for implementation in the field. Thus, substantial opportunity for future research to develop the most suitable type of intervention for mobile phone distracted pedestrians is ensured.

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 did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abdesslem, B., Phillips, & Henderson (2009). Less is more: energy-efficient mobile sensing with senseless. *Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds,* 61–62. https://doi.org/10.1145/1592606.1592621.

Aghabayk, K., Esmailpour, J., Jafari, A., & Shiwakoti, N. (2021). Observational-based study to explore pedestrian crossing behaviors at signalized and unsignalized crosswalks. *Accident Analysis & Prevention, 151*. https://doi.org/10.1016/j.aap.2021.105990 105990.

Alejalil, N., & Davoodi, S. R. (2017). Mobile phone usage and its effects on pedestrians' distraction. *International Journal of High Risk Behaviors and Addiction, 6*(3), 1–6. https://doi.org/10.5812/ijhrba.35431.

Ashqar, H. I., Elhenawy, M., Masoud, M., Rakotonirainy, A., & Rakha, H. A. (2019). Vulnerable road user detection using smartphone sensors and recurrence quantification analysis. *IEEE Intelligent Transportation Systems Conference (ITSC), 2019*, 1054–1059. https://doi.org/10.1109/ITSC.2019.8917520.

Barin, E. N., McLaughlin, C. M., Farag, M. W., Jensen, A. R., Upperman, J. S., & Arbogast, H. (2018). Heads up, phones down: A pedestrian safety intervention on distracted crosswalk behavior. *Journal of Community Health, 43*(4), 810–815. https://doi.org/10.1007/s10900-018-0488-y.

Basch, C. H., Ethan, D., Rajan, S., & Basch, C. E. (2014). Technology-related distracted walking behaviors in Manhattan's most dangerous intersections. *Injury Prevention, 20*(5), 343–346. https://doi.org/10.1136/injuryprev-2013-041063.

Basch, C. H., Ethan, D., Zybert, P., & Basch, C. E. (2015). Pedestrian behavior at five dangerous and busy Manhattan intersections. *Journal of Community Health, 40* (4), 789–792. https://doi.org/10.1007/s10900-015-0001-9.

Bruyneel, A.-V., & Duclos, N. C. (2020). Effects of the use of mobile phone on postural and locomotor tasks: A scoping review. *Gait & Posture, 82*, 233–241. https://doi.org/10.1016/j.gaitpost.2020.09.014.

Byington, K. W., & Schwebel, D. C. (2013). Effects of mobile Internet use on college student pedestrian injury risk. *Accident Analysis & Prevention, 51*, 78–83. https://doi.org/10.1016/j.aap.2012.11.001.

Center, P. R. (2021). *Mobile Fact Sheet*. Pew Research Center. Retrieved 24th April 2022 from https://www.pewresearch.org/internet/fact-sheet/mobile/.

Chang, J. (2022). *90 Smartphone Addiction Statistics You Must See: 2022 Usage and Data Analysis*. FinancesOnline. Retrieved 11th May 2022 from https://financesonline.com/smartphone-addiction-statistics/.

Chen, Zhu, Y., & Wang, G. (2012). Evaluating a mobile pedestrian safety application in a virtual urban environment. *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry,* 175–180. https://doi.org/10.1145/2407516.2407561.

Chen, J., Zhang, Y., & Xue, W. (2018). Unsupervised Indoor Localization Based on Smartphone Sensors, iBeacon and Wi-Fi. *Sensors, 18*(5), 1378. https://doi.org/10.3390/s18051378.

Courtemanche, F., Labonté-LeMoyne, E., Léger, P.-M., Fredette, M., Senecal, S., Cameron, A.-F., Faubert, J., & Bellavance, F. (2019). Texting while walking: An expensive switch cost. *Accident Analysis & Prevention, 127*, 1–8. https://doi.org/10.1016/j.aap.2019.02.022.

Davidson, L. (2012). *Utah'distracted walking'ordinance is working, UTA says*. Retrieved 11th August 2022 from https://archive.sltrib.com/article.php?id=53895472&itype=CMSID.

Davis, S. J., & Barton, B. K. (2017). Effects of secondary tasks on auditory detection and crossing thresholds in relation to approaching vehicle noises. *Accident Analysis & Prevention, 98*, 287–294. https://doi.org/10.1016/j.aap.2016.10.024.

Erkan, I. (2017). Cognitive analysis of pedestrians walking while using a mobile phone. *Journal of Cognitive Science, 18*, 301–319.

Foerster, K.-T., Gross, A., Hail, N., Uitto, J., & Wattenhofer, R. (2014). Spareeye: enhancing the safety of inattentionally blind smartphone users. *Proceedings of the 13th international conference on mobile and ubiquitous multimedia,* 68–72.

Goh, Kim, Han, Han, & Noh (2020). Smombie Forecaster: Alerting Smartphone Users about Potential Hazards in Their Surroundings [Article]. *IEEE Access, 8*, 153183-153191, Article 9170547. https://doi.org/10.1109/ACCESS.2020.3017653.

Gruenefeld, U., Stratmann, T. C., Jung, J., Lee, H., Choi, J., Nanda, A., & Heuten, W. (2018). Guiding Smombies: Augmenting Peripheral Vision with Low-Cost Glasses to Shift the Attention of Smartphone Users. *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct),* 127–131. https://doi.org/10.1109/ISMAR-Adjunct.2018.00050.

Hamann, C., Dulf, D., Baragan-Andrada, E., Price, M., & Peek-Asa, C. (2017). Contributors to pedestrian distraction and risky behaviors during road crossings in Romania [Article]. *Injury Prevention, 23*(6), 370–376. https://doi.org/10.1136/injuryprev-2016-042219.

Holländer, Krüger, & Butz (2020). Save the smombies: App-assisted street crossing [Conference Paper]. *Conference Proceedings - 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services: Expanding the Horizon of Mobile Interaction, MobileHCI 2020*, 1–11, Article 3403547. https://doi.org/10.1145/3379503.3403547.

Horberry, T., Osborne, R., & Young, K. (2019). Pedestrian smartphone distraction: Prevalence and potential severity. *Transportation Research Part F: Traffic Psychology and Behavior, 60*, 515–523. https://doi.org/10.1016/j.trf.2018.11.011.

Hou, M., Cheng, J., Xiao, F., & Wang, C. (2021). Distracted behavior of pedestrians while crossing street: A case study in China. *International Journal of Environmental Research and Public Health, 18*(1), 353.

Hwang, J., Ji, Y., Kwak, N., & Kim, E. Y. (2016). Outdoor context awareness device that enables mobile phone users to walk safely through urban intersections. *ICPRAM 2016 - Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods,* 526–533. https://doi.org/10.5220/0005664705260533.

Hwang, Y., Kim, K. H., & Choi, H. K. (2017). Effects of the providing an augmented reality information on the cognitive reaction behavior under the reduced visibility traffic environment. *International Conference on Information and Communication Technology Convergence: ICT Convergence Technologies Leading the Fourth Industrial Revolution, ICTC 2017, 2017-December,* 1031–1033. https://doi.org/10.1109/ICTC.2017.8190845.

Jain, S. (2015). Gradient profiling for pedestrian services. *PhDForum 2015 - Proceedings of the 2015 MobiSys PhD Forum,* 11–12. https://doi.org/10.1145/2752746.2752784.

Jain, S., Borgiattino, C., Ren, Y., Gruteser, M., Chen, Y., & Chiasserini, C. F. (2015). Lookup: Enabling pedestrian safety services via shoe sensing. *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services,* 257–271. https://doi.org/10.1145/2742647.2742669.

Jain, S., & Gruteser, M. (2019). Recognizing Textures with Mobile Cameras for Pedestrian Safety Applications [Article]. *IEEE Transactions on Mobile Computing, 18*(8), 1911–1923, Article 8454285. https://doi.org/10.1109/TMC.2018.2868659.

Jiang, K., Ling, F., Feng, Z., Ma, C., Kumfer, W., Shao, C., & Wang, K. (2018). Effects of mobile phone distraction on pedestrians' crossing behavior and visual attention allocation at a signalized intersection: An outdoor experimental study. *Accident Analysis & Prevention, 115*, 170–177. https://doi.org/10.1016/j.aap.2018.03.019.

Kalra, P., Mittal, D., & Sidhu, E. (2017). Autonomous solar smart cap (ASSC) for pedestrian safety. *International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), 2017*, 307–311. https://doi.org/10.1109/ICBDACI.2017.8070854.

Kang, H., & Han, J. (2020). SafeXR: Alerting walking persons to obstacles in mobile XR environments. *The Visual Computer, 36*(10), 2065–2077. https://doi.org/10.1007/s00371-020-01907-4.

Kang, H., Lee, G., & Han, J. (2019). Obstacle detection and alert system for smartphone ar users. *25th ACM Symposium on Virtual Reality Software and Technology,* 1–11. https://doi.org/10.1145/3359996.3364256.

Kim, Han, Sim, & Noh (2018). Smombie guardian: We watch for potential obstacles while you are walking and conducting smartphone activities [Article]. *PLoS ONE, 13*(6), Article e0197050. https://doi.org/10.1371/journal.pone.0197050.

Kim, E., Kim, H., Kwon, Y., Choi, S., & Shin, G. (2021). Performance of ground-level signal detection when using a phone while walking [Article]. *Accident; Analysis and Prevention, 151*. https://doi.org/10.1016/j.aap.2020.105909 105909.

Kim, J., Jo, Y., & Jeong, J. P. (2015). Design and evaluation of a smartphone-based alarming system for pedestrian safety in vehicular networks. *International Conference on Internet of Vehicles, 221–233*. https://doi.org/10.1007/978-3-319-27293-1_20.

Kumar, A., Singh, A., & Sidhu, E. (2015). Smart-cap for pedestrian safety. *2015 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT),* 583–587. https://doi.org/10.1109/ICCICCT.2015.7475346.

Larue, & Watling (2021). Acceptance of visual and audio interventions for distracted pedestrians [Article]. *Transportation Research Part F: Traffic Psychology and Behavior, 76*, 369–383. https://doi.org/10.1016/j.trf.2020.12.001.

Larue, Watling, Black, & Wood (2021). Improving the safety of distracted pedestrians with in-ground flashing lights. A railway crossing field study. *Journal of Safety Research, 77*, 170–181. https://doi.org/https://doi.org/10.1016/j.jsr.2021.02.014.

Larue, Watling, Black, Wood, & Khakzar. (2020). Pedestrians distracted by their smartphone: Are in-ground flashing lights catching their attention? A laboratory study [Article]. *Accident Analysis & Prevention, 134*, 105346. https://doi.org/10.1016/j.aap.2019.105346.

Law, I. (2013). *What's as dangerous as distracted driving? Distracted walking*. Wheels. ca. Retrieved 18th May 2021 from https://www.wheels.ca/news/whats-as-dangerous-as-distracted-driving-distracted-walking/.

Lennon, A., Oviedo-Trespalacios, O., & Matthews, S. (2017). Pedestrian self-reported use of smart phones: Positive attitudes and high exposure influence intentions to cross the road while distracted [Article]. *Accident Analysis & Prevention, 98*, 338–347. https://doi.org/10.1016/j.aap.2016.10.028.

Lewandowski, Bocker, Koster, & Wietfeld (2013). Design and performance analysis of an IEEE 802.15. 4 V2P pedestrian protection system. *2013 IEEE 5th international symposium on wireless vehicular communications (WiVeC),* 1–6. https://doi.org/10.1109/wivec.2013.6698232.

Li, Xue, Fan, Qu, & Zhou. (2018). Pedestrian walking safety system based on smartphone built-in sensors. *IET Communications, 12*(6), 751–758. https://doi.org/10.1007/978-3-030-32150-5_88.

Liu, X., Cao, J., Wen, J., & Tang, S. (2016). Infrasee: An unobtrusive alertness system for pedestrian mobile phone users. *IEEE Transactions on Mobile Computing, 16*(2), 394–407. https://doi.org/10.1109/TMC.2016.2550447.

Liu, Z., Pu, L., Meng, Z., Yang, X., Zhu, K., & Zhang, L. (2015). POFS: A novel pedestrian-oriented forewarning system for vulnerable pedestrian safety. *International Conference on Connected Vehicles and Expo (ICCVE), 2015*, 100–105. https://doi.org/10.1109/ICCVE.2015.63.

Malathy, N., Nandha, S. S., Praveen, B., & Kumar, K. P. (2019). Pedwarn-Enhancement of Pedestrian Safety Using Mobile Application. *International Conference on Emerging Current Trends in Computing and Expert Technology, 35*, 877–887. https://doi.org/10.1007/978-3-030-32150-5_88.

Marsalia, A. C., Ferris, T. K., Benden, M. E., & Zheng, Q. (2016). Evaluation of vibrotactile warning systems for supporting hazard awareness and safety of distracted pedestrians. *IIE Transactions on Occupational Ergonomics and Human Factors, 4*(4), 222–235. https://doi.org/10.1080/21577323.2016.1214767.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Oct). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement.

Journal of Clinical Epidemiology, 62(10), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005.
Moudon, A. V., Lin, L., Jiao, J., Hurvitz, P., & Reeves, P. (2011). The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington. *Accident Analysis & Prevention, 43*(1), 11–24. https://doi.org/10.1016/j.aap.2009.12.008.
Mwakalonge, J., Siuhi, S., & White, J. (2015). Distracted walking: Examining the extent to pedestrian safety problems [Article]. *Journal of traffic and transportation engineering (English edition), 2*(5), 327–337. https://doi.org/10.1016/j.jtte.2015.08.004.
Narváez, Y. V., Parra Sierra, V., Peña Cárdenas, F., Ruíz Ramos, L., Zamorano González, B., Vargas Martínez, J. I., & Monreal Aranda, O. (2019). Road risk behaviors: Pedestrian experiences. *Traffic injury prevention, 20*(3), 303–307.
Osborne, Horberry, & Young (2020). Pedestrian distraction from Smartphones: An end-user perspective on current and future countermeasures. *Transportation Research Part F: Traffic Psychology and Behavior, 73*, 348–361.
Ou, R., Peng, F., Zhang, T., Li, J., Xu, J., Chen, Y., & Yin, X. (2019). WatchOUT: A Practical Secure Pedestrian Warning System Based on Smartphones. *IEEE Global Communications Conference (GLOBECOM), 2019*, 1–6. https://doi.org/10.1109/GLOBECOM38437.2019.9014250.
PCA (2012). *LAMBS TO THE SLAUGHTER - WAIT FOR THE GREEN*. Pedestrian Council of Australia. Retrieved 19th August 2022 from https://www.walk.com.au/pedestriancouncil/page.asp.
Pešić, D., Antić, B., Glavić, D., & Milenković, M. (2016). The effects of mobile phone use on pedestrian crossing behavior at unsignalized intersections–Models for predicting unsafe pedestrians behavior. *Safety Science, 82*, 1–8. https://doi.org/10.1016/j.ssci.2015.08.016.
Proctor, R. W., & Vu, K.-P.-L. (2019). How psychologists help solve real-world problems in multidisciplinary research teams: Introduction to the special issue. *American Psychologist, 74*(3), 271. https://doi.org/10.1037/amp0000458.
Rahimian, P., O'Neal, E. E., Yon, J. P., Franzen, L., Jiang, Y., Plumert, J. M., & Kearney, J. K. (2016). Using a virtual environment to study the impact of sending traffic alerts to texting pedestrians. *IEEE Virtual Reality (VR), 2016*, 141–149. https://doi.org/10.1109/VR.2016.7504697.
Rahimian, P., O'Neal, E. E., Zhou, S., Plumert, J. M., & Kearney, J. K. (2018). Harnessing vehicle-to-pedestrian (V2P) communication technology: Sending traffic warnings to texting pedestrians. *Human Factors, 60*(6), 833–843. https://doi.org/10.1177/0018720818781365.
Russo, B. J., James, E., Aguilar, C. Y., & Smaglik, E. J. (2018). Pedestrian behavior at signalized intersection crosswalks: Observational study of factors associated with distracted walking, pedestrian violations, and walking speed. *Transportation Research Record, 2672*(35), 1–12. https://doi.org/10.1177/0361198118759949.
Savolainen, P. T., Gates, T. J., & Datta, T. K. (2011). Implementation of targeted pedestrian traffic enforcement programs in an urban environment. *Transportation Research Record, 2265*(1), 137–145.
Schwebel, Hasan, R., Griffin, R., Hasan, R., Hoque, M. A., Karim, M. Y., Luo, K., & Johnston, A. (2021). Reducing distracted pedestrian behavior using bluetooth beacon technology: a crossover trial. *Accident Analysis & Prevention, 159*, 106253. https://doi.org/10.1016/j.aap.2021.106253.
Schwebel, McClure, & Porter (2017). Experiential exposure to texting and walking in virtual reality: A randomized trial to reduce distracted pedestrian behavior. *Accident Analysis & Prevention, 102*, 116–122. https://doi.org/10.1016/j.aap.2017.02.026.
Scopatz, R. A., & Zhou, Y. (2016). *Effect of electronic device use on pedestrian safety: a literature review* (DOT HS 812 256). (National Highway Traffic Safety Administration), Issue.
Seeing Eye People. (2013). Improve Everywhere. Retrieved 18th May 2021 from https://improveverywhere.com/2013/04/30/seeing-eye-people/.
Shaaban, K., Muley, D., & Mohammed, A. (2018). Analysis of illegal pedestrian crossing behavior on a major divided arterial road. *Transportation Research Part F: Traffic Psychology and Behavior, 54*, 124–137. https://doi.org/10.1016/j.trf.2018.01.012.
Silver, L. (2019). *Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally*. Pew Researcher Center. Retrieved 24th April 2022 from https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/.
Simmons, S. M., Caird, J. K., Ta, A., Sterzer, F., & Hagel, B. E. (2020). Plight of the distracted pedestrian: A research synthesis and meta-analysis of mobile phone use on crossing behavior. *Injury Prevention, 26*(2), 170–176. https://doi.org/10.1136/injuryprev-2019-043426.
Sobhani, & Farooq (2018). Impact of smartphone distraction on pedestrians' crossing behavior: An application of head-mounted immersive virtual reality [Article]. *Transportation Research Part F: Traffic Psychology and Behavior, 58*, 228–241. https://doi.org/10.1016/j.trf.2018.06.020.
Sobhani, Farooq, & Zhong, Z (2017). Distracted pedestrians crossing behavior: Application of immersive head mounted virtual reality. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC),* 1–6. https://doi.org/10.1109/ITSC.2017.8317769.
Solah, M. S., Deros, B. M., Jawi, Z. M., Harun, N. Z., Hamzah, A., & Ariffin, A. H. (2016). The effects of mobile electronic device use in influencing pedestrian crossing behavior. *Malaysian Journal of Public Health Medicine, 16*(Special Volume (1)), 61–66.
Sun, C., Su, J., Ren, W., & Guan, Y. (2019). Wide-View Sidewalk Dataset Based Pedestrian Safety Application [Article]. *IEEE Access, 7*, 151399-151408, Article 8867891. https://doi.org/10.1109/ACCESS.2019.2947165.
Syazwan, S. M., Baba, M. D., Zarifah, H. N., Hafeez, A. A., & Faradila, P. N. (2017). Prevalence of distracted pedestrians while crossing: A study of Malaysia's situation. *MATEC Web of Conferences, 90*, 01031. https://doi.org/10.1051/matecconf/20179001031.
Tang, M., Nguyen, C.-T., Wang, X., & Lu, S. (2016). An efficient walking safety service for distracted mobile users. *2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS),* 84–91.
Tapiro, H., Oron-Gilad, T., & Parmet, Y. (2018). The effect of environmental distractions on child pedestrian's crossing behavior. *Safety Science, 106*, 219–229. https://doi.org/10.1016/j.ssci.2018.03.024.
Thompson, L. L., Rivara, F. P., Ayyagari, R. C., & Ebel, B. E. (2013). Impact of social and technological distraction on pedestrian crossing behavior: An observational study [Article]. *Injury Prevention, 19*(4), 232–237. https://doi.org/10.1136/injuryprev-2012-040601.
Transportation, U. S. D. o. (2022). *NHTSA Releases 2020 Traffic Crash Data*. NHTSA Media. Retrieved 9th September 2022 from https://www.nhtsa.gov/press-releases/2020-traffic-crash-data-fatalities.
Tung, Y. C., & Shin, K. G. (2018). Use of Phone Sensors to Enhance Distracted Pedestrians' Safety [Article]. *IEEE Transactions on Mobile Computing, 17*(6), 1469–1482. https://doi.org/10.1109/TMC.2017.2764909.
Violano, Roney, & Bechtel (2015). The incidence of pedestrian distraction at urban intersections after implementation of a Streets Smarts campaign [Article]. *Injury epidemiology, 2*(1), 1–7. https://doi.org/10.1186/s40621-015-0050-7.
Wang, Q., Guo, B., Peng, G., Zhou, G., & Yu, Z. (2016). CrowdWatch: Pedestrian safety assistance with mobile crowd sensing. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct,* 217–220. https://doi.org/10.1145/2968219.2971433.
Wang, T., Cardone, G., Corradi, A., Torresani, L., & Campbell, A. T. (2012). Walksafe: a pedestrian safety app for mobile phone users who walk and talk while crossing roads. *Proceedings of the twelfth workshop on mobile computing systems & applications,* 1–6. https://doi.org/10.1145/2162081.2162089.
Wen, J., Cao, J., & Liu, X. (2015). We help you watch your steps: Unobtrusive alertness system for pedestrian mobile phone users. *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom),* 105–113.
WHO. (2013). *Pedestrian safety*. WHO. Retrieved 19/03/2021 from https://apps.who.int/iris/bitstream/handle/10665/79753/9789241505352_eng.pdf?sequence=1.
Won, M., Shrestha, A., Park, K.-J., & Eun, Y. (2020). SaferCross: Enhancing Pedestrian Safety Using Embedded Sensors of Smartphone. *IEEE Access, 8*, 49657–49670. https://doi.org/10.1109/access.2020.2980085.
Wong, J. (2017). *Melbourne's footpaths light up in the name of pedestrian safety*. Drive.com.au. Retrieved 26th April 2022 from https://www.drive.com.au/news/melbournes-footpaths-light-up-in-the-name-of-pedestrian-safety/.
Wundersitz, L., & Hutchinson, T. (2011). What can we learn from recent evaluations of road safety mass media campaigns? *Journal of the Australasian College of Road Safety, 22*(4), 40–50.
Xia, S., de Godoy Peixoto, D., Islam, B., Islam, M. T., Nirjon, S., Kinget, P. R., & Jiang, X. (2019). Improving pedestrian safety in cities using intelligent wearable systems. *IEEE Internet of Things Journal, 6*(5), 7497–7514. https://doi.org/10.1109/JIOT.2019.2903519.
Zegeer, C. V., & Bushell, M. (2012). Pedestrian crash trends and potential countermeasures from around the world. *Accident Analysis & Prevention, 44*(1), 3–11.
Zhou, Z. (2015). Headsup: Keeping pedestrian phone addicts from dangers using mobile phone sensors. *International Journal of Distributed Sensor Networks, 11*(5). https://doi.org/10.1155/2015/279846 279846.
Zhuang, Y., & Fang, Z. (2020). Smartphone zombie context awareness at crossroads: A multi-source information fusion approach. *IEEE Access, 8*, 101963–101977. https://doi.org/10.1109/ACCESS.2020.2998129.

Md Eaysir Arafat is doing his Master of Philosophy at the Queensland University of Technology under the Centre for Accident Research and Road Safety Queensland (CARRS-Q). Alongside Eaysir Studies, he also works as a Research Officer in different road safety projects. Before joining CARRS-Q, he worked as an intern at the Australian Heavy Engineering Company. He is passionate about road safety and road user behaviour. Currently, his work focuses on evaluating the effectiveness of warning interventions for mobile phone-distracted pedestrians in signalised intersections. His professional interests focus on transportation modelling, road safety, vehicle crash prediction, and prevention.

Associate Professor Gregoire Larue is a principal research fellow at the Road Safety Research Collaboration. He has a PhD in road safety from the Queensland University of Technology, which investigated driver vigilance impairment on monotonous Australian roads. He has worked in the field of road and rail safety since 2007. His expertise combines human factors, engineering and mathematics to evaluate interventions designed for road users. He has published widely in renowned journals, and he serves in different Standard Development Groups (Rail Industry Safety and Standards Board, Standards Australia) for including human factors research findings into standards.

Dr Sepehr G. Dehkordi is a Senior Technology Leader of the Australian Road Research Board (ARRB). Before joining ARRB, he was a Research Associate at the Queensland University of Technology, Centre for Accident Research and Road Safety – Queensland (CARRS-Q). He has a background in Control and Automation Engineering, and he is currently researching the Connected Automated Vehicle and Intelligence Transportation System (ITS). His research and experience cover several areas, including optimisation techniques in control theory, intelligent control, road safety approaches in ITS, ecological driving behaviours, system identification and dynamic system modelling.

Journal of Safety Research 84 (2023) 212–217

Contents lists available at ScienceDirect

Journal of Safety Research

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

Effects on accidents of technical inspections of heavy goods vehicles in Norway: A re-analysis and a replication

Rune Elvik

Institute of Transport Economics, Gaustadalleen 21, 0349 Oslo, Norway

ARTICLE INFO

Article history:
Received 27 May 2022
Accepted 26 October 2022
Available online 4 November 2022

Keywords:
Heavy goods vehicle
Technical inspections
Accidents
Evaluation study

ABSTRACT

Introduction: This paper presents a re-analysis of a previous study of the effects on accidents of technical inspections of heavy vehicles in Norway and a replication of the study using more recent data. *Method:* Increasing the number of technical inspections is associated with a reduction in the number of accidents. Reducing the number of inspections is associated with an increase in the number of accidents. The relationship between changes in the number of inspections and changes in the number of accidents is well described by means of logarithmic dose–response curves. *Results:* These curves show that inspections had a larger effect on accidents in the recent period (2008–2020) than in the first period (1985–1997). Based on recent data, a 20% increase in the number of inspections is associated with a 4–6% reduction in the number of accidents. A 20% reduction of the number of inspections is associated with a 5–8% increase in the number of accidents.

1. Introduction

Heavy goods vehicles that have technical defects have a higher rate of accident involvement than heavy goods vehicles that do not have technical defects. Jones and Stein (1989) found a relative risk of about 1.7, and a population attributable risk of 0.32, meaning that by eliminating technical defects, the number of accidents could be reduced by 32%. Teoh, Carter, Smith, and McCartt (2017) found a relative risk of 3.1 and a population attributable risk of 0.51. Technical defects are therefore an important risk factor for heavy goods vehicles.

In accordance with EU Directive 2014/47, Norway has implemented technical roadside inspections of heavy goods vehicles. Vehicles are inspected by vehicle experts employed by the Public Roads Administration. Vehicles are inspected in roadside inspection stations that have equipment for measuring, for example, vehicle weight and braking performance. Defects are graded as minor, major, or dangerous. A defect graded as dangerous results in vehicle impoundment, that is, the vehicle is out of service until the defect has been repaired.

The number of technical inspections varies from year to year but has been between 0.8 and 1.2 per registered heavy goods vehicle per year in recent years. Thus, on average, a heavy goods vehicle can expect to be inspected about once per year. Elvik (2002) evaluated the effects on accidents of technical inspections carried out in Norway between 1985 and 1997. He found a statistically non-significant association between the number of inspections per vehicle per year and accident rate. The association indicated that by doubling the number of inspections, accident rate would decline by 5–10%.

The statistical technique used by Elvik (2002) was ordinary least-squares linear regression, using various indicators of accident rate as dependent variable. Technical inspections, as well as other variables, were measured in terms of annual percentage changes, that is, increases or decreases from the year before. This approach may have had low statistical power, resulting in non-significant findings.

This paper has two objectives. The first is to re-analyze the 2002-study, using a more appropriate count regression model (negative binomial regression). The second is to replicate the study, by doing a similar analysis of data covering the years from 2008 to 2020.

2. Previous studies

The study published in 2002 (Elvik, 2002) is one of very few studies of the effects on accidents of technical inspections of heavy vehicles. Some studies, notably Jones and Stein (1989), Moses and Savage (1992), and Teoh et al. (2017) have estimated the increase in the risk of accidents associated with technical defects. These studies suggest that by eliminating or reducing technical defects, the number of accidents can be reduced. However, the studies

E-mail address: re@toi.no

https://doi.org/10.1016/j.jsr.2022.10.021
0022-4375/© 2022 The Author(s). Published by the National Safety Council and Elsevier Ltd.

say nothing about the type or intensity of technical inspections needed to substantially reduce technical defects.

A Canadian study (Gou, Clément, Birikundavyi, Bellavigna-Ladoux, & Abraham, 1999) estimated that technical defects contributed to about 15% of accidents involving heavy goods vehicles. An American study (Thakuriah, Yanos, Lee, & Sreenivasan, 2001) found a weak tendency for vehicles that had many technical defects when inspected to be more often involved in accidents the following year, compared to vehicles that had no technical defects when inspected. A recent Spanish study (Diaz Lopez, 2019) estimated that periodic motor-vehicle inspections prevent about 7% of all injury accidents. This estimate, however, did not include heavy goods vehicles.

Elvik (2002) estimated that doubling the number of roadside technical inspections per vehicle could reduce the number of accidents involving heavy vehicles (both buses and trucks) by about 7%. This estimate was associated with large uncertainty (95% confidence interval from −18.4% to +5.1%).

3. Re-analysis of previous study

The data used in the study by Elvik (2002) are reproduced in Table 1. The table does not include data on light vehicles, as these are not used in the re-analysis presented in this paper. Heavy vehicles include both buses and trucks.

To estimate the effect of technical inspections on the number of accidents, a negative binomial regression model was developed. The model was fitted in four stages. The first stage included only the constant term and a term for technical inspections. In the next three stages, other variables were added to the model. Estimated coefficients are shown in Table 2.

The final model (model 4) contained four independent variables. With only 13 units of observation, it was not possible to include more variables.

The coefficient for technical inspections was negative and statistically significant in all model specifications, suggesting that inspections reduce the number of accidents. The final model explained 94.75% of the systematic variation in the number of accidents according to the Elvik-index of goodness-of-fit (Fridstrøm, Ifver, Ingebrigtsen, Kulmala, & Krogsgård Thomsen, 1995). There was no statistically significant autocorrelation of the residual terms for lags from 1 to 11. Fig. 1 shows the data and model predictions.

There was a downward trend in the number of accidents until about 1990 and a weak upward trend after 1990. To identify the contribution of variations in the number of technical inspections to the annual changes in the number of accidents, annual changes were computed under two conditions: (1) Based on the number of accidents predicted by the model including technical inspections; (2) Based on the number of accidents predicted by a model not including technical inspections. The latter model is intended to establish the counterfactual (i.e., describe the annual changes in the number of accidents that would have occurred if technical inspections did not exist). The differences between the annual differences identifies the contribution to changes in the number of accidents from year N to year N + 1 from changes in the number of technical inspections per vehicle from year N to year N + 1. Table 3 shows these computations.

The first column in Table 3 shows the recorded number of accidents. The next two columns show the number of accidents predicted with and without technical inspections. The fourth and fifth columns show the annual differences in the number of accidents as predicted with and without technical inspections. In the sixth column, the differences between differences are computed in order to identify the annual contribution of technical inspections. It is seen that this contribution is sometimes negative (i.e., a reduction of the number of accidents) and sometimes positive (i.e., an increase in the number of accidents).

One would expect an increase in the number of inspections to be associated with a reduction in the number of accidents and a reduction in the number of inspections to be associated with an increase in the number of accidents. The last two columns of Table 3 provide information to assess whether there is such an association. These columns state the relative change in the number of inspections and the relative change in the number of accidents. The latter was computed relative to the predicted number of accidents in the model, including technical inspections. Values above 1 indicate increases, values below 1 indicate decreases.

The data in the two rightmost columns of Table 3 are plotted in Fig. 2. Fig. 2 shows the association between annual changes in the number of technical inspections and annual changes in the number of accidents.

The association between changes in the number of technical inspections and changes in the number of accidents is well described by a logarithmic function. The function passes straight through the equilibrium point for effects of enforcement, that is, no change in enforcement is associated with no change in accidents (Bjørnskau & Elvik, 1992). The standard error of the coefficient for the logarithmic term is 0.015. It is highly statistically significant. A 95% confidence interval for the estimated effect on accidents of a 20% reduction in the number of inspections is (+0.9%; +2.2%). For a 20% increase in the number of inspections, the 95% confidence interval is (−0.8%; −1.8%).

The curve in Fig. 2 is strongly influenced by the data point in the upper left corner. Would the results be different if this data point is

Table 1
Data used in original study.

Year	Number of Inspections	Accidents involving heavy	Number of vehicles	Million vehicle kilometers	New drivers	All drivers	Inspections per vehicle	New drivers as proportion	Change (%) in GDP/capita
1985	39,134	1180	90,270	2486	9214	304,416	0.434	0.030	5.2
1986	42,940	1232	94,963	2971	9838	317,250	0.452	0.031	3.6
1987	47,708	1202	98,203	3182	10,818	331,992	0.486	0.033	2.0
1988	69,039	1064	98,131	3387	10,631	353,696	0.704	0.030	−0.1
1989	93,490	974	96,587	3475	3805	345,678	0.968	0.011	0.9
1990	113,259	943	95,505	3552	7952	367,262	1.186	0.022	2.0
1991	128,920	1027	95,412	3634	8238	375,938	1.351	0.022	3.1
1992	182,768	995	97,028	3728	7766	383,344	1.884	0.020	3.3
1993	58,310	1008	97,494	3820	7350	389,496	0.598	0.019	2.7
1994	55,990	1046	98,257	3957	6932	395,519	0.570	0.018	5.5
1995	50,143	1074	100,219	4127	7026	400,730	0.500	0.018	3.8
1996	48,340	1082	103,331	4197	11,167	407,403	0.468	0.027	4.9
1997	42,543	1068	107,763	4636	8502	409,593	0.395	0.021	4.7

Table 2
Negative binomial regression of data for 1985–1997. Regression coefficients and standard errors.

Term	Regression coefficients. Standard errors in parentheses. P-value in square brackets			
	Model 1	Model 2	Model 3	Model 4
Constant term	7.061 (0.0196) [0.000]	7.307 (0.0655) [0.000]	6.993 (0.1271) [0.000]	6.978 (0.1153) [0.000]
Inspections per vehicle	−0.115 (0.0227) [0.000]	−0.114 (0.0225) [0.000]	−0.082 (0.0265) [0.002]	−0.070 (0.0251) [0.005]
Million vehicle kilometers		0.000068 (0.000017) [0.000]	0.000029 (0.000023) [0.209]	0.000037 (0.000021) [0.084]
Proportion of new drivers			6.348 (2.1304) [0.003]	6.271 (1.9357) [0.001]
Change in GDP per capita				0.011 (0.0062) [0.082]
Elvik-index of goodness-of-fit				0.9475

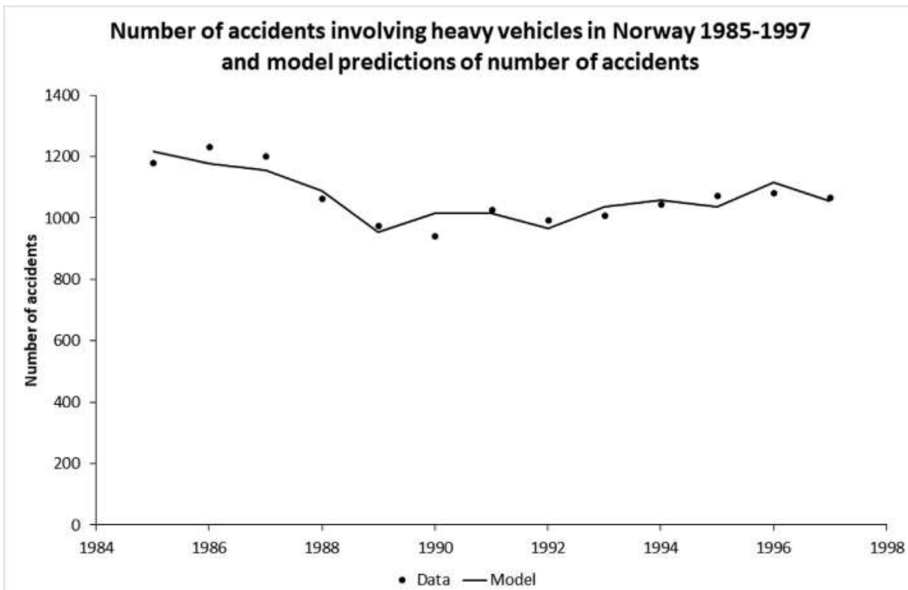


Fig. 1. Number of accidents involving heavy vehicles in Norway 1985–1997 and model prediction of number of accidents.

Table 3
Differences-in-differences estimate of effects on accidents of technical inspections 1985–1997.

Year	Accidents	Predicted with inspections	Predicted without inspections	Annual differences with inspections	Annual differences without inspections	Difference in differences	Relative change in inspections	Relative change in accidents
1985	1180	1216.43	1206.03					
1986	1232	1178.12	1167.28	−38.31	−38.75	0.44	1.043	1.000
1987	1202	1157.34	1146.43	−20.79	−20.85	0.06	1.074	1.000
1988	1064	1088.09	1079.78	−69.25	−66.65	−2.60	1.448	0.998
1989	974	955.27	933.71	−132.82	−146.07	13.25	1.376	1.014
1990	943	1015.13	1035.84	59.86	102.13	−42.27	1.225	0.958
1991	1027	1014.28	1053.80	−0.85	17.96	−18.81	1.139	0.981
1992	995	965.88	1039.91	−48.39	−13.89	−34.50	1.394	0.964
1993	1008	1037.29	1015.66	71.40	−24.25	95.65	0.318	1.092
1994	1046	1057.54	1045.61	20.25	29.95	−9.70	0.953	0.991
1995	1074	1036.58	1013.18	−20.96	−32.43	11.47	0.878	1.011
1996	1082	1115.92	1117.06	79.34	103.88	−24.54	0.935	0.978
1997	1068	1056.33	1040.61	−59.59	−76.45	16.86	0.844	1.016

omitted? If omitted, there is still a negative relationship between the number of inspections and the number of accidents. The relationship is, however, considerably weaker. The coefficient for the logarithmic function is –0.035 (–0.071 when all data points are included), with a standard error of 0.032.

As the data in Table 1 show, there was a sharp reduction of the number of technical inspections from 1992 to 1993 and an increase in the number of accidents. It is therefore concluded that the data point referring to changes from 1992 to 1993 should be included, although it is located far from the other data points.

4. Replication

The replication copied the re-analysis presented above as far as possible. The data used in the replication are shown in Table 4. The replication included only heavy goods vehicles.

The variables of principal interest are the number of accidents and the number of technical inspections per vehicle. The number of technical inspections per vehicle per year fluctuates between about 0.85 and 1.20. This variation is considerably smaller than during 1985–1997.

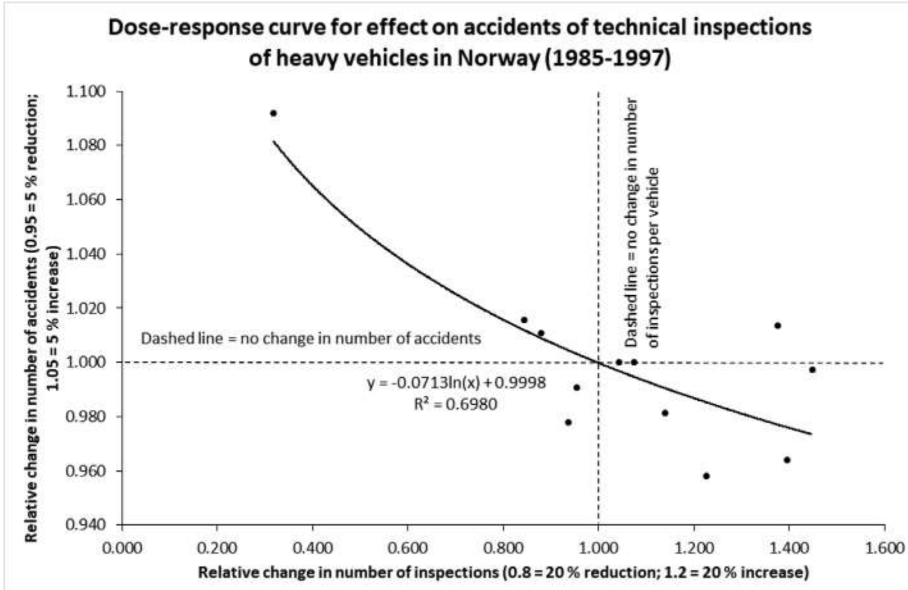


Fig. 2. Dose-response curve for effect on accidents of technical inspections of heavy vehicles in Norway (1985–1997).

Table 4
Data used in replication study.

Year	Number of vehicles	Million vehicle kilometers	Technical inspections	Inspections per vehicle	Change (%) in GDP/capita	Proportion of young drivers	Accidents involving heavy
2007	84,742	1722			1.0	0.007	786
2008	84,350	1886	82,032	0.973	−0.8	0.007	675
2009	82,694	1737	76,783	0.929	−3.0	0.008	581
2010	81,330	1799	83,784	1.030	−0.5	0.008	602
2011	80,160	1832	68,181	0.851	−0.3	0.012	521
2012	79,857	1938	73,409	0.919	1.4	0.011	551
2013	79,437	1950	69,824	0.879	−0.2	0.014	459
2014	78,668	1909	86,571	1.100	0.8	0.013	384
2015	77,120	1883	70,404	0.913	0.9	0.013	312
2016	75,238	1818	83,160	1.105	0.2	0.014	335
2017	73,808	1855	88,313	1.197	1.5	0.014	327
2018	72,405	1760	82,611	1.141	0.5	0.014	333
2019	72,078	1798	77,734	1.078	0.1	0.014	319
2020	70,670	2035	79,042	1.118	−1.3	0.014	251

A negative binomial regression model was developed in four stages, as in the re-analysis. Exploratory analysis found that when kilometers driven was entered, the coefficient for technical inspections indicated an implausibly large effect. Kilometers driven was therefore replaced by number of trucks. Estimated coefficients are presented in Table 5.

The coefficient for technical inspections was negative in three of the four models, but was not statistically significant in any of models 2–4. It nevertheless approached statistical significance in model 4. Fig. 3 shows how the model fits the data. There was a large decline in the number of accidents from 2008 to 2020, and the model captures this decline. It explained 94.21% of the systematic variation in the number of accidents (Elvik-index).

There was some autocorrelation of residual terms. It was statistically significant at the 5% level for lags 4–8, but not for lags 1–3 and 9–11. The contribution of technical inspections to the annual changes in the number of accidents was estimated the same way as in the re-analysis. The number of accidents predicted by models including and not including technical inspections was estimated and annual differences taken. Differences between differences show the annual contribution of changes in the number of technical inspections (see Table 6).

The data in the two rightmost columns of Table 6 serve as the basis for the data presented in Fig. 4. Fig. 4 shows a dose–response curve for the association between technical inspections and the number of accidents during 2008–2020.

Table 5
Negative binomial regression of data for 2008–2020. Regression coefficients and standard errors.

	Regression coefficients. Standard errors in parentheses. P-value in square brackets			
Term	Model 1	Model 2	Model 3	Model 4
Constant term	7.841 (0.6459) [0.000]	0.398 (0.5092) [0.435]	3.219 (1.0691) [0.003]	4.058 (1.1768) [0.001]
Inspections per vehicle	−1.752 (0.6316) [0.006]	0.123 (0.1829) [0.502]	−0.208 (0.2136) [0.331]	−0.339 (0.2262) [0.134]
Number of heavy goods vehicles		0.000071 (0.0000047) [0.000]	0.000045 (0.0000099) [0.000]	0.000038 (0.000011) [0.000]
Proportion of young drivers			−38.148 (12.7529) [0.003]	−52.698 (15.3647) [0.001]
Change in GDP per capita				0.027 (0.0157) [0.089]
Elvik-index of goodness-of-fit				0.9421

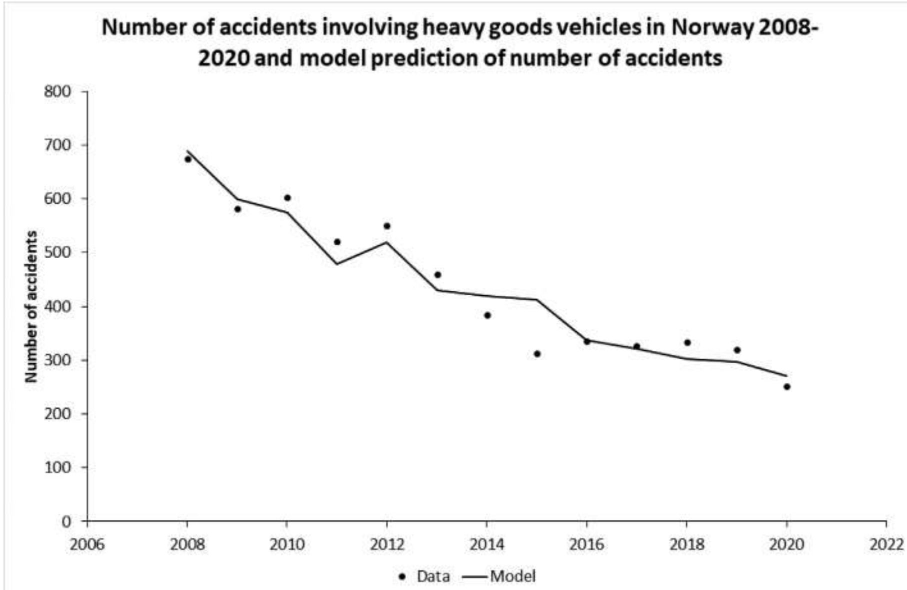


Fig. 3. Number of accidents involving heavy goods vehicles in Norway 2008–2020 and model prediction of number of accidents.

Table 6
Differences-in-differences estimate of effects on accidents of technical inspections 2008–2020.

Year	Accidents	Predicted with inspections	Predicted without inspections	Annual differences with inspections	Annual differences without inspections	Difference in differences	Relative change in inspections	Relative change in accidents
2007	786	688.08	709.23					
2008	675	599.41	596.39	−88.67	−112.84	24.17	0.955	1.040
2009	581	575.62	588.22	−23.79	−8.17	−15.62	1.109	0.973
2010	602	478.18	468.41	−97.44	−119.81	22.37	0.826	1.047
2011	521	519.75	515.47	41.57	47.06	−5.49	1.081	0.989
2012	551	429.39	425.98	−90.36	−89.49	−0.87	0.956	0.998
2013	459	419.32	441.72	−10.07	15.74	−25.81	1.252	0.938
2014	384	411.53	401.01	−7.79	−40.71	32.92	0.830	1.080
2015	312	337.58	344.78	−73.95	−56.23	−17.72	1.211	0.948
2016	335	320.62	333.18	−16.96	−11.60	−5.36	1.083	0.983
2017	327	302.18	302.01	−18.44	−31.17	12.73	0.954	1.042
2018	333	296.68	289.64	−5.50	−12.37	6.87	0.945	1.023
2019	319	271.56	262.72	−25.12	−26.92	1.80	1.037	1.007

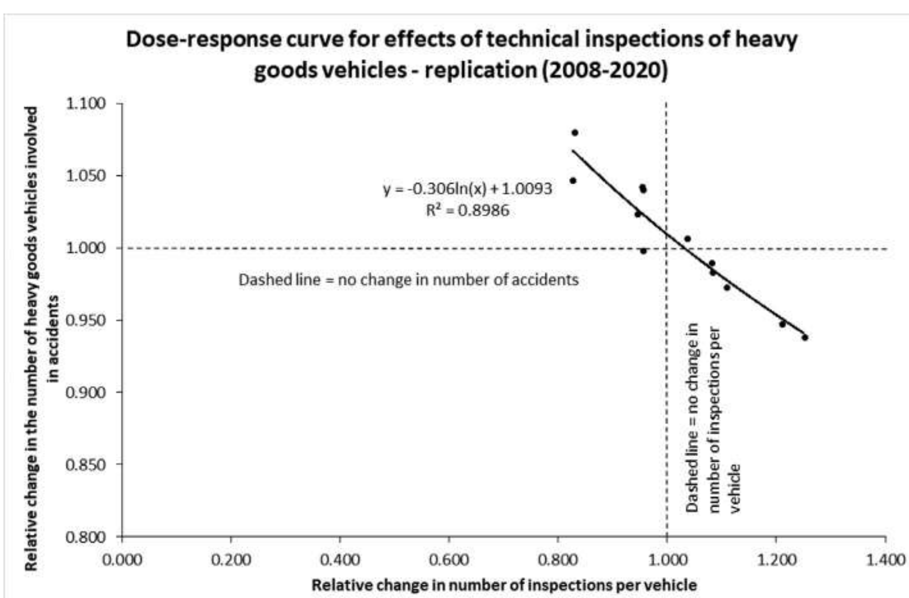


Fig. 4. Dose-response curve for effects of technical inspections of heavy goods vehicles – replication (2008–2020).

A logarithmic function fits the data well. The curve passes close to the general equilibrium point for effects of enforcement (i.e., the intersection of the lines showing no change in number of inspections and no change in number of accidents). The standard error of the coefficient for the logarithmic term is 0.032. A 95% confidence interval for the estimated effect of a 20% reduction in the number of inspections is (+5.4%; +8.2%). A 95% confidence interval for the estimated effect of a 20% increase in the number of inspections is (–4.4%; –6.3%).

5. Discussion

The re-analysis found a somewhat smaller effect on technical inspections of heavy vehicles than the original study. The original study estimated that doubling the number of inspections would reduce the number of accidents by 6.7% (95% CI: –18.4%; +5.1%). The re-analysis estimated that doubling the number of inspections would reduce the number of accidents by 4.9% (95% CI: –2.9%; –5.0%). The confidence interval is much smaller in the re-analysis than in the original analysis.

The replication indicates a larger effect of technical inspections. According to the replication, doubling the number of inspections would reduce the number of accidents by 21.2% (95% CI: –16.8%; –25.6%). It is reasonable to believe that technical inspections have become more effective in recent years. The roadside inspection stations have been upgraded with more advanced technology for measuring, for example, the performance of braking systems. Leaks and uneven braking forces between axles can be detected more easily than in the past.

Nevertheless, the current level of technical inspections in Norway is insufficient to eliminate technical defects. To achieve a reduction of accidents consistent with an elimination of the risk attributable to technical defects, as indicated by the population attributable risks based on Jones and Stein (1989) and Teoh et al. (2017), the number of inspections would have to increase by a factor of nine. While inspections were at a higher level than now in some years of the first period, they were never close to nine times the current level.

6. Conclusions

Technical inspections of heavy goods vehicles are associated with a reduction in the number of accidents. If inspections did not exist, there would be a higher number of accidents involving heavy goods vehicles. Most years, the number of technical inspections varies within plus or minus 20% from the previous year. Variations in this range are associated with a variation in the number of injury accidents involving heavy goods vehicles of about +8% (for a 20% reduction) to –7% (for a 20% increase).

Credit author statement

The author is the sole author of this paper.

Declaration of conflict of interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was funded by the Norwegian Ministry of Transport, grant number 5137.

References

Bjørnskau, T., & Elvik, R. (1992). Can road traffic law enforcement permanently reduce the number of accidents? *Accident Analysis and Prevention, 24*, 507–520.

Diaz Lopez, V., san Roman Garcia, J. L., Gomez Amador, A. M., Santos Cuadros, S., & Olmeda Santamaria, E. (2019). *Roadworthiness testing contribution to vehicle safety and environment*. Report. Universidad Carlos III Madrid.

Elvik, R. (2002). The effect on accidents of technical inspections of heavy vehicles in Norway. *Accident Analysis and Prevention, 34*, 753–762.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., & Krogsgård Thomsen, L. (1995). Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident Analysis and Prevention, 27*, 1–20.

Gou, M., Clément, B., Birikundavyi, S., Bellavigna-Ladoux, O., & Abraham, E. (1999). Effect of heavy-vehicle mechanical condition on road safety in Quebec. *Paper presented at the 1999 Annual meeting of the Transportation Research Board, Washington D. C.*

Jones, I. S., & Stein, H. S. (1989). Defective equipment and tractor-trailer crash involvement. *Accident Analysis and Prevention, 21*, 469–481.

Moses, L. N., & Savage, I. (1992). The effectiveness of motor carrier safety audits. *Accident Analysis and Prevention, 24*, 479–496.

Teoh, E. R., Carter, D. L., Smith, S., & McCartt, A. T. (2017). Crash risk factors for interstate large trucks in North Carolina. *Journal of Safety Research, 62*, 13–21.

Thakuriah, P., Yanos, G., Lee, J.-T., & Sreenivasan, A. (2001). Motor carrier safety. Crash patterns of inspected commercial vehicles. *Transportation Research Record, 1779*, 150–156.

Rune Elvik is a senior research political scientist. He joined the Institute of Transport Economics in 1980. He served as editor-in-chief (together with Karl Kim) of Accident Analysis and Prevention from 2005 to 2013.

Journal of Safety Research 84 (2023) 418–434

Contents lists available at ScienceDirect

Journal of Safety Research

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

Heterogeneous ensemble learning for enhanced crash forecasts – A frequentist and machine learning based stacking framework

Numan Ahmad [a], Behram Wali [b,*], Asad J. Khattak [c]

[a] Department of Civil & Environmental Engineering, The University of Tennessee, Knoxville, TN 37996, USA
[b] Urban Design 4 Health, Inc., 24 Jackie Circle, East Rochester, NY 14612, USA
[c] University of Tennessee, Knoxville, TN 37996, USA

ARTICLE INFO

Article history:
Received 12 September 2021
Received in revised form 16 July 2022
Accepted 5 December 2022
Available online 14 December 2022

Keywords:
Crash frequency
Crash prediction
Count data models
Machine learning
Base-learners
Meta-learner
Stacking

ABSTRACT

Introduction: This study aims to increase the prediction accuracy of crash frequency on roadway segments that can forecast future safety on roadway facilities. A variety of statistical and machine learning (ML) methods are used to model crash frequency with ML methods generally having a higher prediction accuracy. Recently, heterogeneous ensemble methods (HEM), including "stacking," have emerged as more accurate and robust intelligent techniques providing more reliable and accurate predictions. *Methods*: This study applies "Stacking" to model crash frequency on five-lane undivided (5 T) segments of urban and suburban arterials. The prediction performance of "Stacking" is compared with parametric statistical models (Poisson and negative binomial) and three state-of-the-art ML techniques (Decision tree, random forest, and gradient boosting), each of which is termed as the base-learner. By employing an optimal weight scheme to combine individual base-learners through stacking, the problem of biased predictions in individual base-learners due to differences in specifications and prediction accuracies is avoided. Data including crash, traffic, and roadway inventory were collected and integrated from 2013 to 2017. The data are split into training (2013–2015), validation (2016), and testing (2017) datasets. After training five individual base-learners using training data, prediction outcomes are obtained for the five base-learners using validation data that are then used to train a meta-learner. *Results:* Results of statistical models reveal that crashes increase with the density (number per mile) of commercial driveways whereas decrease with average offset distance to fixed objects. Individual ML methods show similar results – in terms of variable importance. A comparison of out-of-sample predictions of various models or methods confirms the superiority of "Stacking" over the alternative methods considered. *Conclusions and practical applications:* From a practical standpoint, "stacking" can enhance prediction accuracy (compared to only one base-learner with a particular specification). When applied systemically, stacking can help identify more appropriate countermeasures.

1. Introduction

Safety performance functions (SPFs) or crash prediction models are extensively used to predict the expected level of safety on specific roadway types. These models help evaluate the safety performance of specific countermeasures on a particular type of roadway or intersection. These practices are well discussed in the Highway Safety Manual (HSM, 2010), which presents SPFs for various roadway types (AASHTO, 2010). HSM (2010) was developed by AASHTO to provide a coherent and rigorous methodology to evaluate the safety performance on national roads (AASHTO, 2010). HSM SPFs were developed using data from specific states - and given the variations in geographical conditions, driving behaviors, and design practices nationally (AASHTO, 2010; Khattak, Ahmad, Mohammadnazar, Mahdinia, Wali, & Arvin, 2020), HSM highly recommends calibration of HSM SPFs to local conditions or developing jurisdiction-specific SPFs (AASHTO, 2010).

Traditionally, count data models (Poisson and negative binomial models) have been extensively used to model the relationships between crash frequency and key correlates, such as annual average daily traffic (AADT) and segment length (Abdel-Aty & Radwan, 2000; Caliendo, Guida, & Parisi, 2007; Hauer, Council, & Mohammedshah, 2004; Khattak et al., 2020; Mohammadnazar, Mahdinia, Ahmad, Khattak, & Liu, 2021; Shankar, Mannering, & Barfield, 1995; Srinivasan & Carter, 2011; Thakali, Fu, & Chen,

* Corresponding author.
E-mail addresses: nahmad1@vols.utk.edu (N. Ahmad), bwali@ud4h.com (B. Wali), akhattak@utk.edu (A.J. Khattak).

https://doi.org/10.1016/j.jsr.2022.12.005
0022-4375/© 2022 Published by the National Safety Council and Elsevier Ltd.

2016; Wali et al., 2018, 2020; Zhong, Sisiopiku, Ksaibati, & Zhong, 2011). Compared to Poisson models, negative binomial variants are well-suited to capture potential over-dispersion in crash data. These models provide rich inferential insights into the mechanisms through which associated factors correlate with safety outcomes. However, given the intrinsic parametric nature of the models and the subsequent assumptions, the prediction accuracy of count data models is often a concern. Growing evidence of the role of more accurate crash predictions in designing more appropriate safety countermeasures has led to an increased interest in machine learning methods. Unlike count data models, machine learning methods do not place strong restrictions on the specifications of the model (Pan, Fu, & Thakali, 2017). Machine learning methods are more adequate for modeling complex non-linear relationships that frequently arise in crash data modeling. Tree-based regression (TBR) is one of the most popular and widely-used machine learning methods that does not require variable transformations and parametric assumptions (Breiman, Friedman, Stone, & Olshen, 1984; Saha, Alluri, & Gan, 2015). TBR determines significant non-linear relationships among various predictor variables as well as computes the relative influence of predictors on response outcome (Karlaftis & Golias, 2002; Saha et al., 2015). However, the TBR technique is prone to instability leading to estimation results with higher variance (Saha et al., 2015; Zhang, Xu, Daeyaert, Lewi, & Massart, 2005). Ensemble methods like random forest regression (RFR) and gradient boosting regression (GBR) combine the estimates of numerous trees compared to a single tree, leading to improved stability and prediction accuracy (Das, Abdel-Aty, & Pande, 2009; De'ath, 2007; Saha et al., 2015). GBR technique ensembles numerous trees in a sequential way with a slower learning rate that captures a higher variance in data compared to the RFR method (Saha et al., 2015). While the prediction accuracy of machine learning methods usually is greater than the count data models, it lacks a holistic inferential framework providing little to no information about the safety mechanisms that link unsafe outcomes with key risk factors. Also, almost all of the machine learning (ML) methods explicitly relate to the bias-variance trade-off contour with different methods minimizing bias or variance. There is no escaping the relationship between bias and variance in machine learning models. Thus, the use of the single supervised or the unsupervised ML method could lead to relatively less accurate predictions.

While traditional count data models and ML methods have been extensively used in the safety literature, studies that combine the predictive (and inferential) strengths of both paradigms or the strengths of multiple ML methods are rare. The prediction performance of ML methods can be further improved by using more robust and heterogeneous ensemble methods (HEM), such as composite systems, stacking, or blending (Bhatt et al., 2017; Dietterich, 2000; Sigletos, Paliouras, Spyropoulos, & Hatzopoulos, 2005; Tewari & Dwivedi, 2020; Thapa, Gupta, Gupta, Reddy, & Kaur, 2018). HEMs including "stacking" have emerged as more accurate and reliable intelligent techniques in pattern recognition issues. The idea of "Stacking" essentially helps in harnessing the gains simultaneously from less biased and low-variance predictions offered by different ML methods. For example, the gradient-boosting regression method builds on so-called "weak classifiers" - reducing prediction error mainly by reducing bias (and to some extent variance, by aggregating the predictions from many trees). Through heterogeneous ensemble methods such as "Stacking," predictive gains from different methodologies can be combined. For example, the predictive gains from low bias in the gradient boosting method can be combined with predictive gains from lowering variance through the random-forest method via stacking. Studies also suggest other sophisticated approaches like the information Entropy-Bayesian network method to enhance crash severity prediction (Zong, Chen, Tang, Yu, & Wu, 2019). In recent years, the stacked generalization approach – a more robust and accurate ML method, has been used in transportation safety (Bugusa & Patil, 2019; Ghandour, Hammoud, & Al-Hajj, 2020; Tang, Liang, Han, Li, & Huang, 2019; Zahid et al., 2020). However, very few studies have applied this robust ML method to solve problems related to road safety (Bugusa & Patil, 2019; Ghandour et al., 2020; Tang et al., 2019; Zahid et al., 2020). Note that the stacking approach can be used in the contexts of both regression and classification problems. However, most of the aforementioned studies applied the stacking approach to solving classification (e.g., injury severity analysis) problems (Ghandour et al., 2020; Tang et al., 2019). For instance, Tang et al. (2019) applied the stacked generalization approach to predict crash severity with the severity levels of no injury, invisible injury, no-capacitating injury, and highest injury severity. The predictions obtained from three individual ML classifiers like the random forest, adaptive boosting, and gradient boosting decision tree were combined via a stacked model using logistic regression in the second layer (Tang et al., 2019). The prediction accuracy of the stacked model was significantly higher compared to individual ML methods such as random forest classifier (Tang et al., 2019). A similar stacked classification approach was used in one of the recent studies, which adopted a hybrid combination of homogeneous and heterogeneous ensemble methods to explore factors associated with fatal road crashes (Ghandour et al., 2020). That study reveals that prediction accuracy can significantly improve via the stacked generalization approach (Ghandour et al., 2020). Studies also reveal that crash risk prediction can be significantly improved via stacking predictions from individual ML algorithms (Bugusa & Patil, 2019). In another recent study, it was found that the prediction of risky and aggressive driving behavior among taxi drivers can significantly improve via stacked generalization approach compared to individual ML classifier (Zahid et al., 2020). It can be seen that stacking approach was mostly applied to solve classification problems related to transportation safety (Bugusa & Patil, 2019; Ghandour et al., 2020; Tang et al., 2019; Zahid et al., 2020).

Note that ensembles including RFR, GBR, and stacking are used to improve out-of-sample prediction accuracy and can be classified into (a) Homogeneous ensembles, and (b) Heterogeneous ensembles (Chali, Hasan, & Mojahid, 2014; Fernández-Alemán, Carrillo-De-Gea, Hosni, Idri, & García-Mateos, 2019; Rooney, Patterson, Anand, & Tsymbal, 2004; Sabzevari, Martínez-Muñoz, & Suárez, 2018). The homogeneous ensemble (e.g., RFR and GBR) uses the same feature selection algorithm with different training or learning datasets distributed over various nodes (Chali et al., 2014; Rooney et al., 2004). Instead, the heterogeneous ensemble (i.e., stacking) uses different feature selection algorithms (e.g., Poisson, Negative binomial, TBR, RFR, and GBR) where the stacking meta-learner (which can be any statistical or ML method) blends the optimal combinations of predictions by base-learners and acts as a single decision-maker in the second-stage (Chali et al., 2014; Elish, Helmy, & Hussain, 2013; Sabzevari et al., 2018). Both homogeneous and heterogeneous ensembles can be used in regression as well as classification contexts. Compared to homogeneous ensembles, heterogeneous ensembles typically show significant performance gains (Chali et al., 2014; Sabzevari et al., 2018). Both types of ensembles are used in diverse fields (such as medicine) where their application provides more accurate and reliable predictions of a specific disease in patients (Fernández-Alemán et al., 2019; Osareh & Shadgar, 2013; Petrakova, Affenzeller, & Merkurjeva, 2015). Studies suggest that heterogeneous ensembles not only outperform the conventional statistical models and other ML methods but also show superior prediction performance compared to homogeneous ensembles (Fernández-Alemán et al., 2019; Osareh & Shadgar, 2013; Petrakova et al., 2015). In transportation safety, homogeneous ensembles have been widely used for predicting crash frequency (Farid, Abdel-Aty, & Lee, 2019; Saha et al., 2015; Wang, Simandl, Porter, Graettinger, & Smith, 2016) and the severity given a crash (Iranitalab & Khattak, 2017; Yu & Abdel-Aty, 2014). Some studies used heterogeneous ensembles (e.g., stacking) in a classification context to predict injury severity (Ghandour et al., 2020; Tang et al., 2019). However, the application of heterogeneous ensemble (stacking) to predict crash frequency on roadways has not been or is very lightly explored to the best of the authors' knowledge. Given the prevalent gaps in the literature discussed above, this study contributes by:

- Applying a rigorous and robust HEM scheme to model and predict crash frequency on five-lane (5 T) undivided segments on urban and suburban arterials, including two-way left-turn lanes (2WLTL)
- Comparing the prediction performance of "Stacking" with traditional statistical models and three state-of-the-art machine learning techniques (decision trees, random forest, and gradient boosting regression).

The statistical (Poisson and negative binomial models) and ML models used in this study are considered "base-learners." To obtain valid search ranges and optimal values of corresponding tuning parameters for individual base- and stacked learners, grid-search optimization and 10-fold cross-validation procedures are used. It is shown that using more accurate, reliable, and robust intelligent techniques can extract more useful information compared to individual count data or ML methods. To achieve the study objectives, detailed crash and roadway geometric data are extracted from the Enhanced Tennessee Roadway Information Management System (ETRIMS).

2. Methodology

2.1. Conceptual architecture: Heterogeneous ensemble methods (STACKING)

The idea of HEM, including "Stacking" was first introduced almost 30 years ago (Wolpert, 1992). In stacked regression, predictions from various individual models (base-leaners) are combined and used as input in second-stage learning (Güneş, Wolfinger, & Tan, 2017). Stacking generally provides higher prediction accuracy compared to base-learners (Güneş et al., 2017). Suppose Y is the response outcome, X is the set of predictors used in individual models (briefly discussed in subsequent sections), and $g_1, g_2, \cdots, g_L$ are the predictions obtained using base-learners (Güneş et al., 2017). The prediction function for the linear ensemble (stacked) model can be given as (Güneş et al., 2017):

$$b(g) = (w_1 * g_1) + (w_2 * g_2) + \cdots + (w_L * g_L) \quad (1)$$

Note that w_i indicates the weight assigned to an individual model in the stacking technique (Güneş et al., 2017). The model weights (w_i) are used to minimize MSE between actual response variable (y_i) and prediction outcome of meta-learner (stacked ensemble technique) as shown (Güneş et al., 2017):

$$\min \sum_{i=1}^{N} \Big(y_i - \big(w_1 * g_{1i} + w_2 * g_{2i} + \cdots + w_1 * g_{Li}\big)\Big)^2 \quad (2)$$

The conceptual design of this study is presented in Fig. 1. First, we manually extracted crash, traffic, and roadway geometry data using various software made available by the Tennessee Department of Transportation (TDOT) for a randomly selected subsample containing 304 roadway segments of 5 T urban and suburban arterials for a period of five years (2013–2017). Next, we split data into training (2013–2015), validation (2016), and testing (2017) datasets (Fig. 1). Note that in all the three datasets, only crashes and average annual daily traffic may change while all other factors remain the same. We follow this splitting procedure to develop a crash prediction model that can be reused with updated data to forecast crashes in the future. First, five individual base-learners are trained using training data to model crash frequency per year (2013–2015). Next, prediction outcomes obtained from these five base-learners using the validation dataset are obtained and combined with actual crashes reported in 2016, which generates a new training dataset for the meta-learner (stacking). Note that grid-search optimization and 10-fold cross-validation procedures are used to obtain valid search ranges and optimal values of corresponding tuning parameters for individual machine learning techniques and stacking (Fig. 1). In the 10-fold cross-validation procedure, an algorithm splits the available data (used to train the model) into 10 subsamples of equal sizes, and nine of those subsamples are used for training, while one subsample is used for testing to determine the optimal model for prediction accuracy. The algorithm repeats the process 10 times, during which each of the subsamples is once used as a testing subsample. The results are finally averaged to get a single estimation. Note that studies commonly use a 10-fold cross-validation procedure for the tuning of the machine learning models (Mclachlan, Do, & Ambroise, 2005). Finally, we apply individual base-learners (trained using the training dataset) and meta-learners or the stacked model (trained using validation dataset) to the new data (2017) to accurately compare their prediction performance (Fig. 1).

This study applies stacking where a meta-learner is used to combine multiple predictions obtained from various base-learners, as explained below.

- **Base-learner:** Stacking is a two-stage process where individual statistical models and/or ML methods are applied in the first stage. Any statistical model or ML method when applied in the first stage of stacking is termed a "base-learner" in this study. For instance, this study applies five base-learners, which include two statistical models (Poisson and Negative binomial) and three ML methods (TBR, RFR, and GBR). The base-learners applied in this study also include homogeneous ensembles (RFR and GBR), which use the same feature selection algorithm with different training datasets. In homogeneous ensembles, the results and/or predictions are averaged.
- **Meta-learner:** The stacking meta-learner algorithm is an ensemble technique that combines predictions from two or more than two base-learners specifically to further enhance prediction accuracy. This study uses three ML methods including TBR, RFR, and GBR as meta-learners to combine predictions for the five base-learners (Poisson, negative binomial, TBR, RFR, and GBR). Finally, after comparing the out-of-sample RMSE and MAE of all the base-learners and three meta-learners, one model was selected that has the lowest out-of-sample RMSE and MAE. Note: that stacking is termed as a "heterogeneous ensemble" that combines different feature selection procedures (Poisson, Negative binomial, TBR, RFR, and GBR). In stacking, a meta-learner can also be termed a super-learner (Van Der Laan, Polley, & Hubbard, 2007).

2.2. Count data models: Poisson and negative binomial regression

Studies suggest using count data (Poisson and negative binomial regression) models to explore the relationship of the crash frequency with explanatory variables (Anastasopoulos & Mannering, 2009; Wali et al., 2018, 2020). Poisson regression was first introduced by a French mathematician named Siméon-Denis Poisson in 1830. The mathematical formula of Poisson

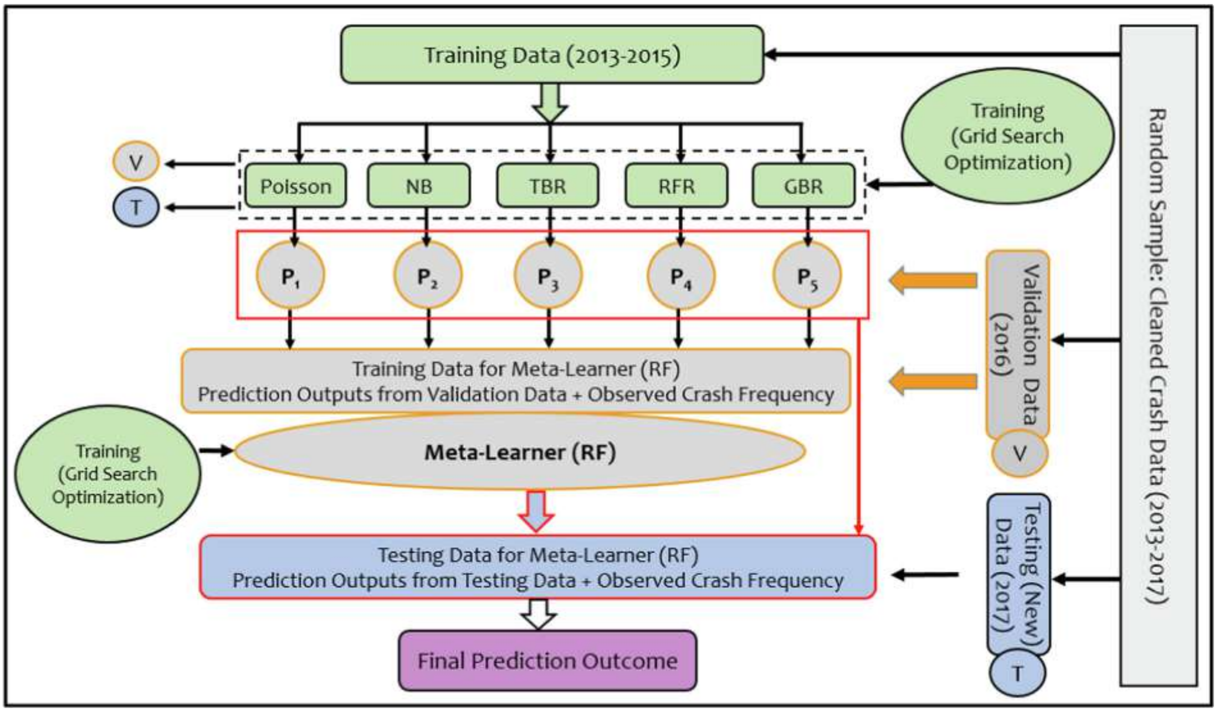


Fig. 1. Conceptual Design of Stacking Ensemble Utilized for Crash Frequency Modeling. **Notes:** In Fig. 1, NB indicates a negative binomial model, TBR indicates a Tree-based regression, RFR indicates a random forest regression, and GBR indicates a gradient boosting regression. P_1, P_2, P_3, P_4, and P_5 are the prediction outcomes obtained while applying Poisson, NB, TBR, RFR, and GBR models to the validation dataset, respectively. V and T indicate validation and testing datasets, respectively

regression is given below (Washington, Karlaftis, & Mannering, 2010).

$$P(n_i) = \frac{exp(-\lambda_i)\lambda_i^n}{n_i!} \tag{3}$$

where $P(n_i)$ is the probability of a crash occurring on a specific road segment (i), (n) is the frequency of a crash on a specific road segment at a particular time, and (λ_i) is the expected number of crashes occurring on a particular road segment (i) in a specific duration. The expected number of crashes (λ_i) is linked to its key contributing factors as below (Washington et al., 2010; Wali, Khattak, Waters, Chimba, & Li, 2018; Wali, Ahmad, Khattak, & Nazar, 2020):

$$ln(\lambda_i) = \beta(X_i) \tag{4}$$

where X_i indicates a set of explanatory variables, and β are their associated parameter estimates.

The equations (3) and (4) can be maximized using the standard maximum likelihood procedure (Washington et al., 2010; Wali et al., 2018):

$$L(\beta) = \prod_i^n \frac{exp[-exp(\beta X_i)][exp(\beta X_i)]^{n_i}}{n_i!} \tag{5}$$

In the case of over-dispersion, Poisson regression is not preferable due to violation of its basic assumption, therefore negative binomial regression is suggested as below (Washington et al., 2010; Wali et al., 2018).

$$ln(\lambda_i) = \beta(X_i) + \in_i \tag{6}$$

where $exp(\in_i)$ is an error term with gamma distribution "mean equals one and variance (α)" (Washington et al., 2010; Wali et al., 2018). The conditional probability for crashes can be given as (Poch & Mannering, 1996; Wali et al., 2018):

$$P(\in) = \frac{exp\,[-\lambda_i exp\,(\in_i)\,]\,[\lambda_i exp\,(\in_i)\,]^{ni}}{n_i!} \tag{7}$$

The error term $(\in_i)$ can be integrated out to determine the unconditional distribution of n_i as given below (Poch & Mannering, 1996; Wali et al., 2018):

$$P(n_i) = \frac{\Gamma(\theta + n_i)}{[\Gamma(\theta).n_i!]}.u_i^\theta(1-u_i)^{n_i} \tag{8}$$

where u_i equals $\theta(\theta+\lambda_i)$ and $\theta = \frac{1}{\alpha}$, and Γ is a gamma function. In the case of (α) approaching zero, the negative binomial simply becomes a Poisson regression (Washington et al., 2010). Negative binomial regression is preferred over Poisson regression when it is significantly different from zero (Anastasopoulos & Mannering, 2009; Washington et al., 2010; Saeed, Hall, Baroud, & Volovski, 2019). To evaluate the goodness of fit performance of the count data models, McFadden R^2 value (Wali et al., 2018), Akaike Information Criteria (Akaike, Petrov, & Csaki, 1973; Bozdogan, 1987; Wali et al., 2018), and Bayesian Information Criteria (Schwarz, 1978; Wali et al., 2018) can be used.

2.3. Machine learning methods

2.3.1. Decision-tree regression

The decision tree uses a fast algorithm that recursively splits training data into smaller subsets (Torgo, 1999). However, instability and reliability issues are key weaknesses of this method (Breiman, 1996; Torgo, 1999). The algorithm searches to determine a splitting point with the lowest value of mean square error (MSE). At the optimal splitting point, the parent node is further split into two child nodes and the process continues until the optimal tree length is determined (reducing impurity associated with the terminal node). The algorithm chooses the best splitter (S^*) considering deviance (D) or MSE at a particular node as:

$$D(t) = \sum_{x \in t}(Y_n - \hat{\mu})^2 \tag{9}$$

where, $\hat{\mu}$ is a sample mean $(\bar{y})$ or mean estimate, t indicates a specific node, and X indicates a set of predictors. Referring to the generalized linear models, deviance (D) is also termed log-likelihood ratio statistics and can be written as:

$$D = 2 * l(\mu_{max}; y) - l(\hat{\mu} : y) \tag{10}$$

where, μ_{max} is the maximum likelihood estimate. Deviance of a tree (T) can be determined as below:

$$D(T) = \sum_{t \in \hat{\tilde{T}}} D(t) = \sum_{t \in \hat{\tilde{T}}} \sum_{x \in t} \left(Y_n - \bar{y}(t)\right)^2 \tag{11}$$

where T is the tree, $\hat{\tilde{T}}$ is a set of terminal nodes of T. For a binary partition via splitter (s), the difference is:

$$\Delta D(s,t) = D(t) - D(t_L) - D(t_R) \tag{12}$$

where t_L and t_R indicate left and right child of the parent node (t) respectively. The difference is maximized to determine the best splitter (s^*) as:

$$\Delta D(s^*, t) = max_{s \in S} \Delta D(s,t) \tag{13}$$

Note size selection and tree pruning are carried out using 10-fold cross-validation to select the optimal tree size with the lowest MSE.

2.3.2. Random forest regression

Studies suggest ensemble methods like RFR and GBR to mitigate instability issues related to a single decision (Louppe, Wehenkel, Sutera, & Geurts, 2013; Malekipirbazari & Aksakalli, 2015). The RFR algorithm works on a similar principle to the single decision tree; however, the key difference is that RFR assembles an enormous number of trees. The RFR algorithm selects a predictor at each node to maximize homogeneity at successive nodes (Hastie, Tibshirani, & Friedman, 2009; Liaw & Wiener, 2002). Regularization parameters considered for RFR include (Chung, 2013; Heung, Bulmer, & Schmidt, 2014; Liaw & Wiener, 2002):

- Number of predictors selected at each node for split-up $(\boldsymbol{m_{try}})$
- Number of trees in the forest (n_{tree})
- Number of maximum nodes in the forest

Studies suggest using the following trials to select the optimal number of predictors $(\boldsymbol{m_{try}})$ at each node (Breiman, 1996):

- $m_{try} = \frac{p}{3}$
- $m_{try} = \frac{1}{2} * \frac{p}{3}$
- $m_{try} = 2 * \frac{p}{3}$

While p is the total number of predictor variables considered in RFR regression. Note that $\boldsymbol{m_{try}}$ indicates the number of variables/predictors available for splitting at each node. It is considered an important regularized or tuning parameter (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008). To determine the optimal value of $\boldsymbol{m_{try}}$, we use an extended grid-search optimization and 10-fold cross validation procedure. To select an optimal pair of n_{tree} and $\boldsymbol{m_{try}}$, two performance criteria including MSE and R^2 values are usually used (Liaw & Wiener, 2002):

$$MSE \approx MSE_{OOB} = \frac{1}{n} \sum_{i \in OOB} (y_i - \hat{y}_i)^2 \tag{14}$$

$$R^2 = 1 - \frac{MSE_{OOB}}{Var(y_i)} \tag{15}$$

where MSE_{OOB} is the MSE for the out-of-bag (OOB) sample, y_i is the observed number of crashes occurring on i^{th} roadway segment in OOB sample, $\hat{y}_i$ is predicted crashes on i^{th} road segment in OOB sample, n is the number of roadway segments in OOB sample, $Var(y_i)$ is the variance of response outcomes (y) determined as $\frac{1}{n}\sum_{i \in OOB}\left(y_i - \bar{y}\right)^2$, while $\bar{y}$ is the mean value of y_i in the OOB sample.

Similar to the TBR approach, variable importance relates to the reduction in node impurity at each split; however, the RFR technique uses the average reduction of all trees in the forest to determine the overall reduction in impurity. Importance $Imp(X_m)$ of any particular predictor variable X_m, is computed while summing the weighted reduction in node impurities, $\Delta_i(s,t)_{X_m}$, for all nodes t where X_m is used for splitting (Louppe et al., 2013):

$$Imp(X_m) = \frac{1}{N_T} \sum_{T} \sum_{t \in T} \frac{N_t}{N} \Delta_i(s,t)_{X_m} \tag{16}$$

where N_T is the number of trees, N_t is the number of data points at a specific node (t), and N is the sample size.

2.3.3. Gradient boosting regression

Similar to the RFR approach, GBR is a pool procedure to enhance prediction accuracy (De'ath, 2007; Elith & Leathwick, 2017; Hastie et al., 2009). The algorithm calculates residuals after fitting the first tree to the l} due to which the GBR algorithm assigns more weight to such observations while fitting the next tree and so on (Saha et al., 2015). In this straightforward and stagewise process, the GBR algorithm keeps the existing tree unchanged while re-estimating residuals for every observation to reveal contributions to the new tree (Elith & Leathwick, 2017; Saha et al., 2015). Let $f(x)$ be an approximation function of response outcome (y) as predicted by a set of predictor variables (x). In GBR approach, an additive expansion of the basic functions $(x : \gamma_m)$ can be given as:

$$f(x) = \sum_m f_m(x) = \sum_m \beta_m b(x : \gamma_m) \tag{17}$$

Note that $\beta_m(m = 1, 2, 3, \cdots, M)$ indicates the expansion coefficients, $b(x : \gamma_m)$ indicates single regression trees having parameter (γ_m) as a split variable, and β_m are the weights assigned to every tree (Saha et al., 2015). The algorithm estimates parameters like β_m and γ_m to minimize loss function $L(y(f(s))$ indicating prediction performance in term of deviance (Saha et al., 2015). Note that while GBR may nicely fit the data, it can also lead to overfitting (Saha et al., 2015). To cure this issue, studies suggest selecting appropriate regularization parameters including the number of trees, shrinkage (learning rate), the minimum number of observations in the tree's terminal nodes, and complexity which help in achieving a balance between variance and bias (Saha et al., 2015). The learning rate is usually smaller ranging from 0.0001 to 0.1 (Saha et al., 2015). Note that smaller values of shrinkage parameters are good but require more trees. The complexity parameter refers to tree depth which shows interactions among predictor variables (Saha et al., 2015).

2.4. Model performance

To evaluate the prediction performance of individual models (Poisson, negative binomial, TBR, RFR, and GBR) and stacked regression, we compare their Root Mean Square Error (RMSE) (Wali et al., 2018) and Mean Absolute Error (MAE) (Wali et al., 2018; Washington, Karlaftis, & Mannering, 2010) based on the testing dataset (2017):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2} \tag{18}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e| \tag{19}$$

The value of n is the total number of roadway segments, and f_i and y_i indicate predicted and observed crash frequency, respectively. Low values of RMSE and MAE indicate higher prediction accuracy.

3. Results and discussion

3.1. Data processing and descriptive statistics

Data used in this research was extracted from the Enhanced Tennessee Roadway Information Management System (ETRIMS), which is a roadway inventory and crash database maintained by the Tennessee Department of Transportation (TDOT). We identified the five-lane (5 T) roadway segments of urban and suburban arterials by selecting the attributes of interest including the number of through lanes (five lanes), the presence of two-way-left-turn lanes (2WLTL), and functional class (urban arterials). The roadway segments are pre-defined in ETRIMS where a segment refers to a portion of the roadway that either connects two nodes (i.e., intersections) or has uniform features (e.g., lane width, shoulder width, number of lanes, and median width) as compared to neighboring (proceeding and succeeding) roadway segments. ETRIMS showed a total of 3,208 (753.97 miles) segments of state-maintained 5 T urban and suburban arterials in Tennessee. Fig. 2 shows the distribution of the total 5 T roadway segments (N = 3,208) of urban and suburban arterials in TN identified in ETRIMS, which was first cleaned and then a random sample (N = 304) was selected for analysis.

Following the HSM guidelines (AASHTO, 2010), segments shorter than 0.1 mile were removed leading to a reduced dataset containing 1,519 segments (totaling 523.93 miles). First, we determined the sample size to be selected from the population (1,519 segments) using 95 % confidence level criteria. A random sample of 317 segments (105.78 miles) was selected for which crash (2013–2017), roadway geometry, and traffic data (2013–2017) were extracted using ETRIMS and TDOT Traffic History Application. Finally, 304 (103.27 miles) segments with complete data are considered in the analysis. The distribution of roadway segments of 5 T urban and suburban arterials (random sample "N = 304") in Tennessee based on the total number of crashes that have occurred on these segments during the five-year (2013–2017) period is shown in Fig. 3. The segments with a higher number of total crashes during the five years are mostly located in Region 3, which contains the Nashville area (Fig. 3). Notably, segments with a low number of crashes over the 5 years are mostly located in the suburbs of the major cities or other urban areas and small cities (Fig. 3). Note that each circle refers to a roadway segment of 5 T urban and suburban arterials in TN, and the size of the circle depicts the total number of crashes that have occurred on a roadway segment during the five-year (2013–2017) period.

A study by AASHTO revealed that factors like density (number per mile) of major/minor driveways based on various land uses (e.g., commercial and industrial) and average offset distance to fixed objects significantly influence crash frequency on a roadway segment (AASHTO, 2010). Fixed objects (utility poles, traffic signs, trees, and billboards) along roadway segments are considered potential safety risks (Albuquerque & Awadalla, 2019; Albuquerque & Awadalla, 2020; Safety, 2011; Wolf, 2006). Such objects are more prevalent along with urban roadway segments (Albuquerque & Awadalla, 2019, 2020; Wolf, 2006). Distance to fixed objects along roadway segments is critical as the risk of fixed-object collisions increases as the offset distance to roadside fixed objects decreases (AASHTO, 2010). In HSM (2010), SPFs for all types of urban roadways include offset to roadside fixed objects as an important factor to predict crashes on specific roadway segments (AASHTO, 2010). We consider it important to include the average offset distance to fixed objects along with the roadway segments of 5 T urban and suburban arterials in the models. The offset distance (measured in feet) to every fixed object along the roadway segment may vary; therefore, we calculated and used the average of the offset distances to fixed objects along with the roadway segments in the models.

To achieve the study objective, the data are split into three subsets: training, validation, and testing. Table 1 presents descriptive statistics of key variables. Statistics reveal an average of 11.026 crashes (standard deviation of 14.020) across the three years on 5 T segments of urban and suburban arterials. Crash distributions for validation (2016) and testing (2017) are shown in Table 1, revealing similar distributions across the three (training, validation, testing) streams. Statistics for traffic measures and roadway geometric features are provided in Table 1. Table 1 provides the distribution of AADT in the three subsamples including training, validation, and test samples; however, we have computed the vehicle miles traveled (VMT) in millions (=segment length*AADT*365*10^{-6}), which is used in the analyses. In 2017, the mean AADT was 19,903, which is slightly higher than the yearly AADT in 2016 and the average AADT per year from 2013

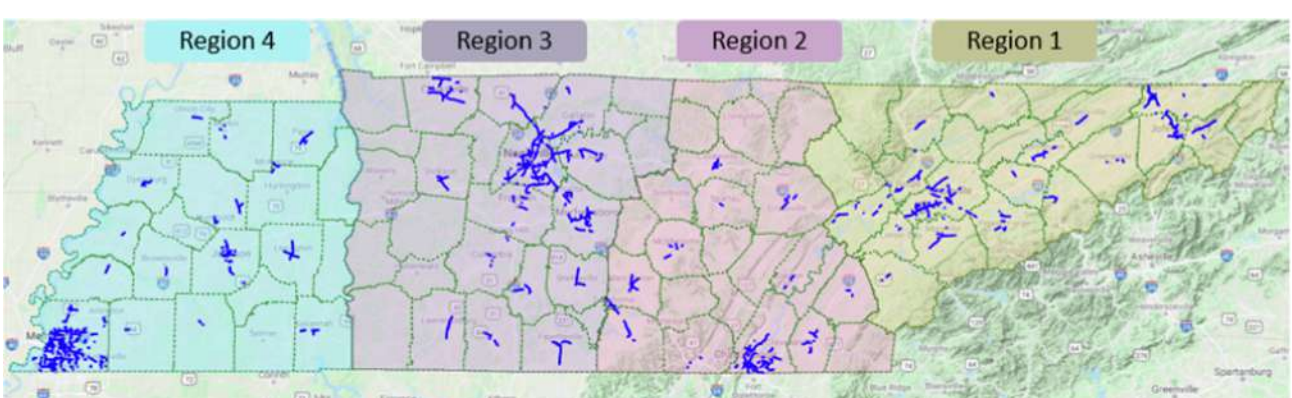


Fig. 2. Distribution of the Overall 5 T roadway segments of Urban and Suburban arterials in TN. **Note:** Tennessee has 95 counties, which are divided into four TDOT regions, as shown on the map.

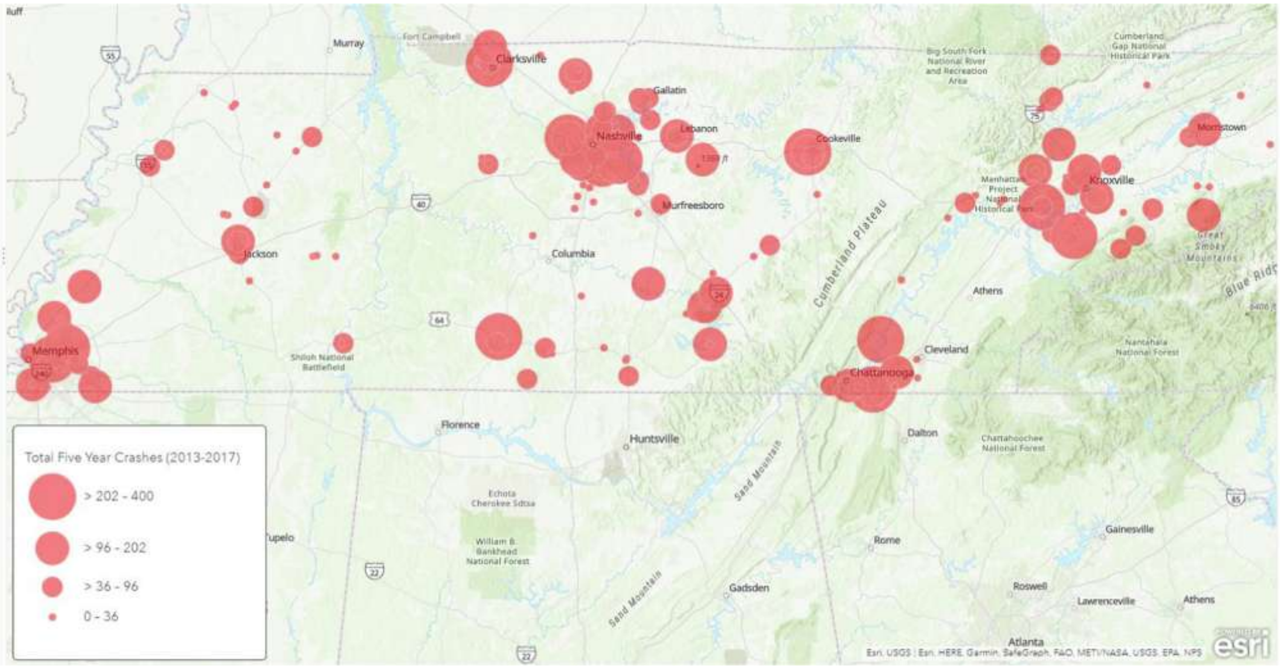


Fig. 3. Distribution of 5 T Urban and Suburban Arterial Segments based on Five-Year (2013–2017) Crashes in TN.

Table 1
Descriptive statistics of key variables: 5 T segments of urban and suburban arterials.

Variables	Obs.	Mean	Std. Dev.	Min	Max
Average Three-years Crashes (2013–15)	304	11.026	14.020	0.000	72.000
Total Crashes (2016)	304	11.010	14.651	0.000	90.000
Total Crashes (2017)	304	11.072	14.618	0.000	100.000
AADT Per Year (2013–15)	304	19,098	8937.771	3182	49,766
AADT (2016)	304	19,644	9209.835	3613	54,360
AADT (2017)	304	19,903	9153.563	3811	54,564
Segment length (mile)	304	0.340	0.279	0.100	1.809
VMT Per Year (2013–15) (in Millions)	304	2.326	2.359	0.116	19.823
VMT (2016) (in Millions)	304	2.390	2.409	0.132	19.769
VMT (2017) (in Millions)	304	2.414	2.439	0.139	20.989
Density (frequency per mile) of Commercial Driveways	304	1.214	1.941	0	12
Density (frequency per mile) of Industrial Driveways	304	1.747	2.167	0	12
Density (frequency per mile) of Residential Driveways	304	1.243	2.992	0	30
Average Offset Distance (feet) to Roadside Fixed Objects	304	14.266	8.188	0.000	30.000

Note: In Table 1, AADT and VMT stands for annual average daily traffic and vehicle miles travelled respecitvley both of which have already been defined in the texts above. Furthermore, obs., Std. Dev., Min, and Max refer to number of observations, standard deviation, minimum, and maximum, respectively.

to 2015 (Table 1). This shows that, on average, AADT per year has increased slightly compared to the previous years. The sample statistics show the mean segment length to be 0.340 miles, including no segment with a length less than 0.1 miles (Table 1). Referring to VMT (millions), the mean VMT (millions) per year in the training sample is found to be 2.326, which is lower than the corresponding values in the validation (2.390) and test (2.414) samples. The mean offset distance to roadside fixed objects is found to be 14.26 feet (Table 1). Referring to the density (i.e., frequency per mile) of driveways based on various land uses, the density of industrial (including both minor and major) driveways was found to be the highest density with a mean value of 1.747 driveways per mile (along both sides of the roadway segment) followed by residential (including both minor and major) driveways (1.243 per mile) and commercial driveways (1.214 per mile), as shown in Table 1. The density of minor and major driveways (for each of the three types) in Tennessee was not extensive due to which this study has considered the density of total (including both minor and major) driveways for each of the three land uses (residential, commercial, and industrial) without splitting them into minor and major categories. The descriptive statistics seem reasonable because the dataset contains little to no outliers.

3.2. Estimation results

3.2.1. Count data models: Poisson and negative binomial regression

As a first step, we apply Poisson and negative binomial models to explore the average three-year (2013–2015) crash frequency. Both models come from a series of trials evaluated based on statistical significance, parsimony, and intuition. To select more appropriate models (with superior fit), several trials were made based on the specifications of explanatory variables. Initially, Poisson and negative binomial models were estimated including all the significant variables (including VMT in millions) in their original forms (Model 1 and Model 2). Next, logarithmic forms of VMT (in millions) were included while keeping all other covariates (e.g., the density of commercial and residential driveways and average offset to roadside fixed objects) in their original forms (Model 3 and Model 4). Finally, logarithmic forms of all significant variables were tested. Including logarithmic forms of all variables in the model did not lead to improvements (results not shown for brevity). Poisson and Negative Binomial models with log-transformed VMT (in millions) variables (Model 3 and Model 4) outperformed their counterparts with untransformed variables based on AIC, BIC, and log-likelihood values at convergence (Table 2). Thus, Model 3 and Model 4 (including logarithmic forms of VMT in millions) were selected as the best models compared to their counterparts. Similar specifications for the key variables (ln forms of VMT) were used while training machine learning methods. To quantify the effects of significant variables on crash frequency, we present marginal effects (MEs) in Table 2. According to the estimation results, VMT (2013–2015) per year was positively correlated with the average three-years crash frequency (Table 2). In terms of geometric factors, density (number per mile) of commercial (including both minor and major) driveways is also positively correlated with crash frequency on 5 T segments of urban and suburban arterials (Table 2). On other hand, the density of residential driveways showed a negative correlation with crash frequency on 5 T segments of urban and suburban arterials (Table 2). It is important to mention that the density of industrial or institutional driveways was also tried in the model, but it did not show any statistical significance and was therefore excluded. We also found that the average offset distance (feet) to fixed objects along these segments is negatively associated with average three-years (2013–2015) crash frequency (Table 1). The over-dispersion parameter in negative binomial models is found to be statistically significant, indicating that the negative binomial model is preferred over the Poisson regression (Table 2).

To understand the relationship between key variables and crash frequency, we discuss the marginal effects of variables for the best statistical model (negative binomial model with VMT (millions) in ln forms), which has the best in-sample fit (Table 2). Our findings indicate that yearly crash frequency increases by almost 6.851 units with a unit increase in yearly VMT in millions (ln form) while keeping all other variables at their means (Table 2). The estimation results of the best-fit model suggest that commercial driveways have a stronger association with crash frequency (i.e., a unit increase in density of commercial driveways is associated with an increase in yearly crashes by 0.728 units; Table 2). Moreover, yearly crash frequency is lower by 0.375 with a unit increase in the density of residential driveways (Table 2). Other studies suggest similar findings (AASHTO, 2010; Dixon, Avelar, Brown, Mecham, & Van Schalkwyk, 2012; Khattak et al., 2020). These findings were expected as an increase in commercial driveways increases potential conflict points and creates a potential for gap acceptance errors. On other hand, the findings related to the density of residential driveways were expected as there could be lower traffic coming from these driveways (compared to commercial driveways) thus having a lower chance of potential crashes with the through traffic compared to commercial driveways. These findings highlight the need for investigating proactive access management strategies that can potentially reduce crashes specifically on 5 T roadway segments of urban and suburban arterials. Our findings indicate that a higher offset distance to roadside fixed objects is associated with fewer crashes. Crash frequency is lower by 0.229 with a unit increase in average offset distance (feet) to roadside fixed objects (Table 2). This was expected as fixed objects (i.e., utility poles, traffic signs, trees, and billboards) along roadway segments are potential safety risks specifically for errant vehicles (Albuquerque & Awadalla, 2019; Albuquerque & Awadalla, 2020; Safety, 2011; Wolf, 2006). Such objects are more prevalent along with urban roadway segments (Albuquerque & Awadalla, 2019, 2020; Wolf, 2006). The distance to fixed objects along the roadway segments is a critical factor because the risk of fixed-object collisions is lower with higher offset distances to roadside fixed objects (AASHTO, 2010).

3.2.2. Machine learning techniques

3.2.2.1. Single decision tree regression. First, we apply a single TBR to predict average crash frequency per year on 5 T urban and suburban arterials using a training dataset (2013–2015). Using one standard-error rule, we do not observe a significant reduction in

Table 2
Estimation results of Poisson and negative binomial models.

Variables	Poisson (Model 1)			Negative Binomial (Model 2)			Poisson (Model 3)			Negative Binomial (Model 4)		
	Data (2013–2015)			Data (2013–2015)			Data (2013–2015)			Data (2013–2015)		
	Coeff.	t-stat	MEs	Coeff.	t-stat	MEs	Coeff.	t-stat	MEs	Coeff.	t-stat	MEs
Constant	2.229	54.81	–	1.854	15.26	–	2.169	50.23	–	1.996	17.79	–
VMT Per Year (2013–15) in Millions	0.155	32.38	1.402	0.256	10.67	2.149	–	–	–	–	–	–
Density (frequency per mile) of Commercial Driveways	0.103	16.58	0.931	0.129	4.61	1.084	0.053	8.05	0.421	0.093	3.51	0.728
Density (frequency per mile) of Residential Driveways	−0.032	−5.15	−0.289	−0.042	−2.19	−0.353	−0.039	−6.64	−0.307	−0.048	−2.74	−0.375
Average Offset Distance (feet) to Roadside fixed objects	−0.033	−12.96	−0.298	−0.030	−4.46	−0.250	−0.038	−14.72	−0.302	−0.029	−4.60	−0.229
Key Variables (ln form)												
VMT Per Year (2013–15) in Millions(ln form)	–	–	–	–	–	–	0.875	35.93	6.912	0.873	12.58	6.851
Over-dispersion Parameter	–	–	–	1.446	10.33	–	–	–	–	1.743	10.08	–
Summary												
Sample Size	304			304			304			304		
Log likelihood at Convergence	−1707.267			−963.620			−1465.016			−939.771		
AIC	3424.533			1939.241			2940.032			1891.542		
BIC	3443.118			1961.543			2958.617			1913.844		

Note: AIC is Akaike Information Criterion, BIC is Bayesian Information Criterion, while MEs indicate marginal effects. Furthemore, Coeff. Refers to the coefficients. The variance inflation factor (VIF) for each of the predictor variables in Model 4 was computed which shows no significant sign of multicollinearity among the predictor variables. The VIF for VMT (millions) (ln form), the density of commercial driveways, the density of residential driveways, and offset distance to fixed objects is found to be 1.3191, 1.2397, 1.1361, and 1.0608, respectively.

error after a tree size of 7 (with cost complexity ~ 0.01328895). Using the mentioned optimal values of tuning parameters, an optimal tree is grown as shown on the right side in Fig. 4. The key predictor variables used in developing the optimal tree include VMT (2013–2015) per year in millions (ln form), the density of commercial driveways, and offset distance to fixed objects (Fig. 4). Note that the single decision tree is easily interpretable. For instance, it can be seen that if VMT (2013–2015) is lower than $1.00*10^6$ ($e^{0.0032}$ = 3.74 VMT in millions), the estimated number of crashes on average is 3.13 (Fig. 4). The optimal TBR model may assign only 1 of the 12 values (3.13, 6.57, 9.60, 4.83, 18.54, 13.58, 29.57, 56, 18.86, 40, 50, 53.1) of crashes to roadway segments based on the attributes (mean VMT (millions), the density of commercial driveways, and average offset distance to fixed objects) selected by the optimal tree-based regression model. Note that logarithmic forms of VMT (millions) along with other key covariates (e.g., the density of commercial driveways and average offset distance to fixed objects) in their original forms were used to train the tree-based model. Once the results from tree-based regression were obtained, we took the anti-log of the values of VMT (millions) to interpret the results – as shown in Fig. 4).

Note: In Fig. 4, the values with which "in millions" is written refer the values already converted and shown on a million scale e.g., 1.00 "in millions" refers to a value of 1,000,000 on an actual scale which when multiplied with 10^{-6} results into 1.00 "in millions."

3.2.2.2. Random forest regression. To select optimal values of tuning parameters including the number of predictors considered at each split, the number of trees, and the maximum number of nodes in the random forest, an extended grid-search optimization and 10-fold cross-validation procedure were used (Fig. 5). In the grid search, we assigned different values to each of the tuning parameters, that is, different values of predictors considered at each split, number of maximum nodes, and number of trees assigned were 2 to 4 (with an increment of 1), 1 to 31 (with an increment of 2), and 250 to 1,000 (with an increment of 50) leading to 768 different combinations of the three tuning parameters. Based on RMSE, our comprehensive grid search for all 768 different combinations of the three tuning parameters indicates that optimal values for the number of predictors considered in each split, number of maximum nodes, and number of trees are found to be 2, 23, and 900, respectively (Fig. 5). Fig. 5 shows the RMSE only for the top 20 best combinations of the three tuning parameters (mtry, number of maximum nodes, and number of trees), which showed smaller RMSE compared to the remaining 748 combinations of the three tuning parameters. Using these tuning parameters, we apply the RFR model to predict crash frequency per year using training data (Fig. 5).

The relative importance of predictor variables used in the final random forest model is illustrated in Fig. 6. On basis of relative importance, VMT per year (2013–2015) and density of commercial driveways are found to be the most important predictor variables (Fig. 6). Similarly, average offset distance to fixed objects and density of residential driveways are ranked 3rd and 4th as per the final RFR model using their relative importance (Fig. 6).

3.2.2.3. Gradient boosting regression. As discussed earlier, GBR is prone to overfitting, which can be minimized while achieving a balance between variance and bias through the selection of optimal regularization parameters, such as the number of trees, learning rate (shrinkage), the minimum number of observations in trees' terminal nodes, and complexity parameter (interaction depth). Again, extended grid search and 10-fold cross-validation procedures are used to select optimal values of the regularized parameters. After conducting a grid search with all possible combinations of the number of trees, shrinkage, and interaction depth, a minimum RMSE is achieved when the number of trees, shrinkage, and complexity parameters are equal to 100, 0.1, and 3, respectively (Table 3). The performance of some key combinations of regularization parameters is shown in Table 3.

Once the optimal values of the regularization are determined, a final GBR model is trained. The relative importance of key variables in predicting crash frequency per year on 5 T segments of urban and suburban arterials is shown in Fig. 7. Similar to the RFR model, average three-year VMT (2013–2015) in millions and density of commercial driveways are the most important predictor variables (Fig. 7). Moreover, average offset distance to roadside fixed objects and density of residential driveways are ranked 3rd and 4th in terms of their relative importance in predicting crash frequency (Fig. 7).

Note that GBR (the best performing base-learner) provides variable importance but does not show the magnitude or nature of the relationship between the response outcome and specific explana-

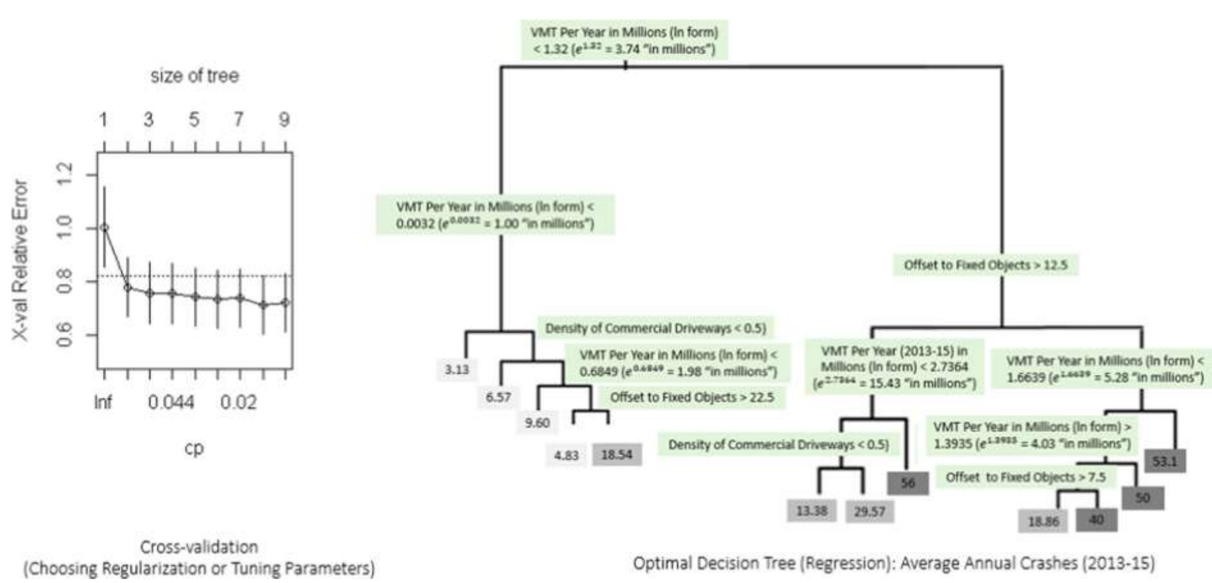


Fig. 4. Illustration of Cross validation (Regularization) and Optimal Decision-Tree Regression.

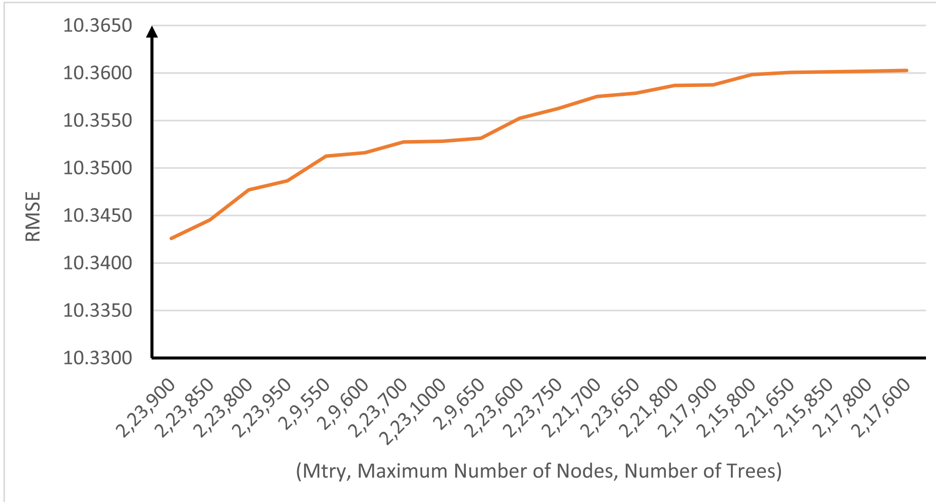


Fig. 5. Selecting Optimal Values of Regularization Parameters for Random Forest.

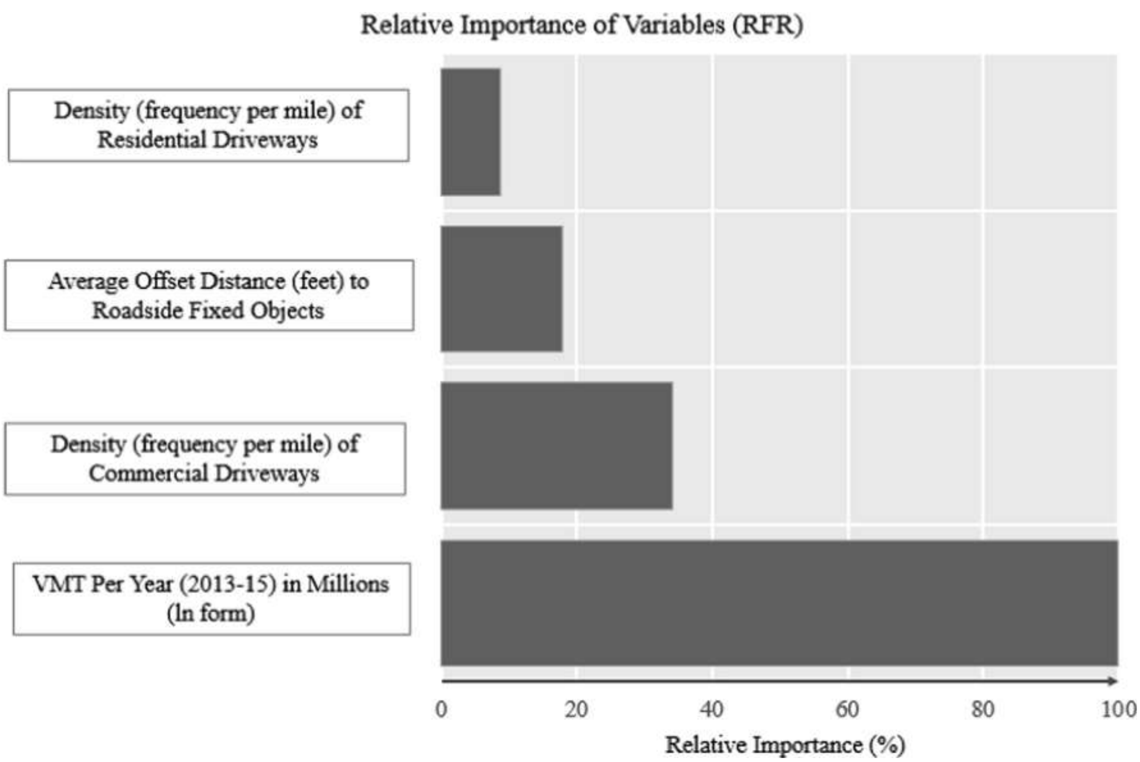


Fig. 6. Variables Relative Importance Plot: Optimal RFR (Base-learner).

Table 3
Selecting optimal combination of regularization parameters for gradient boosting.

Shrinkage	Interaction Depth	Minimum Number of Observations in Trees' Terminal Nodes	Number of Trees	RMSE
0.1	3	10	100	10.4056
0.1	1	10	100	10.4818
0.1	3	5	100	10.5222
0.1	10	10	100	10.5307
0.1	7	10	100	10.5344
0.1	1	5	100	10.5702

Note: The above six combinations are the combinations with smaller RMSE (root mean square error) compared to all other combinations. Note that in our grid search, we assigned a range of values to shrinkage (0.1 to 1 with an increment of 0.2), interaction depth (1, 3, 7, and 10), the minimum number of observations in trees' terminal nodes (2, 5, 10) and the number of trees (100, 300, 500, 1000).

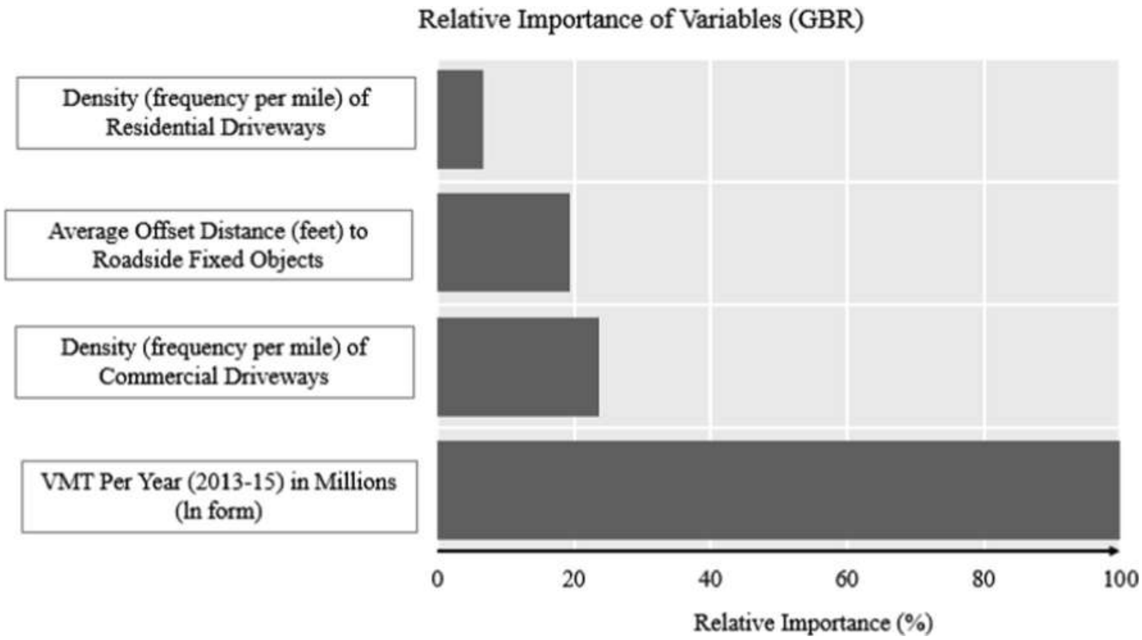


Fig. 7. Variables Relative Importance Plot: Optimal GBR (Base-learner). **Note:** In the figure above, the GBR refers to gradient boosting tree regression.

tory variables (Friedman, 2001). We present the partial dependence plots, which are similar to marginal effects in statistical models, for the two key variables, AADT and segment length (Fig. 8). For consistency with the best statistical model (details can be found in Section 3.2.1 and Table 2), we used the natural log form of VMT per year (in millions), respectively, in all statistical and ML base-learners. The partial dependence plots reveal a non-linear association of average yearly VMT (millions) with average yearly crash frequency (Fig. 8). For instance, there is a sharp increase in crash frequency beyond a VMT of 0.67*10^6 (Fig. 8). With

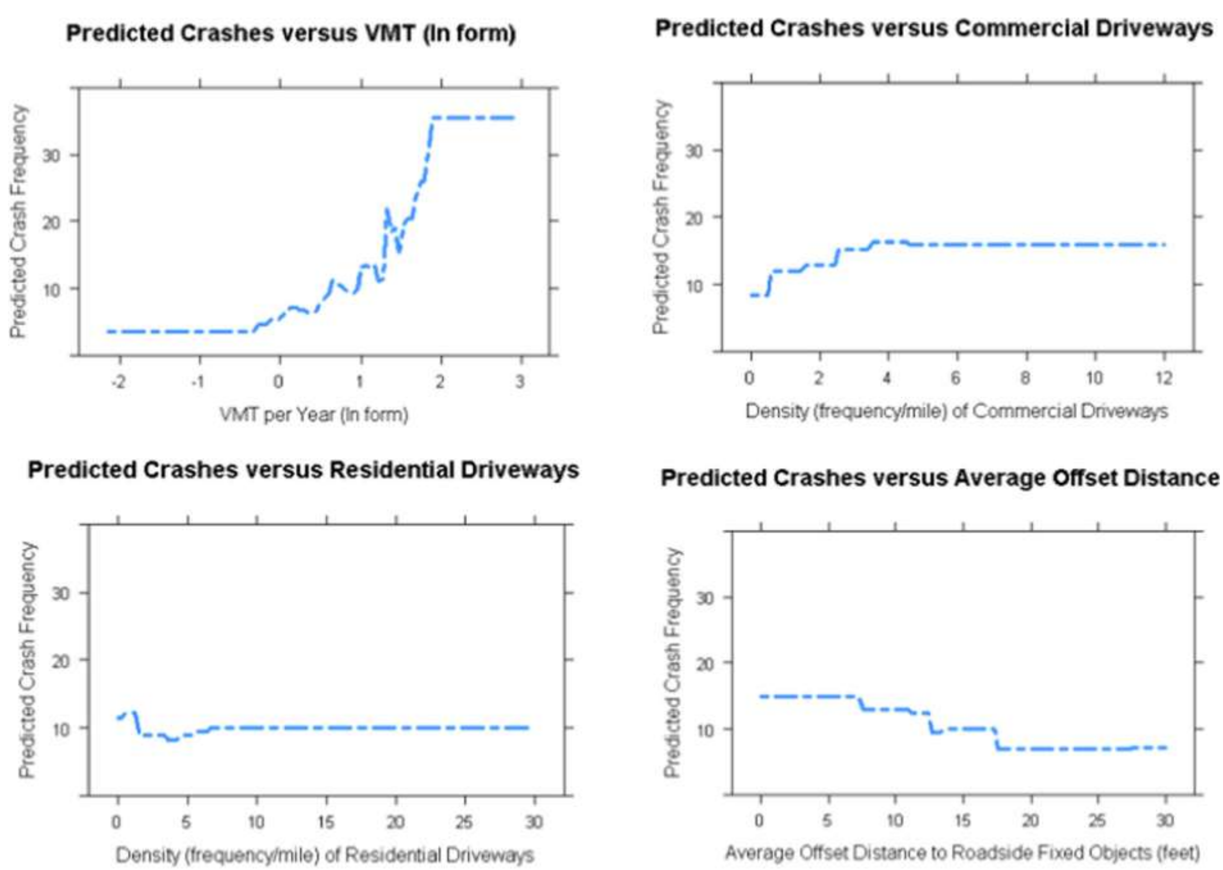


Fig. 8. Yearly Predicted Crashes by GBR (best performing base-learner) for VMT and Other Predictor Variables.

a higher average yearly VMT, the frequency of average crashes (including both injury and non-injury crashes) per year increases. While previous studies reveal that total crash frequency increases with VMT (Jovanis & Chang, 1986), the interesting aspect of the current study is that it captures non-linearities in such a relationship through ML methods. Referring to the partial dependence plots of VMT (millions), the values of −2.000 and 2.000 along the x-axis indicate a VMT of 0.1353 (=$e^{-2.000}$) and 7.3890 (=$e^{2.000}$) in millions respectively (Table 1). From the plots, predicted crashes per year (GBR) increase with an increase in yearly VMT between $0.67*10^6$ (=$e^{-0.4}$) and $6.69*10^6$ (=$e^{1.9}$). Interestingly, if yearly AADT decreases or increases beyond the values of $0.6703*10^6$ and $6.69*10^6$, respectively, the number of predicted crashes by optimal GBR base-learners remain constant (∼4 and ∼ 35 crashes per year, respectively) (Fig. 8). Similarly, the relationship of predicted crashes by GBR with change in density of commercial driveways, average offset distance to fixed objects, and density of residential driveways can be seen in Fig. 8.

3.2.3. Stacking

After the five individual count data and machine learning-based models are developed using training data (2013–2015), the performance is evaluated using a validation dataset (2016). In the next step, the stacked model is trained on the validation dataset (2016) for which observed crash frequency (2016) is used as a response variable. Eventually, the predictions obtained from the five base-learners applied to the validation dataset are used as inputs (predictors) to train the stacked model. Descriptive statistics of predicted and observed crashes for the validation dataset are shown in Table 4. Note that the mean number of crashes (2016) predicted by individual models such as count data models (Poisson and Negative Binomial model) and machine learning models such as TBR, RFR, and GBR (P_3, P_4, and P_5, respectively) are very similar to the mean number of observed crashes occurred during 2016 (Table 4). While using the validation dataset including five new predicted values (P_1, P_2, ..., P_5) and observed crashes, we train an RFR model as a meta-learner (stacked ensemble model) in second-stage regression. Several techniques ranging from simple linear regression to more robust ensemble methods like RFR and GBR can be used to train the stacked model.

The three ML methods (TBR, RFR, and GBR) were used as stacking meta-learners in the second-stage regression to predict crashes using the optimal combination of the base-learners. Our findings suggest that all of the three ML-based stacking meta-learners including RFR, GBR, and TBR significantly reduced the out-of-sample RMSE, and MAE compared to homogeneous ensembles (RFR and GBR) used as base-learners (Table 5). As mentioned, we used the three ML methods (TBR, RFR, and GBR) as meta-learners to predict crashes; however, we present and discuss the results of RFR as a meta-learner (stacked ensemble method) because it led to maximum improvement in out-of-sample prediction accuracy.

Similar to individual machine learning models (TBR, RFR, and GBR), grid search optimization and 10-fold cross-validation procedures were used to select optimal values for regularization parameters of the Stacked model (RFR meta-learner). The tuning parameters in the random forest model include the number of predictors considered at each split, the number of trees, and the maximum number of nodes in the random forest, which were found to be 2, 450, and 15, respectively (Fig. 9). To select the best combination of tuning parameters with the lowest RMSE, different values of the three tuning parameters were assigned, which include: 2 to 5 (with an increment of 1) for the number of predictors considered at each split, 1 to 31 (with an increment of 2) for the maximum number of nodes, and 250 to 1,000 (with an increment of 50) for the number of trees. Based on the values of the tuning parameters, 1,024 different combinations of the three tuning parameters were considered in the grid-search procedure. We illustrate RMSE for only the 20 best combinations of the tuning parameters that result in smaller RMSE compared to the remaining 1,004 combinations (Fig. 9). The optimal stacked RFR model was fitted using the values of 2, 15, and 450, which are found to be the optimal values of number predictors considered for splitting at each node, the maximum number of nodes, and the number of trees, respectively.

The relative importance plot of the predictors (obtained from the five base-learners) for meta-learner (stacked ensemble model) is shown in Fig. 10. The predicted crashes obtained from the individual RFR model (P_4) are found to be the most important predictor variable followed by those predicted via gradient boosting (P_5) (importance = 99.24 %), Poisson model (P_1), negative binomial model (P_2), and TBR model (P_1) (Fig. 10).

3.2.4. Comparing out-of-sample prediction performance

For evaluating the out-of-sample prediction performance of the stacked versus un-stacked models, crash data in 2017 are used – these data were neither used to train base-learners nor the meta-learner (stacked model). Before comparing the out-of-sample prediction performance of various stacking meta-learners, it should be noted that each of the three ML methods (TBR, RFR, and GBR) were used as stacking meta-learners. To select optimal values of regularization parameters for a particular ML meta-learner, we used the same procedure when the method was used as an ML base-learner. Using the optimal values of specific regularization parameters obtained through 10-fold cross-validation and extended grid-search, the three ML methods including TBR, RFR, and GBR were trained as stacking meta-learners. To compare the predictive performance of the five base-learners and meta-learners based on the new dataset, we computed out-of-sample RMSE and MAE (Table 5). Our findings indicate that GBR has the lowest out-of-sample RMSE and MAE among all base-learners (Table 5). Referring to the predictive performance of meta-learners, all of the three stacking meta-learners including RFR, GBR, and TBR further reduced out-of-sample RMSE and MAE compared to the best performing base-learner (GBR) (Table 5). To conclude, RFR, as stacking meta-learner, is found to have the lowest out-of-sample RMSE and out-of-sample MAE among all the base-learners and meta-learners and is selected as the best performing model for out-of-sample crash prediction (Table 5). Among all the meta-learners, the results

Table 4
Descriptive statistics of predicted and observed crashes (Validation Dataset: 2016).

Variables	Obs.	Mean	Std. Dev.	Min	Max
Total Crashes (2016)	304	11.010	14.651	0.000	90.000
Predicted Crashes via Poisson Model (P_1)	304	11.257	10.580	1.079	70.126
Predicted Crashes per Negative Binomial Model (P_2)	304	11.508	12.215	1.086	90.125
Predicted Crashes per Decision Tree Model (P_3)	304	11.416	11.919	3.127	56.000
Predicted Crashes per Random Forest Model (P_4)	304	11.244	9.465	1.242	45.631
Predicted Crashes per Gradient Boosting Model (P_5)	304	11.078	10.482	0.740	57.647

Note: Furthemore, obs., Std. Dev., Min, and Max refer to number of observations, standard deviation, minimum, and maximum, respectively.

Table 5
Comparison Prediction Performance (Out-of-Sample): RMSE and MAE.

Model	Type	RMSE	% Difference in RMSE compared to GBR Model	MAE	% Difference in MAE compared to GBR Model
TBR	Meta-learner	8.96	−9.40	5.88	−8.55
Gradient Boosting	Meta-learner	8.83	−10.72	5.84	−9.18
Random Forest	Meta-learner	8.34	−15.67	5.41	−15.86
Poisson	Base-learner	10.39	5.06	6.60	2.64
Negative Binomial	Base-learner	11.21	13.35	6.87	6.84
TBR	Base-learner	10.69	8.09	6.66	3.58
Random Forest	Base-learner	9.99	1.01	6.44	0.16
Gradient Boosting	Base-learner	9.89	Base	6.43	Base

Note: % Difference in RMSE compared to GBR Model $= \frac{(RMSE_{Model\ X} - RMSE_{GBR\ Model})}{RMSE_{GBR\ Model}} * 100\%$.
% Difference in MAE compared to GBR Model $= \frac{(MAE_{Model\ X} - MAE_{GBR\ Model})}{MAE_{GBR\ Model}} * 100\%$.
Note: Meta-learner refers to the model when it is applied in the second stage to use predictions from the optimal combinations of different base-learners. In stacking, a "meta-leaner" is also termed a "super-learner" (Van Der Laan et al., 2007). As defined in the texts earlier, RMSE, MAE, and GBR refer to root mean square error, mean absolute error, and gradient boosting tree regression, respectively.

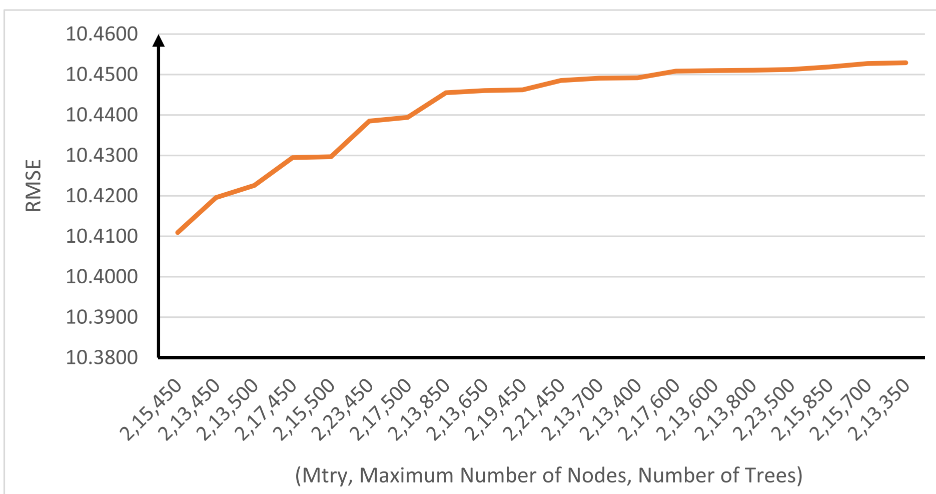


Fig. 9. Selecting Optimal Tuning Parameters for Stacked RFR Model (Second-Stage Regression). **Note:** Mtry and RMSE in the figure above refer to number of predictors used in the stacked RFR (random forsest regression) at each split and root mean square error, respectively.

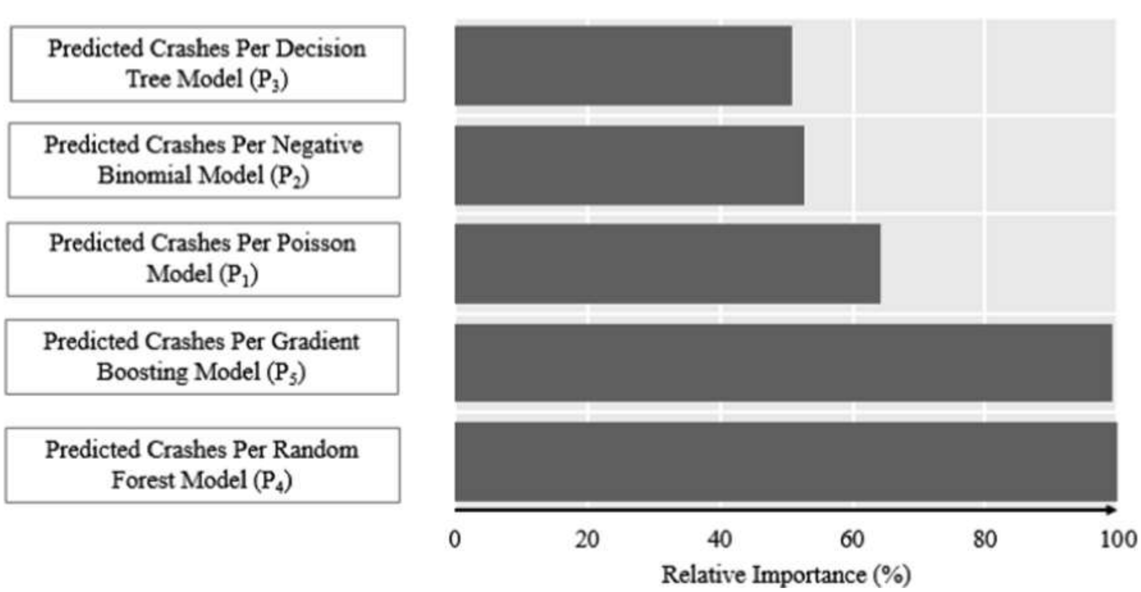


Fig. 10. Variables Relative Importance Plot: Optimal Stacked RFR Model (Meta-Learner).

Table 6
Summary of Absolute Prediction Errors (Out-of-Sample) for Base and Meta-learners.

Model	Type (used as)	N	Absolute (observed crashes - predicted crashes)			
			Mean	Std. Dev.	Minimum	Maximum
TBR	Meta-learner	304	5.88	6.78	0.17	59.00
Gradient Boosting	Meta-learner	304	5.84	6.63	0.02	59.95
Random Forest	Meta-learner	304	5.41	6.35	0.01	52.43
Poisson	Base-learner	304	6.60	8.03	0.00	77.25
Negative Binomial	Base-learner	304	6.87	8.87	0.04	77.58
TBR	Base-learner	304	6.66	8.37	0.12	81.12
Random Forest	Base-learner	304	6.44	7.66	0.01	73.31
Gradient Boosting	Base-learner	304	6.43	7.50	0.03	64.23

Note: Meta-learner refers to the model when it is applied in the second stage to use predictions from the optimal combinations of different base-learners. In stacking, a "meta-leaner" is also termed a "super-learner" (Van Der Laan et al., 2007). In Table 6, TBR, N, and Std. Dev. refer to tree-based regression, number of observations, and standard deviation, respectively.

of RFR as a stacking meta-learner (with the lowest out-of-sample RMSE and MAE among all meta-learners and base-learners) are only discussed in the paper.

To have a deeper understanding of the out-of-sample prediction errors, we also provide distributional statistics of out-of-sample absolute prediction errors (Table 6). RFR as a stacking meta-learner leads to the lowest out-of-sample absolute prediction error (Table 6). The standard deviations of absolute prediction errors for all three stacking meta-learners including RFR, GBR, and TBR are found to be smaller, indicating less spreading-out around the mean value of the error (Table 6).

To visualize and compare the out-of-sample prediction performance of individual models (base-learners) and the stacked ensemble technique (meta-learner), we present plots of predicted versus observed crashes based on the testing data (2017) (Fig. 11). It can be seen that the RFR ensemble model, when used as a meta-learner, shows the best fit, followed by GBR (meta-learner) and TBR (meta-learner) as shown in Fig. 11. Similar findings are obtained in other fields where prediction accuracy for the stacked ensemble model (used for classification) improved by 2 %–4% (Güneş et al., 2017). To conclude, we found that the application of the stacked ensemble technique can help in obtaining more accurate crash predictions in the future. Forecasting error is composed of the inversely related bias and variance errors of the underlying model parameters. The stacked ensemble technique can help achieve a desirable trade-off between bias and variance – pooling predictions from traditional (low bias, high variance) and machine learning (high bias, low variance) models, ultimately leading to more accurate forecasts. Note that none of the studies evaluated the applicability of more accurate, reliable, and intelligent heterogeneous ensemble procedures to determine the crash frequency. Similar rigorous stacked ensemble techniques can be used to predict crash frequency on other types of roadway using local data.

The plot of predicted versus actual crash frequency for the tree-based regression model seems unusual compared to the other five regression models. Note that tree-based regression models only assign a specific number of values to response outcomes for individual observations (roadway segments in this case) based on their attribute values and conditions (for details, refer to Fig. 4) assigned by the optimal tree-based regression model. In our case, the optimal tree-based regression model indicates that 1 of the 12 values (3.13, 6.57, 9.60, 4.83, 18.54, 13.58, 29.57, 56, 18.86, 40, 50, 53.1) of crashes may be assigned to any segment based on its attributes

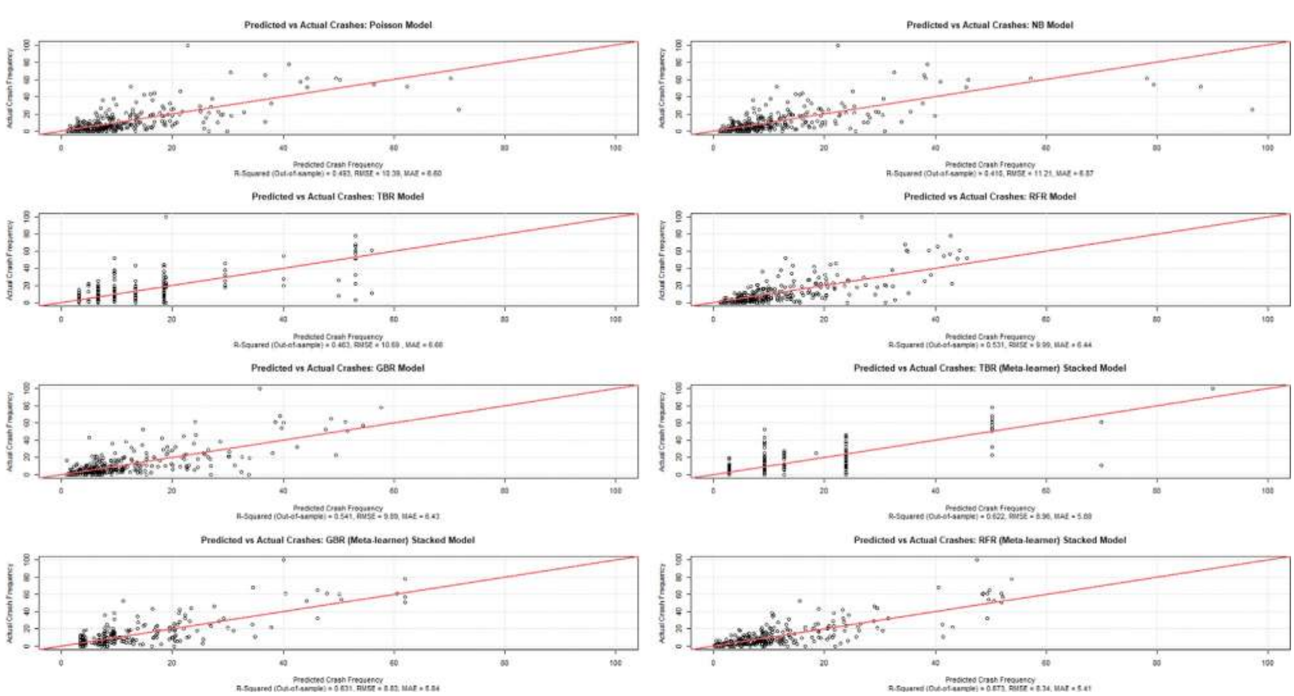


Fig. 11. Out-of-sample prediction: Observed versus Predicted Crashes.

(mean VMT, density of commercial driveways, and average offset distance to fixed objects) based on which the optimal TBR model was developed. Hence, it can be seen in Fig. 11 that crashes are spotted at a few specific points by TBR models (both as base-learner and meta-learner), which seems unusual compared to the remaining models.

4. Limitations and future directions

This study uses 5 T segments of urban and suburban arterials in Tennessee and may not be extended to other states due to variations in driving behavior, socio-demographic, and roadway conditions. Compared to individual machine learning techniques, we recommend using heterogeneous ensemble methods like stacking, which are more accurate, reliable, and intelligent techniques and can help in accessing crash forecasts in the future. While stacking may significantly improve the out-of-sample prediction accuracy, it does not provide the variable importance for the actual predictor variables (e.g., VMT, average offset distance to fixed objects). Note that in this study, stacking is applied to combine multiple predictions (as opposed to combining distributions of coefficients, such as in the Bayesian setup). Thus, the inference is not relevant in Stage 2. However, the inferences are provided in Stage-1 by individual base-learners that include statistical models (like Poisson and negative binomial) and ML methods including TBR, RFR, and GBR. Note that based on our study objectives, we split five years of crash data into training (2013–2015), validation (2016), and testing (2017) datasets - indicting that only crash frequency and VMT (due to AADT) may vary across the datasets while roadway geometry remains similar. In the future, data splitting can also be done using standard splitting procedures rather than the year-wise split, depending on the study design and objectives. The application presented herein is based on year-wise splits, assuming temporal transferability of the models over years. As part of future work, variants of the methods presented herein that relax this assumption can be examined.

5. Conclusions

Safety performance functions are core tools necessary for the accurate prediction of crashes and subsequent development of place-based countermeasures. Traditional count data models and machine learning methods have been extensively used in the safety literature for the development of statistical relationships between crash frequency and associated factors. This study contributes by presenting a rigorous and novel heterogeneous ensemble methods (HEM) scheme to "stack" predictions from competing frequentist and ML models – eventually leading to a more accurate prediction of crashes. By using a more accurate and reliable intelligent pattern recognition scheme, the "Stacking" methodology harnesses the inferential framework provided by traditional count data models and the predictive power offered by ML methods. The objectives are achieved using five-year crash, traffic, and roadway geometric data for urban and suburban arterials extracted from the Enhanced Tennessee Roadway Information Management System (ETRIMS). To the best of the authors' knowledge, no study to date has applied heterogeneous ensemble methods to pool multiple predictions from frequentist and ML methods.

The results suggest the significant potential of "Stacking" in providing more accurate predictions by heterogeneously assembling crash forecasts from individual statistical (Poisson and negative binomial) and machine-learning-based base-learners (tree-based regression, random forests, and gradient boosting regression). Using out-of-sample prediction performance, the gradient boosting model led to the lowest RMSE and MAE values among all the individual base-learners. While individual ML-based base-learners can provide greater predictive accuracy, there is no escaping the relationship between bias and variance underpinning most machine learning models. In other words, using a single supervised or unsupervised ML method could lead to relatively less accurate predictions due to the compromised bias or variance. By superimposing the best machine-learning-based meta-learner on predictions obtained from the five statistical and ML-based base-learners, the RMSE and MAE values of crash forecasts were further reduced by 15.67 % and 15.86 %, respectively, compared to the prediction accuracy of the best-fit gradient boosting based individual base-learner. From an inferential standpoint, the individual base-learners offer insights into the links between crash frequency and associated factors. Count data models show that besides exposure variables (VMT), higher accessibility in commercial areas correlates with higher crash frequency. Contrarily, a larger offset distance to a fixed object correlates with lower crash frequency. In terms of variable importance, the three ML-based base-learners rank VMT, density of commercial driveways, and average offset distance to fixed objects as the three top predictors of crash frequency.

The results of this study have important implications. By using heterogeneous ensemble methods such as Stacking, even more, accurate crash forecasts can be obtained compared to those obtained from individual frequentist or ML methods. With more accurate crash forecasts, roadway segments can be better prioritized in terms of the need for place-based safety countermeasures. From a practical standpoint, the straightforward heterogeneous ensemble method technique can be easily automated for more accurate crash prediction. From a research perspective, the methodology can be expanded by other researchers to include an even broader set of ML methods or consider more rigorous simulation-assisted statistical methods accounting for methodological issues like observed and unobserved heterogeneity.

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.

Acknowledgments

This study was funded by the Tennessee Department of Transportation (Grant No. RES2020-04) and the U.S. DOT through the Collaborative Sciences Center for Road Safety; a consortium led by The University of North Carolina at Chapel Hill in partnership with The University of Tennessee. The views in this paper are solely those of the authors who are responsible for the content of this publication. Following the double-blind review policy, we will duly acknowledge the efforts of the individuals who have helped with some parts of the data collection used in this paper and who have done proofreading of the article once the final decision is made by the journal.

References

AASHTO, **2010**. Highway Safety Manual. Washington, DC.
Abdel-Aty, M. A., & Radwan, A. E. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention, 32*(5), 633–642.
Akaike, H., Petrov, B. N., & Csaki, F. (1973). *Second international symposium on information theory*. Budapest: Akadémiai Kiadó.
Albuquerque, F. D., & Awadalla, D. M. (2019). Roadside design assessment in an urban, low-density environment in the gulf cooperation council region. *Traffic Injury Prevention, 20*(4), 436–441.
Albuquerque, F. D., & Awadalla, D. M. (2020). Roadside fixed-object collisions, barrier performance, and fatal injuries in single-vehicle, run-off-road crashes. *Safety, 6*(2), 27.

Anastasopoulos, P. C., & Mannering, F. L. (2009). A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis & Prevention, 41*(1), 153–159.
Bhatt, S., Cameron, E., Flaxman, S. R., Weiss, D. J., Smith, D. L., & Gething, P. W. (2017). Improved prediction accuracy for disease risk mapping using gaussian process stacked generalization. *Journal of the Royal Society Interface, 14*(134), 20170520.
Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*(3), 345–370.
Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140.
Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC Press.
Bugusa, Y., & Patil, S. (2019). An improved accident crash risk prediction model based on driving outcomes using ensemble of prediction algorithms. *International Journal of Scientific & Technology Research, 8*(11), 603–611.
Caliendo, C., Guida, M., & Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis & Prevention, 39*(4), 657–670.
Chali, Y., Hasan, S. A., & Mojahid, M. (2014). Complex question answering: Homogeneous or heterogeneous, which ensemble is better? In *Proceedings of the International Conference on Applications of Natural Language to Data Bases/ Information Systems* (pp. 160–163).
Chung, Y.-S. (2013). Factor complexity of crash occurrence: An empirical demonstration using boosted regression trees. *Accident Analysis & Prevention, 61*, 107–118.
Das, A., Abdel-Aty, M., & Pande, A. (2009). Using conditional inference forests to identify the factors affecting crash severity on arterial corridors. *Journal of Safety Research, 40*(4), 317–327.
De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology, 88* (1), 243–251.
Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning, 40*(2), 139–157.
Dixon, K., Avelar, R., Brown, L., Mecham, M., & Van Schalkwyk, I. (2012). Quantifying safety performance of driveways on state highways. *Oregon Department of Transportation. Research Section*.
Elish, M. O., Helmy, T., & Hussain, M. I. (2013). Empirical study of homogeneous and heterogeneous ensemble models for software development effort estimation. *Mathematical Problems in Engineering, 2013*.
Elith, J., Leathwick, J., 2017. *Boosted regression trees for ecological modeling*. R Documentation. available online: https://cran.r-project.org/web/packages/dismo/vignettes/brt.pdf.
Farid, A., Abdel-Aty, M., & Lee, J. (2019). Comparative analysis of multiple techniques for developing and transferring safety performance functions. *Accident Analysis & Prevention, 122*, 85–98.
Fernández-Alemán, J. L., Carrillo-De-Gea, J. M., Hosni, M., Idri, A., & García-Mateos, G. (2019). Homogeneous and heterogeneous ensemble classification methods in diabetes disease: A review. In *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3956–3959).
Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.
Ghandour, A. J., Hammoud, H., & Al-Hajj, S. (2020). Analyzing factors associated with fatal road crashes: A machine learning approach. *International Journal of Environmental Research and Public Health, 17*(11), 4111.
Güneş, F., Wolfinger, R., & Tan, P.-Y. (2017). Stacked ensemble models for improved prediction accuracy. In *Proceedings of the Proc. Static Anal. Symp.* (pp. 1–19).
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data Mining, inference and Prediction*. Springer Science & Business Media.
Hauer, E., Council, F. M., & Mohammedshah, Y. (2004). Safety models for urban four-lane undivided road segments. *Transportation Research Record, 1897*(1), 96–105.
Heung, B., Bulmer, C. E., & Schmidt, M. G. (2014). Predictive soil parent material mapping at a regional-scale: A Random Forest Approach. *Geoderma, 214*, 141–154.
Iranitalab, A., & Khattak, A. (2017). Comparison of four statistical and machine learning methods for crash severity prediction. *Accident Analysis & Prevention, 108*, 27–36.
Jovanis, P. P., & Chang, H. L. (1986). Modeling the relationship of accidents to miles traveled. *Transportation Research Record, 1068*, 42–51.
Karlaftis, M., & Golias, I. (2002). Effects of road geometry and surface on speed and safety. *Accident Analysis & Prevention, 34*(34), 357–365.
Khattak, A., Ahmad, N., Mohammadnazar, A., Mahdinia, I., Wali, B., & Arvin, R., 2020. *Highway Safety Manual safety performance functions & roadway calibration factors: Roadway segments phase 2, part*.
Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News, 2*(3), 18–22.
Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. In *Proceedings of the Advances in Neural Information Processing Systems* (pp. 431–439).
Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications, 42*(10), 4621–4631.
Mclachlan, G. J., Do, K.-A., & Ambroise, C. (2005). *Analyzing microarray gene expression data*. John Wiley & Sons.
Mohammadnazar, A., Mahdinia, I., Ahmad, N., Khattak, A., & Liu, J. (2021). Understanding how relationships between crash frequency and correlates very for multilane rural highways: Estimating geographically and temporally weighted regression models. *Accident Analysis and Prevention*. https://doi.org/10.1016/j.aap.2021.106146.
Osareh, A., & Shadgar, B. (2013). An efficient ensemble learning method for gene microarray classification. *BioMed Research International, 2013*.
Pan, G., Fu, L., & Thakali, L. (2017). Development of a global road safety performance function using deep neural networks. *International Journal of Transportation Science and Technology, 6*(3), 159–173.
Petrakova, A., Affenzeller, M., & Merkurjeva, G. (2015). Heterogeneous versus homogeneous machine learning ensembles. *Information Technology and Management Science, 18*(1), 135–140.
Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering, 122*(2), 105–113.
Rooney, N., Patterson, D. W., Anand, S. S., & Tsymbal, A. (2004). Random subspacing for regression ensembles. *Proceedings of the FLAIRS Conference*.
Sabzevari, M., Martínez-Muñoz, G., & Suárez, A., 2018. *Pooling homogeneous ensembles to build heterogeneous ones*. arXiv preprint arXiv:1802.07877.
Saeed, T. U., Hall, T., Baroud, H., & Volovski, M. J. (2019). Analyzing road crash frequencies with uncorrelated and correlated random-parameters count models: An empirical assessment of multilane highways. *Analytic Methods in Accident Research, 23*, 100101.
Saha, D., Alluri, P., & Gan, A. (2015). Prioritizing highway safety manual's crash prediction variables using boosted regression trees. *Accident Analysis & Prevention, 79*, 133–144.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464.
Shankar, V., Mannering, F., & Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis & Prevention, 27*(3), 371–389.
Sigletos, G., Paliouras, G., Spyropoulos, C. D., & Hatzopoulos, M. (2005). Combining information extraction systems using voting and stacked generalization. *Journal of Machine Learning Research, 6*(Nov), 1751–1782.
Srinivasan, R., & Carter, D., 2011. Development of safety performance functions for North Carolina. North Carolina Deptartment of Transportation. Research and Analysis Group.
Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics, 9*(1), 307.
Tang, J., Liang, J., Han, C., Li, Z., & Huang, H. (2019). Crash injury severity analysis using a two-layer stacking framework. *Accident Analysis & Prevention, 122*, 226–238.
Tewari, S., & Dwivedi, U. (2020). A comparative study of heterogeneous ensemble methods for the identification of geological lithofacies. *Journal of Petroleum Exploration and Production Technology*, 1–20.
Thakali, L., Fu, L., & Chen, T. (2016). Model-based versus data-driven approach for road safety analysis: Do more data help? *Transportation Research Record, 2601* (1), 33–41.
Thapa, R., Gupta, S., Gupta, A., Reddy, D., & Kaur, H. (2018). Use of geospatial technology for delineating groundwater potential zones with an emphasis on water-table analysis in Dwarka River Basin, Birbhum. *India. Hydrogeology Journal, 26*(3), 899–922.
Torgo, L.F.R.A., 1999. *Inductive learning of tree-based regression models*.
Van Der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology, 6*(1).
Wali, B., Ahmad, N., Khattak, A. J., & Nazar, A. M. (2020). Developing safety performance functions for rural multilane highways in Tennessee: Accounting for unobserved heterogeneity. *Presentated in 99th Annual Meeting of the Transportation Research Board (TRB), Washington DC*.
Wali, B., Khattak, A. J., Waters, J., Chimba, D., & Li, X. (2018). Development of safety performance functions: Incorporating unobserved heterogeneity and functional form analysis. *Transportation Research Record, 2672*(30), 9–20.
Wang, K., Simandl, J. K., Porter, M. D., Graettinger, A. J., & Smith, R. K. (2016). How the choice of safety performance function affects the identification of important crash prediction variables. *Accident Analysis & Prevention, 88*, 1–8.
Washington, S., Karlaftis, M. G., & Mannering, F. I. (2010). *Statistical and econometric methods for transportation data analysis*. New York: Chapman & Hall/CRC.
Wolf, K. L. (2006). Urban trees and traffic safety: Considering the US roadside policy and crash data. *Arboriculture and Urban Forestry., 32*(4), 170–179.
Wolpert, D. H. (1992). Stacked generalization. *Neural Networks, 5*(2), 241–259.
Yu, R., & Abdel-Aty, M. (2014). Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science, 63*, 50–56.
Zahid, M., Chen, Y., Khan, S., Jamal, A., Ijaz, M., & Ahmed, T. (2020). Predicting risky and aggressive driving behavior among taxi drivers: Do spatio-temporal attributes matter? *International Journal of Environmental Research and Public Health, 17*(11), 3937.
Zhang, M., Xu, Q., Daeyaert, F., Lewi, P., & Massart, D. (2005). Application of boosting to classification problems in chemometrics. *Analytica Chimica Acta, 544*(1–2), 167–176.
Zhong, C., Sisiopiku, V., Ksaibati, K., & Zhong, T. (2011). Crash prediction on rural roads. *Proceedings of the 3rd International Conference on Road Safety and Simulation, Indianapolis, Indiana*.
Zong, F., Chen, X., Tang, J., Yu, P., & Wu, T. (2019). Analyzing traffic crash severity with combination of information entropy and bayesian network. *IEEE Access, 7*, 63288–63302.

Numan Ahmad is a PhD student in Department of Civil and Environmental Engineering at University of Tennessee, Knoxville. He is also pursuing his M.S. in Statistics at the same university. He holds a master's degree in Transportation Engineering and a bachelor's degree in Civil Engineering from National University of

Sciences and Technology Pakistan and University of Engineering and Technology Peshawar, Pakistan respectively. He has authored/co-authored 10 scientific articles in peer-reviewed journals and 10 conference papers presented at the Transportation Research Board Annual Meetings at Washington D.C. His main research interests relate to transport safety, alternative fuel vehicles, automated vehicles, and safety of vulnerable road users.

Behram Wali received his Ph.D. degree in Transportation Engineering and M.S. in Statistics from University of Tennessee, Knoxville in 2018. He also holds M.S. degree in Transportation Engineering form the National University of Sciences and Technology, Pakistan in 2015 where he was awarded "President of Pakistan Gold Medal Award for Outstanding Graduate Student". Dr. Wali currently works as Lead Research Scientist at Urban Design 4 Health. He has worked as Postdoctoral Scholar at the MIT Senseable City Lab. Dr. Wali has authored/co-authored more than 35 scientific articles in peer-reviewed journals. His research interests include ITS, safety, CAVs and transport & health.

Asad J. Khattak is a Beaman Distinguished Professor and Coordinator of the Transportation Group in the Department of Civil and Environmental Engineering at University of Tennessee, Knoxville. He is also Associate Director, Collaborative Sciences Center for Road Safety, a National University Transportation Center consortium led by University of North Carolina at Chapel Hill. His main expertise and interests are related to ITS, traffic safety, and sustainable transportation. He has authored/co-authored more than 200 articles and technical reports. Professor Khattak is the Editor-in-Chief and Associate Editor of the Journal of Intelligent Transportation Systems and International Journal of Sustainable Transportation Systems respectively.

Journal of Safety Research 84 (2023) 24–32

Contents lists available at ScienceDirect

Journal of Safety Research

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

How do different micro-mobility vehicles affect longitudinal control? Results from a field experiment

Marco Dozza *, Tianyou Li, Lucas Billstein, Christoffer Svernlöv, Alexander Rasch

The Department of Mechanics and Maritime Sciences at Chalmers University of Technology, Sweden

ARTICLE INFO

Article history:
Received 4 February 2022
Received in revised form 1 July 2022
Accepted 17 October 2022
Available online 28 October 2022

Keywords:
E-scooters
Segways
Cycling safety
Bicycle dynamics
Active safety

ABSTRACT

Introduction: While micromobility vehicles offer new transport opportunities and may decrease fuel emissions, the extent to which these benefits outweigh the safety costs is still uncertain. For instance, e-scooterists have been reported to experience a tenfold crash risk compared to ordinary cyclists. Today, we still do not know whether the real safety problem is the vehicle, the human, or the infrastructure. In other words, the new vehicles may not necessarily be unsafe; the behavior of their riders, in combination with an infrastructure that was not designed to accommodate micromobility, may be the real issue. *Method:* In this paper, we compared e-scooters and Segways with bicycles in field trials to determine whether these new vehicles create different constraints for longitudinal control (e.g., in braking avoidance maneuvers). *Results:* The results show that acceleration and deceleration performance changes across vehicles; specifically, e-scooters and Segways that we tested cannot brake as efficiently as bicycles. Further, bicycles are experienced as more stable, maneuverable, and safe than Segways and e-scooters. We also derived kinematic models for acceleration and braking that can be used to predict rider trajectories in active safety systems. *Practical Applications:* The results from this study suggest that, while new micromobility solutions may not be intrinsically unsafe, they may require some behavior and/or infrastructure adaptations to improve their safety. We also discuss how policy making, safety system design, and traffic education may use our results to support the safe integration of micromobility into the transport system.

1. Introduction

New micromobility vehicles (SAE Committee, 2018), compact and electrically powered, are on the rise worldwide (6t-bureau de recherche, 2019; Chang et al., 2019; Fitt & Curl, 2020; O'Hern & Estgfaeller, 2020; Portland Bureau of Transportation, 2018). A few years ago, e-bicycles (i.e., assisted cycles, pedelecs [SAE Committee, 2018]) were a new transport phenomenon that created some concerns in the safety research community (Huertas-Leyva et al., 2018; MacArthur et al., 2014; Schleinitz et al., 2017; Twisk et al., 2021). Today, e-bicycles are conventional, while new micromobility (e-)vehicles with different geometries, number of wheels, and number of tracks present new challenges for the transport system (Abduljabbar et al., 2021; O'Hern & Estgfaeller, 2020). While monowheels, e-skates, and Segways are not very popular yet and, maybe, they will never be, e-scooters are; they outnumber e-bicycles in many urban centers. It is hard not to see a trend toward electrical vehicles, and it is not a given that e-scooters are the peak of this transformation. In any case, micromobility is here to stay (Gössling, 2020), and it may indeed solve some congestion and pollution issues (6t-bureau de recherche, 2019; Abduljabbar et al., 2021; Portland Bureau of Transportation, 2018; Sharkey et al., 2020). Unfortunately, the safety toll that new micromobility vehicles—and e-scooters specifically—take may be hard to mitigate (Santacreu et al., 2020).

Several studies have shown that riding e-scooters is unsafe: the crash risk is 10 times higher than riding a bicycle (Fearnley et al., 2020). E-scooters also cause major injuries (Badeau et al., 2019; Bekhit et al., 2020; Ishmael et al., 2020; Namiri et al., 2020) that are different from the ones experienced by (e-)cyclists (Beck et al., 2020; Cicchino et al., 2021; B. Trivedi et al., 2019; T. K. Trivedi et al., 2019; Wüster et al., 2020). Several factors may contribute to explain these injury variations, including the different demographics and attitudes in wearing helmets between the cyclist and e-scooterist population. However, some differences (e.g., the higher prevalence of lower extremity injuries for e-scooterists compared to cyclists; Cicchino et al., 2021) suggest that the vehicle geometry and control also play a role.

* Corresponding author.
E-mail address: marco.dozza@chalmers.se (M. Dozza).
URL: https://www.chalmers.se/en/staff/Pages/marco-dozza.aspx (M. Dozza).

https://doi.org/10.1016/j.jsr.2022.10.005
0022-4375/© 2022 The Author(s). Published by the National Safety Council and Elsevier Ltd.

Today, we know very little about the causes of e-scooter crashes. The vehicles often take the blame, although road-user behavior and infrastructure may play more important roles in crash causation. Most of the research on e-scooter safety makes use of data collected after the crash has happened, either by the police, hospitals, or insurance companies (Stigson et al., 2020). While these data describe the consequence of a crash, they do not show what happened just before the crash; in other words, they may not show what caused the crash. Data collected in the field, on the other hand, either naturalistically (Dozza & Werneke, 2014) or in controlled experiments (Kovácsová et al., 2016), may complement the crash data collected a posteriori and help us understand why micromobility crashes occur—and how to avoid them (Dozza et al., 2022).

The same data may help educate micromobility riders; after all, 33 % of the injuries happen during the very first trip of novice riders (Austin Public Health, 2019), suggesting that the crash was caused by the riders' inexperience. In addition, field data may contribute to the development of active safety systems, such as emergency braking, which need to predict the rider behavior in order to provide timely and acceptable interventions (Boda et al., 2018). Finally, today's policy making, in the form of bans or geo-fencing, responds to general and static requirements, rather than dynamically changing in time and space according to the actual crash risk at a given moment. However, geo-fencing may have this ability if a sufficiently large amount of field data are available.

In this paper, we follow the procedure proposed by Dozza et al. (2022) for field data collection and analysis, and compare longitudinal control (i.e., acceleration and braking) among e-scooters, Segways, and bicycles (with and without assisted pedaling). Our main hypotheses were that: (1) as urgency increases, riders may be able to achieve larger acceleration and decelerations with all vehicles; (2) not all vehicles may exhibit the same acceleration and braking performance; and (3) braking and acceleration trajectories may be accurately predicted with simple linear models for all micromobility vehicles. By modeling micromobility kinematics, we can improve the threat assessment of active safety systems and promote a better understanding of how new micromobility vehicles differ from bicycles from a safety point of view.

2. Methods

The data collection and analyses in this study adapted the procedure from Dozza et al. (2022) by comparing acceleration (in addition to braking) and including a Segway (in addition to bicycles and e-scooters).

2.1. Participants

Nine female and 25 male subjects participated to this experiment by maneuvering an e-scooter, a Segway, and a bicycle in field trials. The participants' mean age (±standard deviation) was 23.5 y ± 4.2, mean height (±standard deviation) 1.75 m ± 0.08, and mean weight (±standard deviation) 71.5 kg ± 9.5. Participants shorter than 160 cm or heavier than 85 kg were excluded from the study to comply with the suggested heights and weights from the vehicle manufacturers. The inclusion criteria made sure that participants could ride a bicycle, were between 18 and 50 years old, had no disabilities, and had never been in a severe road crash. These criteria were set to control for possible biases in the results as indicated in (Dozza et al., 2022). Participants who had any symptoms of COVID-19 in the two weeks prior to the experiment were not allowed to take part. The maneuvers required the participant to longitudinally control (e.g., accelerate and brake) the vehicles in different conditions. Each subject signed a consent form before the experiment. The study was approved by the Swedish Ethical Review Authority (Etikprövningsmyndigheten; Ref. 2019–04547). An ad-hoc health insurance covered the participants during the experiment.

2.2. Equipment

The e-scooter (Ninebot ES2), Segway (Ninebot S), and bicycle (Monark Karin 3-VXL) were equipped with a logger and sensors for the collection of vehicle kinematics (Fig. 1). Specifically, the logger was based on a Raspberry Pi 3 model B, and kinematics were collected with an inertial measurement unit (IMU: PhidgetSpatial 3/3/3 1044_B). In addition, a light detection and ranging sensor (LiDAR: HOKUYO UXM 30LAH EWA), installed on the proving ground, was used to track the vehicles during the experiment. The data from the IMU and the LiDAR were combined to achieve a more accurate estimation of the vehicle kinematics than either sensor alone could provide. In particular, the longitudinal acceleration from the IMU and the trace of the centroid of the vehicles from the LiDAR were combined to estimate the position and speed of the rider during each maneuver. A Rauch-Tung-Striebel smoother made this combination possible (Rauch et al., 1965). More details about the processing are presented in the work by Billstein and Svernlöv (2021).

2.3. Protocol

After a period of training so the participants could get acquainted with the vehicles' operation, all participants were asked to accelerate and brake the three vehicles in five different tasks.[1] Two acceleration tasks required the participants to bring the vehicle to a constant speed of 17–20 km/h from a standstill either comfortably (comfort task) or harshly (harsh task). There were three braking tasks that all required braking from a constant 17–20 km/h speed. In the comfort braking task, they were asked to brake comfortably. In the harsh planned task, the participant was supposed to brake as late and hard as possible, stopping just before a line on the ground. In the unexpected task, the experimenter gave a command to stop at a random time and the participant was asked to respond by braking as hard as possible. These different braking conditions were chosen to simulate planned and unplanned braking maneuvers (Huertas-Leyva et al., 2018, 2019); the difference between them would help identify the role of expectation on response time (Dozza et al., 2022). The order of the vehicles and tasks was randomized for each subject, but all trials were completed for each of the vehicles before a new vehicle was ridden. The experimental conditions are shown on Fig. 2. The bicycle was used both as an e-bicycle and a conventional bicycle; in other words, each participant performed the experiment on the bicycle twice, with and without electrical assistance. Therefore, although only three vehicles were tested in this study, we present results for four different riding conditions: the e-scooter, the Segway, and the two bicycle configurations (assisted and unassisted).

2.4. Subjective data

After completing the tasks, the participants were asked to fill in a questionnaire that assessed: (1) how much previous experience they had with the different vehicles in the experiment and (2) their opinions of the performance of the vehicles during the experiment. For this second part, taken from works by (Dozza et al., 2022; Rasch et al., 2016), the participants ranked the four vehicles on a 7-level Likert scale (from 1 = Very poor, to 7 = Exceptional). The following riding six categories were ranked: mounting and dismount-

[1] https://www.youtube.com/watch?v=FWWfsQrtDQY.

Fig. 1. Instrumented vehicles with data loggers and inertial measurement units (IMUs).

ing, maintaining balance at low speed, maintaining balance at high speed, braking at low speed, braking at high speed, and accelerating from a standstill. The following four categories were ranked: stability, maneuverability, comfort, and safety.

2.5. Analyses

The accelerations and decelerations in the five tasks were modeled with linear regressions similar to those in previous studies (Kovácsová et al., 2016; Lee et al., 2020). We also computed the coefficient R^2 to verify the goodness of fit of the linear models. For the braking maneuvers, the distance covered to achieve a full stop was also computed. In addition, we compared the difference between the marked line and the actual position where the participants stopped, to determine how accurately they could estimate their braking distance. Finally, we computed the response time (i.e., the time passed between when the experimenter issued the stop command and when the vehicle started decelerating) for the unexpected braking task, to establish whether the vehicle type affected braking response time (Huertas-Leyva et al., 2018, 2019). The braking maneuver was defined as beginning when the vehicle speed dropped below 16 km/h (12 km/h for the Segway) and ending when it dropped below 2.5 km/h. The acceleration maneuver was defined to begin when the vehicle speed exceeded 2.5 km/h and end when it exceeded 16 km/h (12 km/h for the Segway). The reaction time in the unexpected-braking maneuver was defined to begin when the experimenter gave the stop command and end when the speed had dropped by 1 km/h.

Several generalized linear mixed-effect models (including the participant ID as a random effect and gender, vehicle, and maneuver type as fixed factors) were created to verify the significance of the results. Post-hoc tests were run on the results of the model whenever a factor with more than two categories was significant. The threshold for statistical significance was set to $\alpha = 0.05$ and adjusted with the Bonferroni correction to control for multiple tests across different analyses with uncorrelated measures. (All statistical analyses used the Statistics and Machine Learning Toolbox in Matlab and specifically the functions *fitglme* and *coefTest*.).

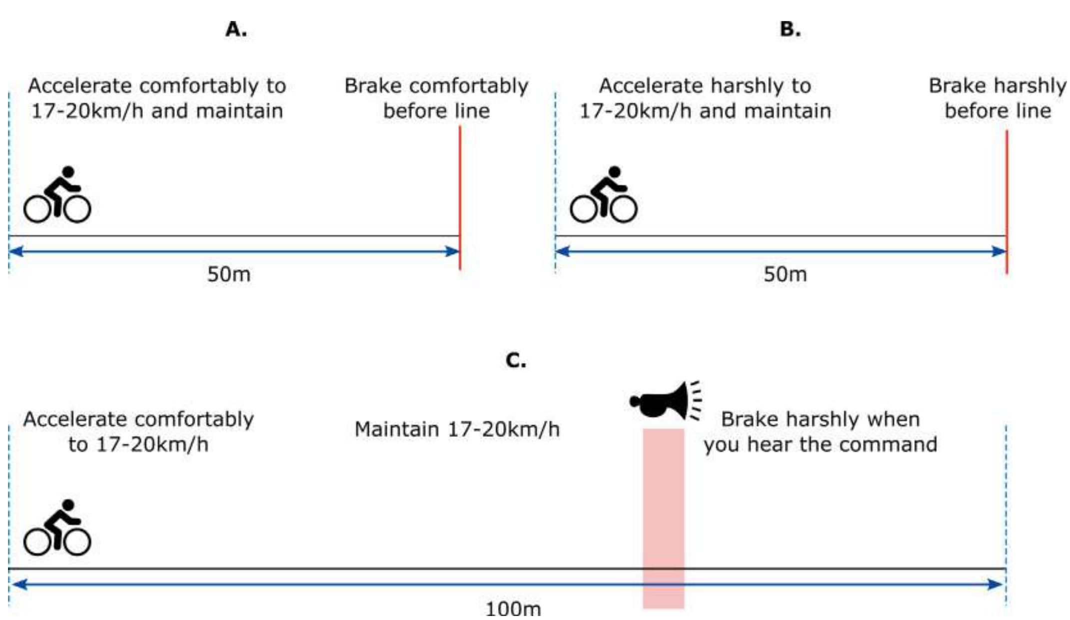


Fig. 2. Experimental protocol. Panel A: accelerating and braking comfortably. Panel B: accelerating and braking harshly. Panel C: braking harshly in response to a command from the experimenter; in this condition the ridden distance was larger than in the other conditions (100 m vs 50 m) to increase the variability of the braking command time.

3. Results

3.1. Dataset

Although we recruited 34 participants, only 25 of them felt comfortable riding the Segway and only 14 out of these 25 provided reliable sensor data for modeling acceleration and braking. Therefore, while the comparisons across the bicycle, e-bicycle, and e-scooter use the same population, the data for the Segway only include a subset of the population (significantly smaller for the kinematics analysis and slightly smaller for the questionnaire analysis). We also experienced other minor data losses. For instance, one of the participants crashed during the experiment; it was then stopped and none of the data were used for the analysis. Data were also excluded from the analysis when the participant did not reach the desired speed before starting braking. It is worth noting that we experienced a significant data loss for technical issues only on the Segway. This was mainly the consequence of a malfunctioning USB drive in the Segway installation.

All participants were very used to riding a conventional bicycle and much less experienced riding with the other vehicles (Table 1). An issue with the Segway was winning the fear of falling when stepping on the vehicle, which is necessary to start riding. In fact, to ride the Segway, the participants had to step with both feet on the vehicle within a short time and balance longitudinally. This action may be uncomfortable (even for experienced riders) because it creates some forward and backward sway that may feel like losing equilibrium. Nine participants did not win, or did not want to try to win, this fear of falling and just refused to ride the Segway; however, most of the data loss was the consequence of technical issues. In the training phase, participants could practice with any vehicle as long as they wanted and, on average, this training phase took 15 min per participant.

3.2. Acceleration maneuvers

Fig. 3 shows the average acceleration across all subjects for comfort and harsh acceleration maneuvers. It can be observed that the assisted bicycle enabled greater accelerations (up to 20 km/h) than the other vehicles; further, the Segway stopped accelerating as it approached 15 km/h, possibly because its design felt unstable at higher speed, so people backed off. Table 1 reports the angular coefficients from the regression models, representing the average acceleration during the trials. Harsh maneuvers resulted in statistically significantly larger accelerations than comfort maneuvers ($t = 7.5$; $p \ll 0.01$; Appendix Table A), suggesting that the participants understood the instructions and could control the vehicles accordingly. While accelerations were not statistically significantly different across gender or age, they were different across vehicles ($F = 16.4$; $p \ll 0.001$; Appendix Table A). Specifically, the assisted bicycle accelerated significantly faster than the e-scooter or the conventional bicycle. The acceleration of the Segway was also very high in the beginning of the maneuver. (Table 2 shows the average acceleration of the Segway until it reached 12 km/h, which may not be directly comparable with that of the other vehicles, which were able to reach 17–20 km/h as instructed.).

3.3. Braking maneuvers

The average speeds over time for each of the three braking maneuvers are presented in Fig. 4. Table 3 complements Fig. 4 by presenting the linear coefficients from the regression models for all vehicles and braking maneuvers. In all maneuvers, the Segway achieved a lower deceleration compared to the other vehicles and the deceleration started at a lower speed. It is very important to keep these differences in mind, especially when comparing the Segway's braking distances with those of the other vehicles.

When riding the bicycle (in both assisted and unassisted modes), the participants were able to brake with larger decelerations than when riding the other vehicles, and this result was statistically significant ($F = 39$; $p \ll 0.001$; Appendix Table B1). The participants' braking performance when riding the Segway was poorer (i.e., deceleration was lower) than for the other vehicles. As expected, the two harsh braking maneuvers resulted in statistically significantly larger braking decelerations for all vehicles ($F = 8.87$; $p \ll 0.001$). What was somewhat surprising is that the unexpected harsh braking task resulted in slightly greater decelerations than the planned harsh braking. No statistically significant difference in braking deceleration was found across ages or genders (Appendix Table B1).

While the braking distances were similar for the assisted and unassisted bicycle modes, the braking distance was statistically significantly longer for the e-scooter than for the bicycle in the harsh braking conditions (Fig. 5; Appendix Table B2-B4). The braking distance was also shorter for Segways than for e-scooters. However, this is not a valid comparison, as participants on the Segway were only able to reach 15 km/h (despite the Segway design allowing for higher speeds), and therefore the shorter distance is likely a consequence of the lower speed (Fig. 5). No statistically significant effect of gender or age was found on the braking distance (Appendix Table B2-B4). The response times were not only similar across gender and age, but also across all vehicles—with the exception of the e-scooter, which induced statistically significantly larger response times (Fig. 6; $F = 6.7$; $p \ll 0.01$; Appendix Table B5). Further, during the harsh (planned) braking task, participants riding the e-scooter crossed the stop line marked on the ground 61% of the time, while this line was exceeded only 18% and 14% of the time for the assisted and unassisted bicycle, respectively. Finally, while they were riding the Segway, they crossed the line in 71% of the trials.

3.4. Subjective data

Table 4 shows some of the results from the questionnaire probing the participants' opinions of the vehicles' performance in different situations. The electrified vehicles, possibly because they required less physical effort, were perceived as more comfortable than the unassisted bicycle when accelerating from a standstill (this result was statistically significant; Appendix Table S1-S6). The assisted and unassisted bicycle tasks were scored similarly in all other situations. The Segway scored statistically significantly lower than the other vehicles for mounting and dismounting, maintaining balance at high speed, and braking at high speed (Appendix Table S1-S6). While all vehicles were similarly rated by the participants at low speed, the e-scooter and the Segway were perceived as less stable as speed increased (both for simply balancing and for braking). Gender did not statistically significantly influence any of the categories in Table 4 (Appendix Table S1-S6). However, age did: the older the subject, the lower the ratings (Appendix Table S1-S6). Nevertheless, the effect of age was small compared to the effect of vehicle type (Appendix Table S1-S6).

The Segway also scored lower than the other vehicles for overall stability, maneuverability, comfort, and safety (Table 4); these differences, too, were statistically significant (Appendix Table S7-S10). The e-scooter was also perceived as less stable and safe than the assisted and unassisted bicycle; however, this result was on the border for statistical significance. The effect of gender was not statistically significant for comfort, stability, maneuverability, or safety, but the effect of age was (Appendix Table S7-S10). Specifically, the older the subject, the less comfortable, stable, maneuver-

Table 1
Experience of the participants with riding vehicles. (For the Segway, we reported the data only from the 25 subjects that contributed to the questionnaire analysis.).

	Bike	e-bike	e-scooter	Segway
Never	2	27	12	24
Few days per year	7	4	7	0
Few days per month	8	2	8	1
Few days per week	9	1	4	0
Everyday	8	0	3	0

able, and safe the vehicle ranking. (Notably, these effects were most pronounced for the safety category and for the Segway.) Table 5 reports the correlation matrix for the four categories presented at the bottom of Table 4; it may be observed that the correlation was high among all categories, particularly between safety and stability.

4. Discussion

In this study, we applied the procedure for data collection and analysis from Dozza et al.'s (2022) field study to the comparison of the longitudinal control of a bicycle (with and without assisted pedaling), an e-scooter, and a Segway. Our results show that, indeed, the same participant may demonstrate different acceleration and braking performance depending on the vehicle. Nevertheless, we also verified that, independently of the vehicle and of the emergency of the maneuver, riders braked with a constant deceleration (this is evident from the very large R^2 coefficients in all models; see Tables 2 and 3). This finding, in line with previous work on bicycle dynamics (Lee et al., 2020), is important for the application of our models to active safety: the linear coefficients from our regression analysis can accurately predict micro-mobility kinematics—specifically, stopping distance. In other words, an (automated) vehicle using our models may estimate whether a rider approaching an intersection is still able to brake and stop in time to avoid a collision and, once the rider starts braking, what the trajectory is going to be (Boda et al., 2020). The data collected from the comfort maneuvers in our experiment may provide a lower bound for these predictions for the threat assessment of an active safety system, and the harsh maneuvers may estimate a higher bound. Further, this paper shows that vehicle classification is essential for an accurate prediction, because the braking and acceleration performances vary largely across the micromobility vehicles tested.

Riders could accelerate almost twice as fast and brake twice as hard when they compromised comfort for urgency (i.e., comfort vs harsh conditions). While this is the first study, to our knowledge, presenting acceleration data from micromobility vehicles, previous studies assessed braking performance. In particular, Dozza et al. (2022) presented results from six cyclists/e-scooterists braking in the same conditions as in this study (i.e., comfort, harsh, and unexpected), and the results are very similar, although the (unassisted) bicycle's harsh braking in their study resulted in a somewhat higher deceleration rate than was found in this study. Because this study had a larger number of subjects, the average value given here is likely to be more accurate than the one presented there. In any case, their results are still well within one standard deviation of this study's, and even this small difference may be explained by the small data sample. Interestingly, both studies found that, in unexpected braking, riders achieve slightly larger deceleration than in planned braking. It is, however, unknown whether the larger deceleration is caused by suggestion (from the expectation of the experimenter's command) or by some other mechanism. The results for bicycle expected braking in this study were similar to those already reported by Lee et al. (2020) from 16 riders, while the results for e-scooter braking were in line with those reported from eight riders by Garman et al. (2020).

Braking performance, in terms of decelerations and braking distances, was similar for the assisted and unassisted bicycle tasks

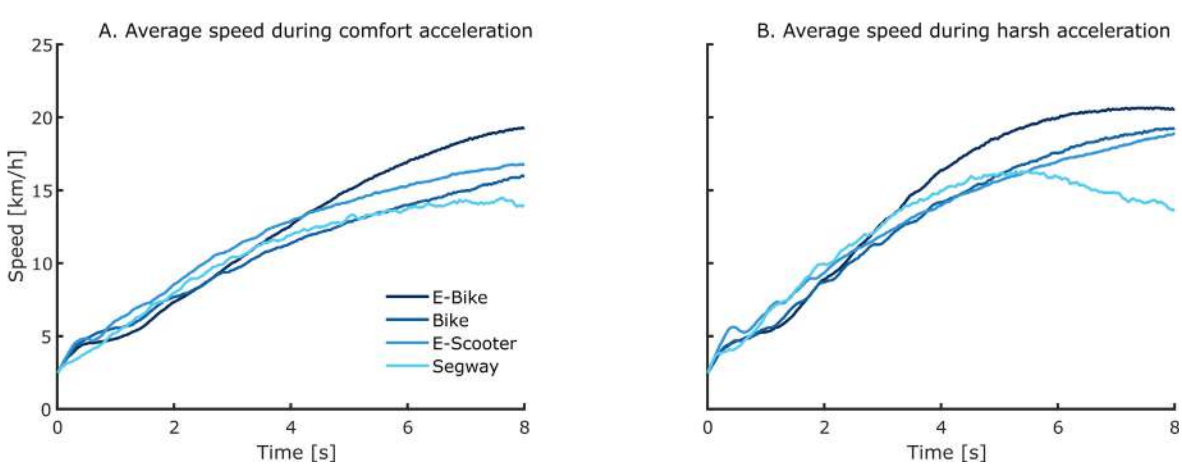


Fig. 3. Average speed for comfort and harsh accelerating maneuvers for all vehicles.

Table 2
Average accelerations (M) with standard deviations (SD) expressed in m/s^2. N indicates the number of trials available for computing averages and standard deviations. We also report the average R2 coefficients to show the goodness of fitness of the linear models.

maneuver	bicycle (M ± SD)	e-bicycle (M ± SD)	e-scooter (M ± SD)	Segway (M ± SD)
Comfort	0.45 ± 0.11 (N = 22; R^2 = 0.96)	0.70 ± 0.12 (N = 26; R^2 = 0.98)	0.56 ± 0.19 (N = 25; R^2 = 0.94)	0.67 ± 0.36 (N = 13; R^2 = 0.93)
Harsh	0.76 ± 0.28 (N = 25; R^2 = 0.96)	0.95 ± 0.14 (N = 26; R^2 = 0.95)	0.70 ± 0.25 (N = 28; R^2 = 0.93)	1.01 ± 0.34 (N = 13; R^2 = 0.95)

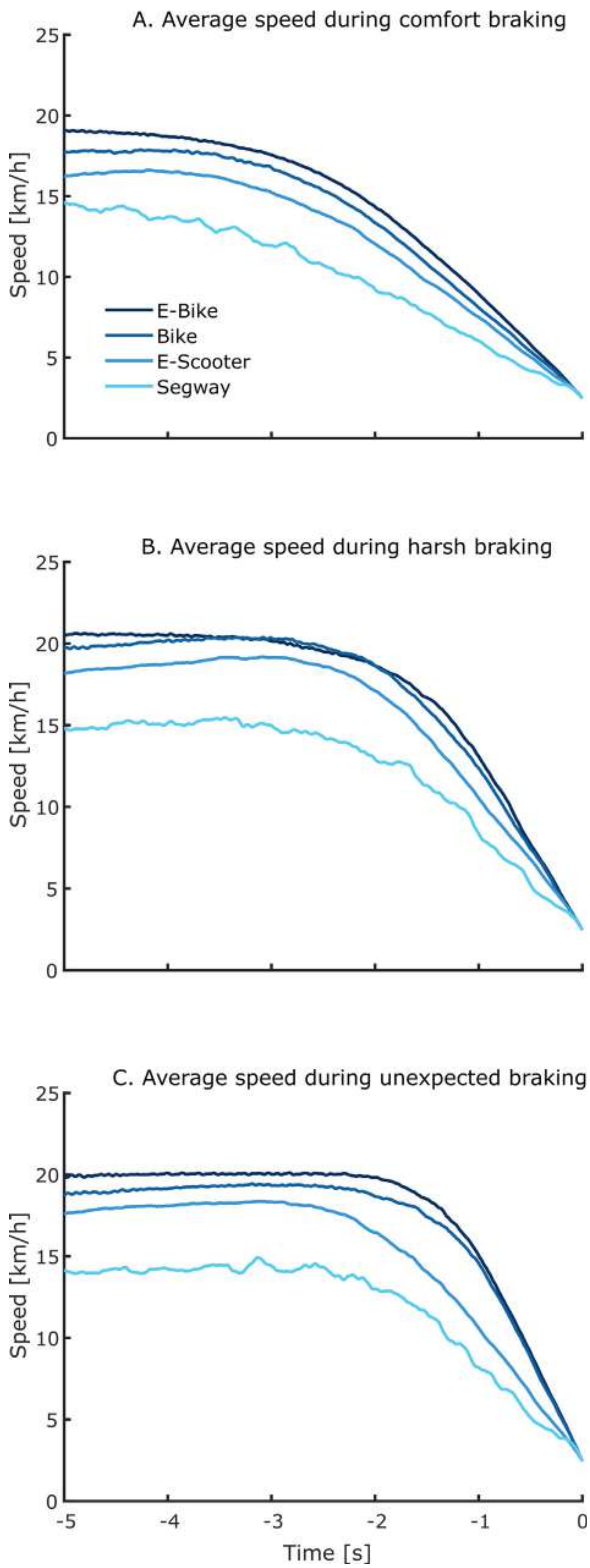


Fig. 4. Average speed for comfort, harsh, and unexpected braking maneuvers for all vehicles.

and poorer for the e-scooter and Segway. Both objective and subjective data suggest that the Segway is less stable and harder to maneuver than the other vehicles. Further, e-scooters seem to be harder to control than bicycles, both because 61% of riders were not able to halt before the stop line and because response times for braking were longer for e-scooters than for all other vehicles. Although steering performance (which is not addressed in this paper) may redeem e-scooters' maneuverability, when it comes to longitudinal control (i.e., crash avoidance by braking), e-scooters and Segways perform much more poorly than bicycles, which raises some concerns about their safety. Riders seem to be aware of these limitations, because the questionnaire data clearly indicate that riders perceive e-scooters and Segways as less stable and safe than bicycles. This result is positive, because riders may be able to use their awareness to compensate for the inferior braking performance by braking in anticipation (earlier), for instance, or using other crash-avoidance strategies.

This study verified that accelerative and braking maneuvers on micromobility vehicles are highly predictable because riders tend to control the vehicles by maintaining constant accelerations. Although the accelerations may change depending on the vehicle and the urgency of the maneuvers, the constancy may make it possible for active safety systems (and automated vehicles) to predict cyclist trajectories. This may be particularly critical at an intersection: a vehicle may estimate the probability that a crossing cyclist will stop at the intersection in time and use this information to warn the driver or apply automated interventions, such as emergency steering and braking (Thalya et al., 2020). Further, by including our models in the threat assessment for warning and intervention systems (SAE J3063), current systems intended to avoid crashes with motorized vehicles (e.g., frontal collision warning, automated emergency braking) may be adapted to also avoid crashes with micromobility vehicles (Boda et al., 2018). As consumer rating programs such as Euro NCAP include new test scenarios with new vulnerable-road-users (Van Ratingen et al., 2016), the results in this paper may be used to derive test scenarios that specify the safety-relevant differences between micromobility solutions. Finally, dynamic geofencing (i.e., algorithms that can remotely control micromobility, for example, by limiting speed) may make use of the data from this study to determine which speeds are safe for different vehicles, depending on the time of day and the location of the rider.

Experience is fundamental for safe riding, especially for new micromobility vehicles (Austin Public Health, 2019). Similarities across vehicles may help a rider to master a new vehicle in a short time. For instance, our participants were much more experienced with traditional bicycles than with electrical bicycles; nevertheless, they perceived the two vehicles similarly and mastered the bicycle equally well with and without assistance. Previous experience from riding a bicycle may not have ported equally well to e-scooters, because the controls and the geometry are very different. Indeed, riding a bicycle is an overlearned skill that required a relatively long time to develop, and we do not know whether riding an e-scooter for the first time would be equally challenging for a rider who does not know how to ride a bicycle. Future studies should investigate the extent to which experience from riding a bicycle may transfer to e-scooter riding and whether, in critical situations, such previous experience may lead to suboptimal avoidance maneuvers (Adams, 1987).

If cycling skills transferred to e-scooter riding, they certainly did not help much with riding a Segway. Only 25 participants completed the experiment with the Segway, and none of them reached the 17–20 km/h speed set by the experimental protocol (although it should have been possible for the Segway to reach this range, according to the specifications from the manufacturer). All participants rated this vehicle lower than all the others for comfort, stability, maneuverability, and safety. Although our correlation analysis shows that these categories are not orthogonal at all, this

Table 3
Average acceleration (M) with standard deviations (SD) expressed in m/s^2. N indicates the number of trials available for computing averages and standard deviations. We also report the average R2 coefficients to show the goodness of fitness of the linear models.

maneuver	bicycle (M ± SD)	e-bicycle (M ± SD)	e-scooter (M ± SD)	Segway (M ± SD)
Comfort	−1.50 ± 0.51 (N = 18; R^2 = 0.97)	−1.65 ± 0.66 (N = 26; R^2 = 0.98)	−1.28 ± 0.42 (N = 20; R^2 = 0.98)	−0.93 ± 0.40 (N = 11; R^2 = 0.96)
Harsh planned	−3.00 ± 1.29 (N = 25; R^2 = 0.98)	−3.10 ± 1.25 (N = 26; R^2 = 0.97)	−2.21 ± 0.59 (N = 28; R^2 = 0.98)	−1.65 ± 0.59 (N = 14; R^2 = 0.93)
Unexpected	−3.60 ± 1.28 (N = 24; R^2 = 0.97)	−3.66 ± 1.07 (N = 24; R^2 = 0.99)	−2.23 ± 0.71 (N = 28; R^2 = 0.99)	−1.60 ± 0.49 (N = 11; R^2 = 0.94)

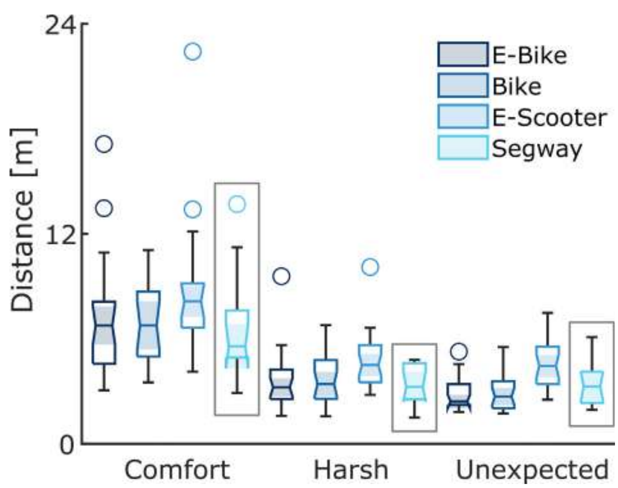


Fig. 5. Boxplots of braking distances for all vehicle types. Circles indicate outliers, whiskers are set by the non-outlier minima and maxima of the distribution, and the center line represents the median, while the horizontal edges of the box are the 25th and 75th percentiles. The notches, highlighted with shading, indicate the confidence intervals. (These boxplots were generated with the boxchart command in Matlab; please refer to its documentation for more detailed information.) The data from the Segway are surrounded by a box to remind the reader that a direct comparison with the other vehicles may be misleading in this specific analysis because the Segway started braking at a lower speed compared to the other vehicles and only few subjects were included in the analysis.

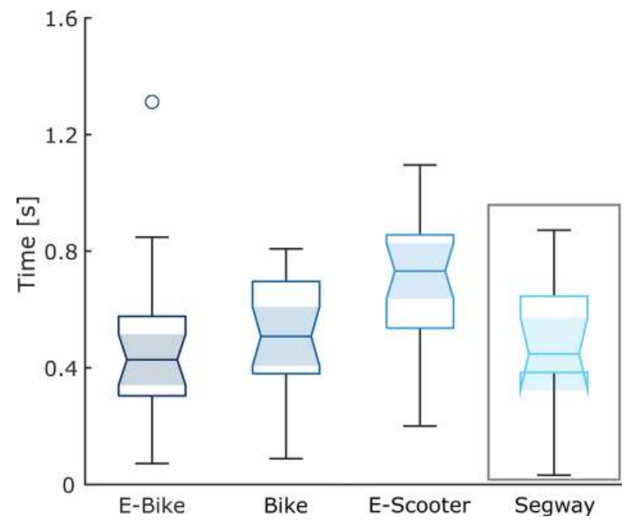


Fig. 6. Response time in unexpected braking across vehicles. Circles indicate outliers, whiskers are set by the non-outlier minima and maxima of the distribution, and the center line represents the median, while the horizontal edges of the box are the 25th and 75th percentiles. The notches, highlighted with shading, indicate the confidence intervals. (These boxplots were generated with the boxchart command in Matlab; please refer to its documentation for more detailed information.) The data from the Segway are surrounded by a box to remind the reader that a direct comparison with the other vehicles may be misleading in this specific analysis because the Segway started braking at a lower speed compared to the other vehicles and only few subjects were included in the analysis.

result is reasonable because the Segway has a different geometry compared to the other vehicles, and its pitch fluctuations may take a while to get used to. None of the participants were acquainted with this vehicle before the experiment, and we do not know whether their inexperience may have affected our results. Nevertheless, this example shows the importance of training on new micromobility vehicles that may look intuitive to ride but are still dangerous, especially on the very first rides, as the report from Austin Public Health (2019) showed for e-scooters. The results presented in this paper suggest that practicing braking to a line marked on the ground and using the possible overshoot distance as feedback may be an easy and useful training for novice Segway users (and possibly for any kind of micromobility vehicle).

E-scooters are mainly ridden by young males (6t-bureau de recherche, 2019; Bjerkan et al., 2020); however, the number of female riders is not negligible. Our study failed to show any statistically significant difference in how female and male riders longitudinally controlled the bicycle, e-bicycle, e-scooter, or Segway. Further, braking distances and response times were similar across gender, and the percentage of females that completed the experiment with the Segway was similar to that of the other vehicles. All in all, we could not verify the common hypothesis that males ride more aggressively or take higher risks than female riders. We did, however, find some effect of age on the subjective data; specifically, the older the subjects were, the lower their ratings were for the e-scooter and the Segway. Although the age span in this study was not large and the effect of age was minor when compared to the effect of vehicle type, our results suggest that younger people are more positive about new micromobility vehicles than older people. This result appears to be in line with previous studies that showed that elderly people are particularly averse to e-scooters (Portland Bureau of Transportation, 2018).

In this study, we lost about 50% of the data from the Segway, in part because the participants were not able to master it; we also lost up to 20% of the data from the other vehicles, mainly because participants had difficulty controlling the speed as they were instructed to. Although this amount of data loss is common in field trials, it may have biased the dataset toward a particularly athletic or daring sub-population of participants, especially for the Segway. It is also worth noting that the experiment was challenging; the one participant who crashed reported a minor injury. While we still believe that the value of this experiment justifies the crash risk that we asked the participants to take, we recommend that the research community not underestimate the risks in these experiments and make sure that the participants are insured.

Although we are not aware of any other study with a larger number of subjects for e-scooter field trials, our sample of 34 participants may not be representative of all ages and geographical locations. In addition, because we collected data in a controlled

Table 4
Average values and ranges of the subjective data for all vehicles (from 1 = Very poor to 7 = Exceptional).

	Bike	E-Bike	E-Scooter	Segway
Accelerating from standing still	4.36 (1–7)	5.64 (2–7)	5.46 (2–7)	5.16 (3–7)
Braking at low speed	5.64 (2–7)	5.70 (2–7)	5.12 (2–7)	4.92 (2–7)
Braking at high speed	5.21 (2–7)	5.33 (2–7)	4.03 (1–7)	3.48 (1–6)
Mounting and dismounting	4.91 (2–7)	5.03 (2–7)	5.67 (2–7)	3.60 (1–7)
Keeping balance at high speed	6.15 (4–7)	6.27 (4–7)	5.70 (3–7)	4.88 (1–7)
Keeping balance at low speed	5.18 (2–7)	5.24 (2–7)	5.30 (2–7)	5.12 (2–7)
Overall comfort	5.33 (2–7)	5.85 (3–7)	5.36 (3–7)	4.60 (2–7)
Overall stability	5.88 (3–7)	5.82 (3–7)	5.33 (3–7)	4.28 (1–7)
Overall maneuverability	5.27 (3–7)	5.46 (3–7)	5.33 (3–7)	4.64 (2–7)
Overall safety	5.85 (3–7)	5.64 (3–7)	4.82 (2–7)	3.80 (1–6)

Table 5
Correlation matrix among the subjective ratings for comfort, stability, maneuverability, and safety. (All coefficients are statistically significant.).

Measure	1	2	3
1.Comfort	–		
2.Stability	0.70	–	
3.Maneuverability	0.66	0.72	–
4.Safety	0.68	0.75	0.61

environment and in repetitive tasks, our results may be biased by the lack of other road users in the surroundings, as well as by the expectancy and habituation that the participants may have developed during the experiment. We presented results for e-bicycles and bicycles; however, we tested the same bicycle with and without electrical assistance. While this choice preserved the vehicle geometry across trials, the e-bicycle was heavier than a conventional bicycle (because of the battery and the motor) and therefore may have been less maneuverable. The e-scooter in this study is representative of the e-scooters that individuals purchase for personal use in Sweden; however, it differs from most of the e-scooters available in sharing systems. Such differences include: suspensions, wheel size, and brakes. Future studies should compare different e-scooter models to determine whether the difference in components affects safety. For instance, the longer response time for e-scooters compared to the other vehicles in this study may be a consequence of the electric braking system of the particular type of e-scooter used.

5. Conclusions and Practical Applications

This study provides further evidence that field data can support the safe integration of micromobility in the transport system. Field data show that different micromobility solutions affect rider behavior in multiple ways and create different constraints for vehicle control. Because e-scooters may brake more poorly than bicycles, steering maneuvers may be a better crash-avoidance strategy for e-scooterists than braking even in situations when a cyclist would be safer braking. Consequently, *infrastructure* that is forgiving of vehicles that run off the road may increase e-scooterists' safety.

The Segway vehicle employed in this study performed poorly in the field trials, and the participants ranked this vehicle as the least comfortable, stable, maneuverable, and safe. Nevertheless, Segways and other two-track vehicles with two wheels are popular, possibly because some of the issues with their safety and stability may disappear with enough training. Therefore, it may be important to *educate* novice users of micromobility vehicles and make sure they ride in real traffic only after a sufficient period of training. The design of the required training to facilitate learning and controlling the new micromobility solutions may be supported by field data such as were presented in this paper.

Because crash avoidance is the best way to avoid injuries when cars share the infrastructure with vulnerable road users, *active safety systems*, and automated emergency braking specifically, should make use of the models from this paper in their threat assessment. This study proved that riders keep accelerations (and decelerations) constant in comfortable and harsh maneuvers; therefore, their trajectories can be reliably predicted by (automated) vehicles. The models presented in this paper provide an indication of the longitudinal control performances (and their variability) that vehicles may expect from micromobility users.

Consumer rating programs, such as the one run by Euro NCAP, may use the models from this study to design the experimental protocols to test crash avoidance systems, such as emergency braking and steering. Further, as tests move toward simulations, the models from this paper may inform the behavior of the *virtual* micromobility users that Euro NCAP may introduce in future test simulations.

As novel micromobility vehicles hit the market and join a transport system where vehicles are increasingly *automated and connected*, it becomes increasingly important to model human behavior so that vehicles may understand and predict it, improving safety for all road users.

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

We would like to thank Christian-Nils Åkerberg Boda for help supervising the students who collected the data. We also thank K. Mayberry for language edits. Finally, we acknowledge Alessio Violin, Estelle Aimar, Ashwij M. Rao, Nithin Revadal, Weicheng Xiao, and Di Xue for their work on the experimental protocol, the instrumentation of the vehicles, and the logging platform.

The data collection was partially sponsored by Trafikverket via the project *Characterizing and classifying new e-vehicles for personal mobility* (2019/21327), while the data analysis and the modeling effort were sponsored by Toyota Motor Europe via the *DICE (driver interaction with cyclists and e-scooterists)* project and Chalmers. The work was carried out at Chalmers University of Technology, Gothenburg, Sweden.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsr.2022.10.005.

References

6t-bureau de recherche. (2019). *Uses and Users of Free-floating Electric Scooters in France*.

Abduljabbar, R. L., Liyanage, S., & Dia, H. (2021). The role of micro-mobility in shaping sustainable cities: A systematic literature review. *Transportation Research Part D: Transport and Environment, 92*. https://doi.org/10.1016/j.trd.2021.102734.

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin, 101*(1), 41.

Austin Public Health. (2019). *Dockless electric scooter - related injuries study*.

Badeau, A., Carman, C., Newman, M., Steenblik, J., Carlson, M., & Madsen, T. (2019). Emergency department visits for electric scooter-related injuries after introduction of an urban rental program. *American Journal of Emergency Medicine, 37*(8), 1531–1533. https://doi.org/10.1016/j.ajem.2019.05.003.

Beck, S., Barker, L., Chan, A., & Stanbridge, S. (2020). Emergency department impact following the introduction of an electric scooter sharing service. *Emergency Medicine Australasia, 32*(3), 409–415. https://doi.org/10.1111/1742-6723.13419.

Bekhit, M. N. Z., Le Fevre, J., & Bergin, C. J. (2020). Regional healthcare costs and burden of injury associated with electric scooters. *Injury, 51*(2), 271–277.

Billstein, L., & Svernlöv, C. (2021). *Evaluating the Safety and Performance of Electric Micro-Mobility Vehicles: Comparing E-bike, E-scooter and Segway based on Objective and Subjective Data from a Field Experiment*.

Bjerkan, A. M., Engebretsen, A., & Steinbakk, R. T. (2020). *Skader på sykkel og elektrisk sparkesykkel i Oslo*.

Boda, C.-N., Dozza, M., Bohman, K., Thalya, P., Larsson, A., & Lubbe, N. (2018). Modelling how drivers respond to a bicyclist crossing their path at an intersection: How do test track and driving simulator compare? *Accident Analysis & Prevention, 111*, 238–250.

Boda, C.-N., Lehtonen, E., & Dozza, M. (2020). A computational driver model to predict driver control at unsignalised intersections. *IEEE Access, 8*, 104619–104631.

Chang, A. Y. J., Miranda-Moreno, L., Clewlow, R., & Sun, L. (2019). Trend or Fad? Deciphering the Enablers of Micromobility in the US. *SAE International*.

Cicchino, J. B., Kulie, P. E., & McCarthy, M. L. (2021). Injuries related to electric scooter and bicycle use in a Washington, DC, emergency department. *Traffic Injury Prevention, 22*(5), 401–406.

Dozza, M., Violin, A., & Rasch, A. (2022). A data-driven framework for the safe integration of micro-mobility into the transport system: Comparing bicycles and e-scooters in field trials. *Journal of Safety Research*.

Dozza, M., & Werneke, J. (2014). Introducing naturalistic cycling data: What factors influence bicyclists' safety in the real world? *Transportation Research Part F: Traffic Psychology and Behaviour, 24*, 83–91.

Fearnley, N., Berge, S. H., & Johnsson, E. (2020). Delte elsparkesykler i Oslo. *Retrieved from Oslo:* https://www.Toi.No/Getfile.Php.

Fitt, H., & Curl, A. (2020). The early days of shared micromobility: A social practices approach. *Journal of Transport Geography, 86*. https://doi.org/10.1016/j.jtrangeo.2020.102779.

Garman, C., Como, S. G., Campbell, I. C., Wishart, J., O'Brien, K., & McLean, S. (2020). Micro-Mobility Vehicle Dynamics and Rider Kinematics during Electric Scooter Riding. *SAE Technical Papers*. https://doi.org/10.4271/2020-01-0935.

Gössling, S. (2020). Integrating e-scooters in urban transportation: Problems, policies, and the prospect of system change. *Transportation Research Part D: Transport and Environment, 79*, 102230.

Huertas-Leyva, P., Dozza, M., & Baldanzini, N. (2018). Investigating cycling kinematics and braking maneuvers in the real world: E-bikes make cyclists move faster, brake harder, and experience new conflicts. *Transportation Research Part F: Traffic Psychology and Behaviour.*. https://doi.org/10.1016/j.trf.2018.02.008.

Huertas-Leyva, P., Dozza, M., & Baldanzini, N. (2019). E-bikers' braking behavior: Results from a naturalistic cycling study. *Traffic Injury Prevention*, 1–6.

Ishmael, C. R., Hsiue, P. P., Zoller, S. D., Wang, P., Hori, K. R., Gatto, J. D., ... Bernthal, N. M. (2020). An early look at operative orthopaedic injuries associated with electric scooter accidents: Bringing high-energy trauma to a wider audience. *JBJS, 102*(5), e18–e.

Kovácsová, N., de Winter, J. C. F., Schwab, A. L., Christoph, M., Twisk, D. A. M., & Hagenzieker, M. P. (2016). Riding performance on a conventional bicycle and a pedelec in low speed exercises: Objective and subjective evaluation of middle-aged and older persons. *Transportation Research Part F: Traffic Psychology and Behaviour, 42*, 28–43. https://doi.org/10.1016/j.trf.2016.06.018.

Lee, O., Rasch, A., Schwab, A. L., & Dozza, M. (2020). Modelling cyclists' comfort zones from obstacle avoidance manoeuvres. *Accident Analysis & Prevention, 144*, 105609.

MacArthur, J. H., Dill, J., & Person, M. (2014). *E-Bikes in the North America: Results from an Online Survey*. 15p. http://docs.trb.org/prp/14-4885.pdf.

Namiri, N., Lui, H., Tangney, T., Allen, I., Cohen, A., & Breyer, B. (2020). Electric Scooter Injuries and Hospital Admissions in the United States, 2014–2018. *JAMA Surgery, 155*. https://doi.org/10.1001/jamasurg.2019.5423.

O'Hern, S., & Estgfaeller, N. (2020). A scientometric review of powered micromobility. *Sustainability, 12*(22), 9505.

Portland Bureau of Transportation. (2018). *2018 E-SCOOTER PILOT User Survey Results*.

Rasch, A., Dozza, M., & Håland, Y. (2016). *Evaluation of a new narrow and tiltable electric tricycle (e-trike) concept*.

Rauch, H. E., Tung, F., & Striebel, C. T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal, 3*(8), 1445–1450.

SAE Committee, S. and D. M. (2018). Taxonomy and Definitions for Terms Related to Shared Mobility and Enabling Technologies. *SAE International*.

Santacreu, A., Yannis, G., de Saint Leon, O., & Crist, P. (2020). *Safe micromobility*.

Schleinitz, K., Petzoldt, T., Franke-Bartholdt, L., Krems, J., & Gehlert, T. (2017). The German Naturalistic Cycling Study – Comparing cycling speed of riders of different e-bikes and conventional bicycles. *Safety Science, 92*, 290–297. https://doi.org/10.1016/j.ssci.2015.07.027.

Sharkey, M., Ahrens, P., Munk, S., & Bogdanowicz, T. (2020). *Micromobility and active travel in the UK: The rise of smaller modes, and resulting infrastructure, safety and regulation implications*.

Stigson, H., Malakuti, I., & Klingegård, M. (2020). *Electric scooters injuries and causes: analysis of Swedish accident data* .

Thalya, P., Kovaceva, J., Knauss, A., Lubbe, N., & Dozza, M. (2020). Modeling driver behavior in interactions with other road users",. *Transportation Research Arena*.

Trivedi, B., Kesterke, M. J., Bhattacharjee, R., Weber, W., Mynar, K., & Reddy, L. V. (2019). Craniofacial injuries seen with the introduction of bicycle-share electric scooters in an urban setting. *Journal of Oral and Maxillofacial Surgery, 77*(11), 2292–2297.

Trivedi, T. K., Liu, C., Antonio, A. L. M., Wheaton, N., Kreger, V., Yap, A., ... Elmore, J. G. (2019). Injuries associated with standing electric scooter use. *JAMA Network Open, 2*(1), e187381–e.

Twisk, D., Stelling, A., Van Gent, P., De Groot, J., & Vlakveld, W. (2021). Speed characteristics of speed pedelecs, pedelecs and conventional bicycles in naturalistic urban and rural traffic conditions. *Accident Analysis & Prevention, 150*, 105940.

Van Ratingen, M., Williams, A., Lie, A., Seeck, A., Castaing, P., Kolke, R., ... Miller, A. (2016). The European New Car Assessment Programme: A historical review. *Chinese Journal of Traumatology - English Edition, 19*(2), 63–69. https://doi.org/10.1016/j.cjtee.2015.11.016.

Wüster, J., Voß, J., Koerdt, S., Beck-Broichsitter, B., Kreutzer, K., Märdian, S., Lindner, T., Heiland, M., & Doll, C. (2020). Impact of the Rising Number of Rentable E-scooter Accidents on Emergency Care in Berlin 6 Months after the Introduction: A Maxillofacial Perspective 1943387520940180–1943387520940180. *Craniomaxillofacial Trauma & Reconstruction*.

Marco Dozza received the Ph.D. degree in bioengineering from the University of Bologna, Italy, in collaboration with Oregon Health & Science University, Portland, OR, USA, in 2007. After graduation, he worked as a System Developer with Volvo Technology, for over two years, a research and innovation company inside the Volvo group. Since 2009, he has been with the Chalmers University of Technology, Gothenburg, Sweden, where he is currently a Professor and leads the Crash Analysis and Prevention Unit. He is an Examiner for the course Active Safety in the Master's Programme for Automotive Engineering. He is working with SAFER, the Vehicle and Traffic Safety Center at Chalmers..

Tianyou Li received the B.E. degree in automation from Huazhong University of Science and Technology, Wuhan, China, in 2019, and the M.Sc. degree in interactive media technology from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2021. Tianyou is currently pursuing the Ph.D. degree on computational models for driver behaviour in interaction with cyclists and e-scooterists, mainly in intersection scenarios. The goal of his research is to develop driver models that can improve advanced driver assistance systems and to inform the design of relevant safety test.

Lucas Billstein received the B. Sc. degree in automation and mechatronics in 2019, and the M. Sc. degree in systems, control and mechatronics in 2021 from Chalmers University of Technology, Gothenburg, Sweden. After graduating he started working as a software engineer at the consulting company, Conmore AB. His current assignment is at Volvo Cars as a software tester within charging of electric cars.

Christoffer Svernlöv received the B. Sc. degree in automation and mechatronics in 2019, and the M. Sc. degree in systems, control and mechatronics in 2021 from Chalmers University of Technology, Gothenburg, Sweden. Christoffer did his master's thesis at the division of vehicle safety where the goal was to collect data and analyze the safety of micro-mobility vehicles. He is currently working as a software engineer consultant in the telecommunication industry.

Alexander Rasch received the B.Sc. degree in mechanical engineering from RWTH Aachen University, Aachen, Germany, in 2016, and the M.Sc. degree in systems, control and mechatronics from the Chalmers University of Technology, Gothenburg, Sweden, in 2018, where he is currently pursuing the Ph.D. degree, on computational models for driver behavior in interaction with pedestrians and cyclists, mainly in overtaking scenarios. The goal of his research is to develop driver models that can improve advanced driver assistance systems.

Journal of Safety Research 84 (2023) 74–85

Contents lists available at ScienceDirect

Journal of Safety Research

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

Injury severity analysis of time-of-day fluctuations and temporal volatility in reverse sideswipe collisions: A random parameter model with heterogeneous means and heteroscedastic variances

Chengying Hua [a], Wei (David) Fan [b,*]

[a] USDOT Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE), Department of Civil and Environmental Engineering, University of North Carolina at Charlotte, EPIC Building, Room 3366, 9201 University City Boulevard, Charlotte, NC 28223-0001, United States

[b] USDOT Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE), Department of Civil and Environmental Engineering, University of North Carolina at Charlotte, EPIC Building, Room 3261, 9201 University City Boulevard, Charlotte, NC 28223-0001, United States

ARTICLE INFO

Article history:
Received 9 March 2022
Received in revised form 17 May 2022
Accepted 17 October 2022
Available online 28 October 2022

Keywords:
Crash injury severity
Sideswipe collisions
Temporal instability
Time of day
Random parameter

ABSTRACT

Problem: Sideswipe collisions in the opposite direction often result in more severe injuries than the typical same-direction crashes, especially when light trucks are involved. This study investigates the time-of-day fluctuations and temporal volatility of potential factors that affect the injury severity of reverse sideswipe collisions. *Methods:* A series of random parameters logit models with heterogeneous means and heteroscedastic variances are developed and utilized to explore unobserved heterogeneity inherent in variables and preclude biased parameter estimation. The segmentation of estimated results is also examined through temporal instability tests. *Results:* Based on crash data in North Carolina, a number of contributing factors are identified that have profound associations with obvious and moderate injuries. Meanwhile, significant temporal volatility is observed in the marginal effects of several factors such as driver restraint, alcohol or drugs impact, Sport Utility Vehicle (SUV) at fault, and adverse road surface across three different periods. Fluctuations in the time of day indicate that restraint with belts is more effective in mitigating the obvious injury in the nighttime, and high-class roadway sustains a higher probability of resulting in more serious injury compared to the daytime. *Practical Applications:* The findings of this study could help further guide the implementation of safety countermeasures related to atypical sideswipe collisions.

1. Introduction

Sideswipe collision refers to the sides of two vehicles sustaining engagement in a parallel path. Typically, it occurs when drivers attempt to change lanes and merge onto a highway in the same direction (Fitch et al., 2009) and result in lane drift or departure through an unintentional manner. Different from the front and rear bumpers that can absorb the collision energy in head-on or rear-end crashes, the exposed sides lack protection, which makes collisions more intense in a lateral manner. According to National Highway Traffic Safety Administration (NHTSA, 2018) statistics, nearly 863,000 sideswipe collisions occurred that year, accounting for 12.8% of all types of crashes. On average, 200 out of 1,000 sideswipe collisions result in varying degrees of injuries.

The potential hazard of sideswipe can be exacerbated in the opposite direction because opposing vehicles often have higher relative speeds (Kusano & Gabler, 2013). Despite accounting for only 20% of the whole sideswipe collisions, such atypical sideswipe occurs unexpectedly and is extremely severe, owing to its unpredictability and seriousness. This impact may be responsible for the initial impetus for a chain reaction where vehicles swerve over the centerline and hit other automobiles or fixed objects. The situation can be aggravated if light trucks (such as the Pickups and SUVs) are involved. Those larger vehicles generally suffer obstructed visibility due to the presence of blind spots, and large inertia makes collisions more intense. Fig. 1 shows the sideswipe collisions in reverse directions involving pickups and SUVs from 2008 to 2016. In general, the fatality or disabled rate for opposite and same direction sideswipe is 2.5% and 0.5%, respectively. However, the opposite direction can cause more evident injuries than the same direction (11% and 3%, respectively).

Previous studies emphasized predicting highway sideswipe-same-direction collision potential that arise from overtaking and lane changing maneuvers (Qu, Wang, Wang, & Liu, 2013). A majority of them utilized traffic parameters as variables in a general

* Corresponding author.
E-mail addresses: chua1@uncc.edu (C. Hua), wfan7@uncc.edu (Wei (David) Fan).

https://doi.org/10.1016/j.jsr.2022.10.009
0022-4375/

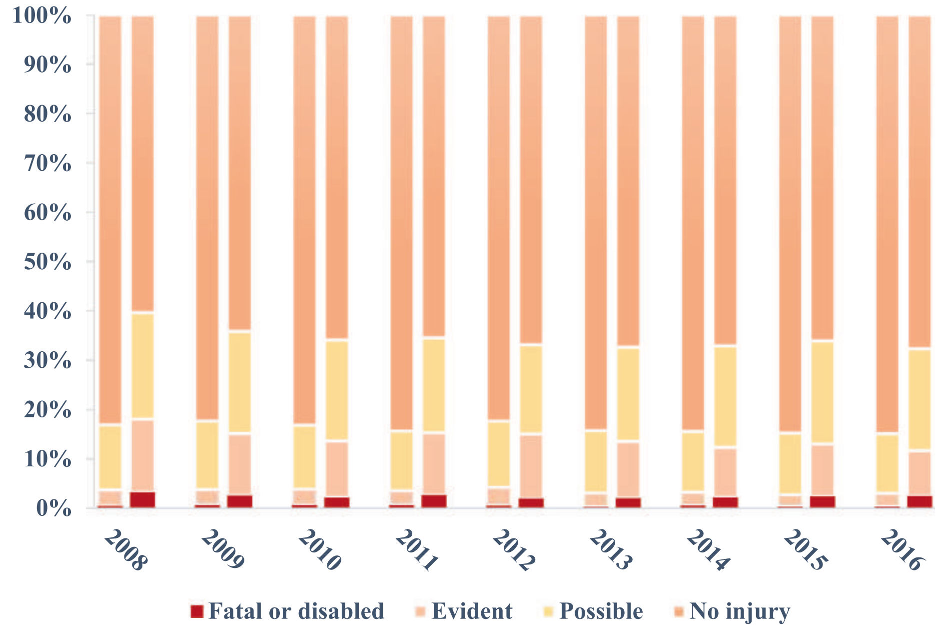


Fig. 1. Comparisons between sideswipe same-direction (left) and opposite-direction (right) from 2008 to 2016 with Highway Safety Information System data.

model (Pande & Abdel-Aty, 2006). Few studies have been undertaken on the modeling of injury severity in sideswipe-opposite-direction collisions, with even less exploration of unobserved heterogeneity among factors such as driver, vehicle, roadway, and environment characteristics. Mannering, Shankar, and Bhat (2016) pointed out that neglecting heterogeneous effects across observations may result in biased parameter estimation and inaccurate interpretation of explanatory variables.

Additionally, Mannering (2018) mentioned that peculiar temporal instability across various periods needs to be addressed. Ignoring such temporal volatility in the effects of contributing factors from year to year may lead to biased results. Another crucial problem is the fluctuations associated with time-of-day. Several studies employing segmented data have clearly revealed such variations in the factors affecting injury severity, including pedestrian-involved crashes (Mokhtarimousavi, 2019), driving under the influence (DUI) of alcohol/drugs crashes (Song, Fan, & Li, 2021), and truck involved crash (Zou, Wang, & Zhang, 2017). Silverstein, Schorr, and Hamdar (2016) indicated that daytime has potential effects in increasing the sideswipe collisions injury severities. However, time-of-day fluctuations for sideswipe collisions are still limited. Moreover, segmenting specific time-of-day and time periods is essential rather than just considering them as a variable in the single model. Such segmentation could reduce the temporal heterogeneity among the dataset due to the accumulation over time (Al-Bdairi, Behnood, & Hernandez, 2020; Islam, Alnawmasi, & Mannering, 2020).

The main objective of this study is to investigate the significant factors influencing injury severities of sideswipe-opposite-direction collisions and associated magnitudes involving pickups and SUVs. To study the time-of-day fluctuations and temporal volatility of factors, a case study of North Carolina is conducted based on the police-reported data during three periods. A series of random parameters logit models with heterogeneous mean and heteroscedastic variance are developed and employed to further explore unobserved heterogeneity and identify the injury severity determinants. The remainder of the paper is organized as follows: Section 2 summarizes the existing literature on the sideswipe collisions and introduces the state-of-art approaches on the injury severity analysis. Section 3 describes the data statistics and empirical settings. Section 4 describes the methodology used to model crash-injury severity. Section 5 presents the temporal instability results of the crash data. Section 6 discusses the model results and marginal effects of several factors. Finally, the insightful findings and implications are concluded in Section 7.

2. Literature review

This section illustrates the significance of this research from three aspects: (a) existing studies on sideswipe crashes in different collision directions; (b) explorations of temporal instability in the analysis of injury severity; and (c) approaches used in modeling the crash injury severity. Table 1 presents the literature reviewed in chronological order, details containing authors and years of publication, models, directions of collision, and corresponding key findings. Much effort has been put into investigating the same direction sideswipe collisions (Adanu et al., 2021; Khattak, Kantor, & Council, 1998; Kim, Washington, & Oh, 2006; Park & Ritchie, 2004; Silverstein et al., 2016), but a scarcity of studies considered the opposite direction scenario. Moreover, most previous studies utilized traffic parameters such as the volume, upstream or downstream speed, and density of adjacent lanes to forecast sideswipe crashes potential (Lee, Abdel-Aty, & Hsia, 2006; Li, Wang, Chen, & Liu, 2014); the effects of contributing factors associated with injury severities have not been fully explored. Actually, factors such as driver characteristics (gender, age, and years of experience), location and surrounding conditions (junction, light, and road surface), and roads feature (road type, number of lanes, and speed limit) should be paid more attention (Shawky, 2020). Hence, this study mainly focuses on various determining factors that might contribute to the establishment of feasible countermeasures that could avert the risky circumstances.

Currently, numerous studies have revealed potential temporal instabilities (if contributing factors vary over years) in crash records (Al-Bdairi et al., 2020; Islam et al., 2020; Mannering, 2018). Although variations in peak and off-peak hours have been examined (Lee et al., 2006; Wang, Zhang, Wang, Weng, & Yan, 2016), there is still a gap in the injury severity analysis of sideswipe collisions considering time-of-day fluctuations embedded in factors, especially for reverse scenarios. Furthermore, limited

Table 1
A summary of the literature on the sideswipe collisions.

Author, year	Model	Direction	Findings
Abdel-Aty, Keller, and Brady (2005)	Tree-based regression	Same and opposite	Daily traffic volume on both major and minor road, number of lanes, presence of median, and exclusive left-turn lane on major road were found important.
Pande and Abdel-Aty (2006)	Neural network based classification models	Same	Difference in occupancy on adjacent lanes, average speeds upstream and downstream of crash site, standard deviation of volume have an impact on the crash occurrence.
Kim et al. (2006)	Poisson and negative binomial regression	Same	Sideswipe is adversely related to median width on major roadways, whereas the existence of a left-turn lane and the density of surrounding driveways have higher frequencies.
Lee et al. (2006)	Logistic regression models	Same	Overall average flow ratio (OAFR) was generally higher for sideswipe than rear-end crashes, variation in flow and peak and off-peak periods were also important.
Kim, Lee, Washington, and Choi (2007)	Hierarchical binomial logistic	Same and opposite	The same direction is less common at junctions with shoulders and horizontal curves, but more frequent at right-angled intersections than skewed ones. The opposite-direction collisions are less likely in the daytime. Clear weather, wet roads, horizontal bends, and vertical curves are less likely to engage.
Ye, Pendyala, Washington, Konduri, and Oh (2009)	Multivariate Poisson regression model with multivariate normal heterogeneity	Same and opposite	For opposite-direction sideswipe, average daily traffic on minor road and number of right turn lanes on major roads were related. For same-direction sideswipe, speed limit, shoulder width, number of left-turn lanes, and sum of absolute change of grade on major road affected the frequency.
Geedipally, Patil, and Lord (2010)	Multinomial logit Model	Same and opposite	Sideswipe-passing crashes increased with lane width increased and decreased as AADT increased. Sideswipe-opposite decreased with AADT increased, but opposite for lane width.
Bham, Javvadi, and Manepalli (2012)	Multinomial logistic regression	Same and opposite	On undivided and divided roads, the probability of sideswipe-same-direction crashes increases while changing lanes and merging. Vision obstruction, night-time, horizontal curve, wet surface, weekends, and alcohol were negatively with sideswipe-same-direction.
Islam and Hernandez (2013)	Random-parameter ordered-probit model	Same	Sideswipes in the same direction are significant and are likely to result in less severe large truck collisions.
Silverstein et al. (2016)	Negative binomial regression and multinomial logit	Same	Work zones are more likely to result in fatalities than nonwork zones. Clear weather, daylight, and straight highways may enhance the likelihood.
Wang et al. (2016)	Case-control logistic regression	Same	It is more common during off-peak hours on straight and flat parts of multilane motorways. In crowded circumstances, high average occupancy, low volume flow, and speed variation upstream of collision spots can enhance the probability.
Chen, Qin, and Shaon (2018)	Case-control logistic regression	Same	The first downstream average flow into the target lane, the second downstream flow ratio, and snow conditions were related.
Shawky (2020)	Binary logistic regression	Same	Gender, nationality, and experience, as well as non-junction, light, and road conditions (road type, number of lanes, and speed limit), all have a substantial impact.
Adanu et al. (2021)	Mixed (random parameter) logit	Same	The darkness, elderly, female, and commercial vehicle drivers all enhance the likelihood of serious injury. Interstates with a greater number of travel lanes have a lower risk of major injury.

investigation has been dedicated to the combination of specific time-of-day fluctuations and temporal volatility of potential factors. Song et al. (2021) have indicated that merging temporal instability with time-of-day variations could result in accurate model estimations and discover extra details that might not be identified.

As is shown in Table 1, extensive research has employed the logistic regression model to model the crash injury severity since the effectiveness in parameter estimations and proper inferences. In terms of the unobserved heterogeneity in variables that impact injury severity, the most commonly applied method is the random parameter logit (RPL) model (Mannering et al., 2016). Nevertheless, parameter overestimation or underestimation may occur, reducing effectiveness when compared to RPL with heterogeneous mean and heteroscedastic variance (Islam et al., 2020). Varying the random parameters' means and variances has been proven statistically superior to the RPL-only method (Al-Bdairi et al., 2020). According to the abovementioned issues, by developing and utilizing a series of RPL models with heterogeneous means and heteroscedastic variances, this research combines the time-of-day fluctuation and temporal volatility of potential factors to analyze the injury severity of reverse sideswipe collisions.

3. Data descriptions and empirical settings

The crash data are extracted from the Highway Safety Information System (HSIS) and North Carolina is selected as a case study; a total of 22,295 observations are obtained. According to the 'Injury Classification Scale and Definitions' by the Federal Highway Administration (FHWA, 2019), this study divides injury severity into three levels (i.e., obvious injury [OI], moderate injury [MI], and no injury [NI]). Due to the relatively small sample size of the fatal or disabling injury, they are combined with the evident injury into the obvious injury, which is all serious enough at the scene. The drivers who were suffering the most serious injury are utilized to determine the severity in all crashes.

In this study, nine years (2008–2016) are retrieved from the database and three periods are defined (2008–2010, 2011–2013, and 2014–2016). Different from the economic cycles based segmentation in Behnood and Mannering (2016), this study is qualitatively based on the consistent fluctuation of crash patterns over time. Furthermore, transferability test results in section 5 and marginal effects of most significant factors in section 6 suggest that the 3-year clustering has robust temporal volatility in a quantitative way. This classification method can not only ensure that volatility will not be ignored due to time accumulation, but also could avoid the insufficient data issue caused by shorter periods. Fig. 2 demonstrates significant time-of-day fluctuations in reverse sideswipe collisions involving pickups and SUVs in North Carolina. The three curves for different periods show apparent variations during the peak hours in the morning and evening, which is similar to the traffic volume pattern in rush hours. According to the crash frequency and two peak hours, the time of day is divided into daytime (7:00 A.M.-6:00 P.M.) and nighttime (6:00 P.M.-7:00 A.M.). How-

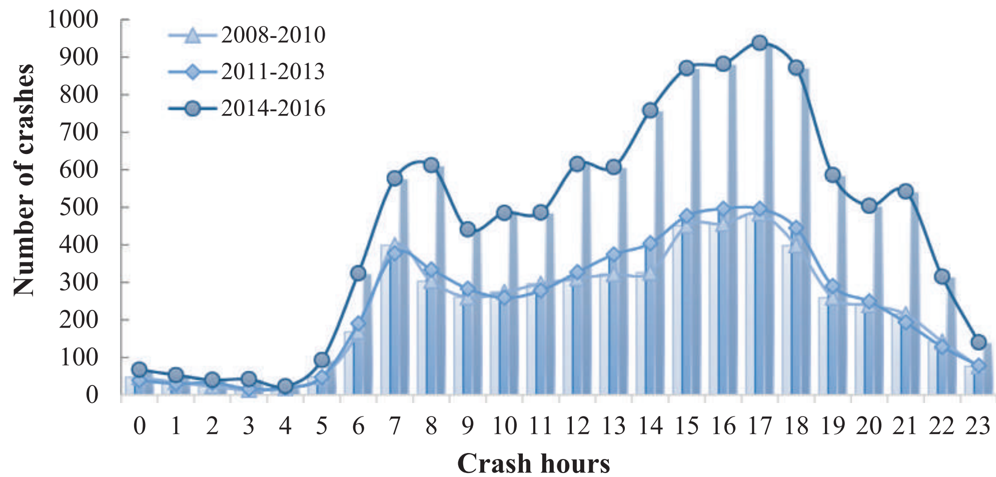


Fig. 2. Time-of-day fluctuations of reverse sideswipe collisions.

ever, the time distribution may not be in accord with the intuition that daytime visibility is clearer and therefore safer. One possible reason for this divergence might be differences in driving behaviors and the external environment, and the precise explanation can be explored by the model results.

Based on the examination of the temporal features, the whole dataset is split into six subgroups: 2008–2010 daytime, 2008–2010 nighttime, 2011–2013 daytime, 2011–2013 nighttime, 2014–2016 daytime, and 2014–2016 nighttime. Table 2 presents the percentage distribution by injury severity for each time of day and time period. It also describes the statistics of explanatory variables, which are classified into driver, vehicle, roadway, and environment characteristics, and 17 categories with 60 parameters are applied. Also, variables marked in bold are chosen as the base in the estimated model.

4. Methodology

The model used to analyze the injury severity of reverse sideswipe collisions is the random parameters logit model with heterogeneous mean and heteroscedastic variance, which is an extension of the random parameter logit model (also known as mixed logit model). The crash-injury-severity function is determined in a linear structure that identifies the severity k ($k = 1,2,3$) for observation i as:

$$U_{ki} = \beta_k X_{ki} + \varepsilon_{ki} \tag{1}$$

where U_{ki} represents the utility function,X_{ki} denotes a vector value of variables, and β_k is the estimated parameters for X_{ki}. In the multinomial logit model, β_k is assumed to be a constant that remains fixed across all observations. ε_{ki} indicates error term for the unobserved effect on the injury severity k, and ε_{ij} is supposed to follow a Gumbel distribution, the probability of severity k for individual i is (McFadden & Train, 2000):

$$P_i(k) = \int \frac{exp(\beta_k X_{ki})}{\sum_{i=1}^{k} exp(\beta_k X_{ki})} f(\beta|\varphi) d\beta \tag{2}$$

where $P_i(k)$ is the probability of observation i with the severity k, $f(\beta|\varphi)$ denotes the probability density function (PDF) of random parameter β, and φ is the mean and variance of the normal distribution for PDF. Moreover, the conventional random parameter models can be further extended by allowing for heterogeneity in the means and variances, based on the assumption that random parameters are distributed randomly across observations (Behnood & Mannering, 2017). The model can be defined as:

$$\beta_{ki} = \beta + \delta_{ki} \boldsymbol{Z_{ki}} + \sigma_{ki} \exp(\omega_{ki} \boldsymbol{W_{ki}}) \gamma_{ki} \tag{3}$$

where β represents the mean and determined over all observations. $\boldsymbol{Z_{ki}}$ is the variable vector that defines the heterogeneous means, and δ_{ki} is the vector of estimated coefficient for $\boldsymbol{Z_{ki}}$. For the heteroscedasticity in variances, $\boldsymbol{W_{ki}}$ is used to represent the variable vector with standard deviation σ_{ki}, ω_{ki} indicates the vector of estimated coefficient linked with $\boldsymbol{W_{ki}}$, and γ_{ki} is the disturbance term. All the significant variables can be defined in the $\boldsymbol{Z_{ki}}$ and $\boldsymbol{W_{ki}}$. If no variable is found statically significant in $\boldsymbol{W_{ki}}$, the model will be collapsed into the random parameter logit model (RPL) with heterogeneity in means only. Further, if no heterogeneous mean is found in the $\boldsymbol{Z_{ki}}$, the model will fall into the conventional RPL model, or the multinomial logit model if no random parameter is identified.

In this study, the parameters are estimated using a simulation-based maximum likelihood method, and 500 Halton draws are used due to the effectiveness. The normal distribution is used to explain the random parameters in the model because it has a better performance at convergence than the lognormal, uniform, and triangular distributions (Moore, Schneider, Savolainen, & Farzaneh, 2011).

In addition, the sign of the coefficients does not always indicate the direction of the intermediate results. Therefore, the marginal effect is utilized to explain the effects of significant variables on the injury severity, which is expressed as follows:

$$E_{X_{kij}}^{P_{ki}} = P_{ij}(X_{kij} = 1) - P_{ij}(X_{kij} = 0) \tag{4}$$

When the j-th binary indicator variable and X_{kij} are identified, the probability related to each severity k for individual i is calculated. For the variables with random parameters, only the mean value of the coefficients is used across all samples.

5. Temporal instability tests

To examine the temporal instability of significant variables that affect the injury severity of reverse sideswipe collisions across the time-of-day and different time ranges, three types of likelihood ratio (LR) transferability tests are utilized (Washington, Karlaftis, & Mannering, 2011). First, equation (5) is used to test the necessity of segmentation between daytime and nighttime,

$$X^2 = -2[LL(\beta_{whole}) - LL(\beta_{daytime}) - LL(\beta_{nighttime})] \tag{5}$$

where $LL(\beta_{whole})$ represents the log-likelihood with the entire time-of-day dataset in the converging model including converged parameters. $LL(\beta_{daytime})$and $LL(\beta_{nighttime})$ have the same expression to indi-

Table 2
Descriptive statistics for reverse sideswipe collisions injury severities.

Variable	Description	Injury severity %			Total
		Obvious	Moderate	No	
Overall Time Segmentation					
Daytime	2008–2010	15.17%	20.31%	64.52%	4278
	2011–2013	15.12%	17.87%	67.01%	4550
	2014–2016	12.30%	20.05%	67.64%	8144
Nighttime	2008–2010	16.07%	20.90%	63.03%	1282
	2011–2013	12.88%	18.06%	69.05%	1312
	2014–2016	12.31%	23.16%	64.53%	2729
Driver characteristic					
Gender	**Female**	13.63%	21.97%	64.40%	6444
	Male	13.70%	19.15%	67.14%	15,851
Age	**Young (<25)**	15.36%	19.51%	65.13%	3860
	Mid-Age (25–50)	13.31%	20.65%	66.04%	12,031
	Old (>50)	13.37%	18.96%	67.68%	6404
Restraint	**None**	34.42%	20.10%	45.48%	398
	Lap Belt	12.76%	19.87%	67.37%	15,991
	Lap and Shoulder Belt	14.77%	20.64%	64.59%	5354
	Shoulder Belt	14.86%	16.12%	69.02%	552
Behavior	**None**	15.00%	21.85%	63.16%	11,397
	Inattention and Disregarded (Sign or Signal)	10.67%	16.91%	72.42%	834
	Exceeded Speed	13.58%	22.97%	63.45%	788
	Improper or Reckless Manner (Turn, Yield, Swerved, Oversteer)	14.50%	23.58%	61.92%	1145
	Wrong Lane Use	16.42%	23.33%	60.25%	3703
	Alcohol or Drugs	28.69%	30.64%	40.67%	359
	Other (Passing, Follow Closely, Defective Equipment)	6.59%	9.73%	83.68%	4069
Vehicle Characteristic					
Vehicle Type	**Pickup**	13.02%	18.10%	68.89%	12,862
	SUV	14.59%	22.52%	62.90%	9433
Roadway Characteristic					
Road Class	**Interstate**	23.70%	22.22%	54.07%	135
	Us Route	17.70%	24.07%	58.23%	2480
	Nc Route	19.23%	21.68%	59.09%	4498
	Secondary	11.72%	17.95%	70.34%	11,667
	Local Street	10.25%	21.99%	67.76%	3365
	Other (Public or Private Road)	1.33%	10.67%	88.00%	150
Road Surface	**Dry**	14.51%	19.58%	65.91%	17,891
	Wet	12.11%	22.09%	65.80%	3088
	Road Adverse (Watery, Icy, Snowy, Sandy, Muddy, Dirty, Or Graveled)	6.08%	20.29%	73.63%	1316
Road Align	**Straight Level**	12.73%	19.96%	67.31%	11,120
	Straight Adverse (Bottom, Grade, Hillcrest)	15.90%	21.27%	62.84%	3202
	Curve Level	13.83%	19.58%	66.59%	4142
	Curve Adverse (Bottom, Grade, Hillcrest)	14.41%	19.34%	66.25%	3831
Road Configuration	**One-Way, Not Divided**	8.59%	22.70%	68.71%	163
	Two-Way, Not Divided	13.68%	19.42%	66.90%	20,367
	Two-Way, Divided	14.11%	26.06%	59.83%	1765
Road Pavement	**Concrete**	13.28%	21.77%	64.94%	271
	Smooth Asphalt	13.72%	20.59%	65.69%	13,525
	Coarse Asphalt	13.96%	19.34%	66.70%	8240
	Other (Gravel, Sand, Soil)	3.47%	5.41%	91.12%	259
Speed Limits	**0–35 Mph**	10.58%	19.15%	70.28%	4586
	36–55 Mph	14.40%	20.19%	65.41%	17,571
	56–70 Mph	25.36%	18.84%	55.80%	138
Traffic Control	**No Control**	13.86%	19.65%	66.49%	20,171
	Partial Control	13.35%	24.81%	61.84%	1161
	Full Control	10.28%	20.87%	68.85%	963
Environment characteristic					
Region	**Urban**	11.58%	22.04%	66.38%	4601
	Rural	14.23%	19.43%	66.34%	17,694
Development	**FWP (Farms, Woods, Pastures)**	14.76%	18.53%	66.71%	12,124
	Residential	13.15%	21.93%	64.92%	6151
	Commercial	11.08%	21.40%	67.52%	3873
	Other (Institutional, Industrial)	14.97%	19.05%	65.99%	147
Terrain	**Flat**	14.05%	20.14%	65.81%	4969
	Rolling	14.67%	20.97%	64.36%	13,745
	Mountainous	9.35%	15.89%	74.76%	3581
Weather	**Clear**	14.11%	19.54%	66.36%	16,086
	Cloudy	14.22%	19.71%	66.07%	3790
	Rain	11.38%	24.07%	64.55%	1687
	Other (Snow, Fog, Smog, Smoke, Sleet, Hall, Drizzle)	6.83%	21.31%	71.86%	732
Light	**Daylight**	14.20%	19.72%	66.08%	16,097
	Dusk, Down	11.44%	18.91%	69.65%	883
	Dark	12.49%	20.88%	66.62%	5315

cate log-likelihood of the respective subsets. The degree of freedom is determined by the number of significant parameters in the whole dataset. The X^2 denotes the $\chi 2$ distribution in the null hypothesis that parameters value are equal, which is 163.46 in this study. This illustrates a 99.99% confidence level to reject the null hypothesis under 88 degrees of freedom, which means that distinct fluctuations between the daytime and nighttime exist.

A further LR test is employed to verify the segmentation efficiency of the three time periods (2008–2010, 2011–2013, 2014–2016), which is presented as follows,

$$X^2 = -2[LL(\beta_{2008-2016}) - LL(\beta_{2008-2010}) - LL(\beta_{2011-2013}) - LL(\beta_{2014-2016})] \tag{6}$$

where $LL(\beta_{2008-2016})$ represents the log-likelihood with the entire periods from 2008 to 2016 in the converging model. It has the same expression to denote the log-likelihood of the other three segmented periods. The X^2 value is 343.74, also illustrates a 99.99% confidence level to reject the null hypothesis, which indicates the distinct temporal volatility across the three periods.

Finally, a series of LR test are applied to test the temporal instability of both time-of-day and three periods as follows,

$$X^2 = -2[LL(\beta_{t_2 t_1}) - LL(\beta_{t_1})] \tag{7}$$

where $LL(\beta_{t_2 t_1})$ denotes the log-likelihood in the converging model with the estimated parameters from t_2, using the data in t_1. $LL(\beta_{t_1})$ denotes the log-likelihood in the converging model with data in t_1. The reversed test is also conducted by substituting t_1 or t_2 for each other. The X^2 value is $\chi 2$ distributed (with a degree of freedom equals to the number of parameters in t_2) and applied to accept or reject the null hypothesis that the parameters of t_1 and t_2 are equal. The results of the temporal instability tests for all period pairs are presented in Table 3 (X^2value with degrees of freedom in parenthesis and confidence level in brackets). Only 2 of these 30 runs show that confidence level is relatively small, but neither of the two reversed X^2 value accepts the null hypothesis concurrently, which indicates a comparatively high confidence to verify the temporal instability. Although one of the confidence levels does not reach 95% (e.g., the 2008–2010 daytime vs 2011–2013 nighttime, which gives a 94.38%), it is still high enough to reject the null hypothesis. Therefore, it is necessary to segment the whole dataset based on the time of day and different time periods to explore the potential factors that impact the injury severity of reverse sideswipe collisions.

6. Model results and discussions

The results for the six period combinations are presented in Tables 4–9 (with Obvious injury [OI]; Moderate injury [MI]; No injury [NI] as the base) by three types of RPL models. Notably, RPL models with heterogeneous means and heteroscedastic variances are obtained in the 2011–2013 daytime, 2014–1016 daytime, and 2014–2016 nighttime. Meanwhile, heterogeneity in means is found in the 2008–2010 daytime. However, there is no random parameter statistically significant in the 2008–2010 nighttime and 2011–2013 nighttime, and therefore, the two models are collapsed into the multinomial logit model. Table 10 displays the marginal effects of each model, which is utilized to explain the potential factors that affect the injury severity of reverse sideswipe collisions. The fluctuations between daytime and nighttime and volatility across different periods can also be observed in Table 10. Detailed results of estimated parameters are discussed in Tables 4–9.

6.1. Driver characteristics

The male driver has been identified as a significant variable in 2008–2010 daytime and 2014–2016 nighttime. From marginal effects in Table 10, both the two periods demonstrate that the male driver slightly increases the probability of obvious injury (0.0045 and 0.0025, respectively) compared to the female. In contrast, the older driver whose age is over 50 slightly decreases the possibility of obvious injury by 0.0045, which indicates that older drivers are more prudent and experienced than aggressive young drivers. The effects of gender and age on the sideswipe injury severity can also be found in the same direction crashes when making lane changes (Bates, Davey, Watson, King, & Armstrong, 2014).

The restraint with belts is statistically significant in all period combinations excluding 2014–2016 daytime. A typical fluctuation between daytime and nighttime and temporal volatility is revealed, and various belt protections significantly reduce the likelihood of obvious injury to varying degrees except in 2014–2016 nighttime. As for the model structure in Table 9 during that time, the lap belt is a random parameter sensitive to the heterogeneity in means and variances, and the residential factor lessens its mean and the male driver expands its variances in obvious injury. Variations in social environment changes are possible reasons for the temporal volatility (Mannering, 2018). Besides, the time-of-day difference is observed in 2008–2010, and the lap and shoulder belt manifest a higher probability of mitigating the obvious injury in nighttime (0.1768) than in daytime (0.0590). This variation also exists in the shoulder-only belt during 2011–2013 with 0.1179 in nighttime and 0.0014 in the daytime.

Regarding the primary behaviors that contribute to the collisions, inattention and ignoring signs or signals could modestly increase the probability of obvious injury by 0.0003 in 2011–2013 daytime and 0.0005 in 2014–2016 nighttime. Exceeding speed limits and improper or reckless manner such as wrong turning, failing to yield or swerve, and oversteering intensify the possibility of obvious injury by 0.1744 and 0.1066, respectively, in 2008–2010 nighttime. Using the wrong lane could also result in the increment of moderate injury in 2008–2010 nighttime by 0.0709 and obvious injury in 2014–2016 nighttime by 0.0026. Remarkable variations in period pairs can be observed in the dri-

Table 3
Transferability test results between period pairs.

t_1	t_2 2008–2010 daytime	2008–2010 nighttime	2011–2013 daytime	2011–2013 nighttime	2014–2016 daytime	2014–2016 nighttime
2008–2010 daytime (26)	-	113.20 [99.99%]	88.71 [99.99%]	139.85 [99.99%]	111.16 [99.99%]	96.17 [99.99%]
2008–2010 nighttime (15)	59.42 [99.98%]	-	49.69 [99.30%]	60.17 [99.99%]	40.11 [96.19%]	42.57 [98.44%]
2011–2013 daytime (28)	119.19 [99.99%]	178.17 [99.99%]	-	129.44 [99.99%]	97.68 [99.99%]	128.67 [99.99%]
2011–2013 nighttime (17)	38.35 [94.38%]	59.92 [99.99%]	20.37 [14.95%]	-	20.72 [24.39%]	39.67 [96.85%]
2014–2016 daytime (26)	134.90 [99.99%]	272.95 [99.99%]	82.73 [99.99%]	198.52 [99.99%]	-	208.86 [99.99%]
2014–2016 nighttime (25)	82.21 [99.99%]	156.46 [99.99%]	59.72 [99.99%]	129.14 [99.99%]	89.53 [99.99%]	-

Table 4
Estimated coefficients of significant variable in 2008–2010 daytime.

Variable	Description	Coefficient	z-value
Intercept	Constant [OI]	0.166	−2.73
	Constant [MI]	−0.698	−7.80
Driver Characteristic			
Gender	Male [MI]	−0.273	−3.18
Restraint	Lap and Shoulder Belt [OI]	−0.688	−2.73
Behavior	Other (Passing, Follow Closely, Defective Equipment) [OI]	−1.402	−6.25
	Other (Passing, Follow Closely, Defective Equipment) [MI]	−0.897	−7.02
	Alcohol or Drugs [MI]	1.111	3.07
Vehicle Characteristic			
Vehicle type	SUV [OI]	0.336	2.86
Roadway Characteristic			
Road class	NC Route [OI]	0.597	3.88
	Us Route [MI]	0.403	3.41
	NC Route [MI]	0.312	3.14
Road Surface	Road Adverse (Watery, Icy, Snowy, Sandy, Muddy, Dirty, or Graveled) [OI]	−1.371	−4.41
Road Align	Curve Adverse (Bottom, Grade, Hillcrest) [OI]	0.313	2.02
Road Pavement	Smooth Asphalt [OI]	−1.735	−2.76
	Standard deviation	2.008	2.85
	Coarse Asphalt [OI]	−0.954	−2.76
	Coarse Asphalt [MI]	−0.219	−2.55
	Other (Gravel, Sand, Soil) [OI]	−1.610	−2.54
	Other (Gravel, Sand, Soil) [MI]	−2.060	−2.84
Environment characteristic			
Development	Commercial [OI]	−3.451	−1.42
	Standard deviation	3.866	1.67
	Commercial [MI]	−0.341	−3.13
Terrain	Mountainous [OI]	−0.595	−3.36
	Mountainous [MI]	−0.461	−4.37
Light	Dusk, Down [MI]	−0.957	−3.08
Heterogeneity in means of random parameters			
Commercial [OI]: Curve Adverse (Bottom, Grade, Hillcrest)		2.819	1.80

Model statistics: number of observations: 4278, Log-likelihood constant only: −3818.570, Log-likelihood at convergence: −3653.911, Akaike information criterion (AIC): 7359.800.

Table 5
Estimated coefficients of significant variable in 2011–2013 daytime.

Variable	Description	Coefficient	z-value
Intercept	Constant [OI]	0.589	0.21
	Constant [MI]	−0.715	−4.71
Driver Characteristic			
Restraint	Lap Belt [OI]	−1.346	−4.78
	Shoulder Belt [OI]	−1.797	−3.82
	Shoulder Belt [MI]	−0.839	−2.19
Behavior	Inattention and Disregarded (Sign or Signal) [OI]	−1.054	−3.24
	Other (Passing, Follow Closely, Defective Equipment) [OI]	−0.977	−6.98
	Inattention and Disregarded (Sign or Signal) [MI]	−0.833	−3.5
	Other (Passing, Follow Closely, Defective Equipment) [MI]	−1.357	−8.14
Vehicle Characteristic			
Vehicle type	SUV [MI]	−0.214	−0.55
	Standard deviation	1.248	1.66
Roadway Characteristic			
Road Class	Us Route [OI]	0.623	4.47
	NC Route [OI]	0.599	5.47
	NC Route [MI]		
Road Surface	Wet [OI]	−0.438	−2.94
	Road Adverse (Watery, Icy, Snowy, Sandy, Muddy, Dirty, or Graveled) [OI]	−1.308	−3.92
	Road Adverse (Watery, Icy, Snowy, Sandy, Muddy, Dirty, or Graveled) [MI]	−0.563	−2.27
Road Configuration	Two-Way, Not Divided [MI]	−0.588	−3.96
Environment characteristic			
Development	Residential [MI]	0.242	2.47
	Commercial [OI]	−2.385	−1.53
	Standard deviation	2.824	1.86
Terrain	Mountainous [OI]	−0.675	−4.88
	Mountainous [MI]	−0.697	−4.19
Weather	Cloudy [MI]	0.258	2.26
Light	Dark [MI]	−0.359	−1.85
Heterogeneity in means of random parameters			
Commercial [OI]: Road Adverse (Watery, Icy, Snowy, Sandy, Muddy, Dirty, Or Graveled)		2.421	2.15
Commercial [OI]: Cloudy		1.095	1.86
Heteroscedasticity of random parameters			
SUV [MI]: Mountainous		0.752	1.79

Model statistics: number of observations:4550, Log-likelihood constant only: −3920.362, Log-likelihood at convergence: −3734.886, Akaike information criterion (AIC): 7525.800.

Table 6
Estimated coefficients of significant variable in 2014–2016 daytime.

Variable	Description	Coefficient	z-value
Intercept	Constant [OI]	−3.090	−7.24
	Constant [MI]	−0.804	−5.95
Driver characteristic			
Age	Old (>50) [OI]	−0.155	−1.97
Behavior	Inattention and Disregarded (Sign or Signal) [MI]	−1.029	−3.84
	Alcohol or Drugs [OI]	0.970	3.69
	Other (Passing, Follow Closely, Defective Equipment) [OI]	−1.184	−10.01
	Other (Passing, Follow Closely, Defective Equipment) [MI]	−1.604	−12.09
Vehicle Characteristic			
Vehicle Type	SUV [MI]	−1.794	−2.64
	Standard deviation	4.118	4.00
Roadway Characteristic			
Road Class	Us Route [OI]	0.605	5.63
	NC Route [OI]	0.541	6.59
Road Surface	Road Adverse (Watery, Icy, Snowy, Sandy, Muddy, Dirty, Or Graveled) [OI]	−1.143	−5.31
Road Align	Straight Adverse (Bottom, Grade, Hillcrest) [OI]	0.351	3.58
	Curve Adverse (Bottom, Grade, Hillcrest) [OI]	0.245	2.54
Road Configuration	Two-Way, Not Divided [MI]	−0.537	−4.07
Road Pavement	Smooth Asphalt [OI]	1.193	2.82
	Coarse Asphalt [OI]	1.210	2.86
Speed Limits	36–55 Mph [OI]	0.368	3.85
	36–55 Mph [MI]	0.367	3.51
Environment characteristic			
Terrain	Mountainous [OI]	−0.890	−7.34
	Mountainous [MI]	−0.490	−0.60
	Standard deviation	3.662	3.31
Weather	Rain [MI]	0.495	3.63
Heterogeneity in means of random parameters			
Mountainous [MI]: Curve Adverse (Bottom, Grade, Hillcrest)		−1.19295	−2.45
Mountainous [MI]: Two-Way, Not Divided		−2.3572	−2.23
Heteroscedasticity of random parameters			
Mountainous [MI]: Coarse Asphalt		0.35022	0.13585

Model statistics: number of observations: 8144, Log-likelihood constant only: −6876.940, Log-likelihood at convergence: −6519.633, Akaike information criterion (AIC): 13091.300.

Table 7
Estimated coefficients of significant variable in 2008–2010 nighttime.

Variable	Description	Coefficient	z-value
Intercept	Constant [OI]	0.478	1.24
	Constant [MI]	−1.055	−11.63
Driver characteristic			
Restraint	Lap and Shoulder Belt [OI]	−1.417	−3.82
	Shoulder Belt [OI]	−1.692	−3.33
Behavior	Exceeded Speed [OI]	1.397	3.73
	Improper or Reckless Manner (Turn, Yield, Swerved, Oversteer) [OI]	0.854	2.8
	Wrong Lane Use [MI]	0.440	2.54
	Alcohol or Drugs [OI]	1.447	4.12
	Alcohol or Drugs [MI]	0.867	2.44
	Other (Passing, Follow Closely, Defective Equipment) [OI]	−0.756	−2.85
	Other (Passing, Follow Closely, Defective Equipment) [MI]	−0.915	−3.67
Roadway Characteristic			
Road Class	Local Street [OI]	−0.774	−2.83
Road Surface	Road Adverse (Watery, Icy, Snowy, Sandy, Muddy, Dirty, or Graveled) [OI]	−0.731	−2.03
Environment Characteristic			
Terrain	Mountainous [MI]	−0.566	−2.17
Light	Dark [OI]	−0.595	−3.47

Model statistics: number of observations: 1282, Log-likelihood constant only: −1169.085, Log-likelihood at convergence: −1110.061, Akaike information criterion (AIC): 2250.100.

vers who are alcohol or drugs impaired. The magnitude of severity shifts from moderate injury in the daytime to obvious injury in nighttime during 2008–2010, and the severity level is reversed between daytime and nighttime during 2014–2016. Although actions like improper passing, following closely, and so forth, only cause property damage including unsightly scrapes and dented doors in sideswipe collisions, psychological impairment and mental anguish resulting from the crash still cannot be neglected.

6.2. Vehicle characteristics

Concerning the vehicle types, this study mainly examines the light trucks involving the pickup and sport utility vehicle owing to higher proportions in reverse sideswipe collisions. Apparent temporal instability is demonstrated, SUVs at fault are more likely to result in obvious injury in 2008–2010 daytime (0.013) and 2011–2013 nighttime (0.0477), but contribute to more moderate

Table 8
Estimated coefficients of significant variable in 2011–2013 nighttime.

Variable	Description	Coefficient	z-value
Intercept	Constant [OI]	−0.875	−2.42
	Constant [MI]	−1.516	−14.56
Driver characteristic			
Restraint	Lap Belt [OI]	−1.400	−3.99
	Shoulder Belt [OI]	−1.121	−2.09
Behavior	Alcohol or Drugs [MI]	0.698	2.26
	Other (Passing, Follow Closely, Defective Equipment) [OI]	−1.056	−3.56
	Other (Passing, Follow Closely, Defective Equipment) [MI]	−0.676	−3.11
Vehicle Characteristic			
Vehicle Type	SUV [OI]	0.453	2.65
Roadway Characteristic			
Road Class	Us Route [OI]	0.896	3.40
	Us Route [MI]	0.471	2.11
	NC Route [OI]	0.692	3.46
Road Align	Straight Adverse (Bottom, Grade, Hillcrest) [OI]	0.869	3.92
	Straight Adverse (Bottom, Grade, Hillcrest) [MI]	0.557	2.78
	Curve Adverse (Bottom, Grade, Hillcrest) [OI]	0.790	2.99
	Curve Adverse (Bottom, Grade, Hillcrest) [MI]	0.463	2.13
Road Configuration	Two-Way, Divided [MI]	0.529	2.30
Terrain	Mountainous [OI]	−0.940	−2.64

Model statistics: number of observations: 1312, Log-likelihood constant only: −1087.379, Log-likelihood at convergence: −1031.438, Akaike information criterion (AIC): 2096.900.

Table 9
Estimated coefficients of significant variable in 2014–2016 nighttime.

Variable	Description	Coefficient	z-value
Intercept	Constant [OI]	−1.628	−3.53
	Constant [MI]	−1.692	−7.99
Driver characteristic			
Gender	Male [MI]	−0.227	−2.08
Restraint	Lap Belt [OI]	−2.715	−2.48
	Standard deviation	2.800	2.40
Behavior	Wrong Lane Use [OI]	0.892	2.70
	Wrong Lane Use [MI]	0.356	2.50
	Alcohol or Drugs [MI]	0.582	2.39
	Other (Passing, Follow Closely, Defective Equipment) [OI]	−1.875	−3.03
	Other (Passing, Follow Closely, Defective Equipment) [MI]	−0.807	−5.33
Vehicle Characteristic			
Vehicle Type	SUV [MI]	0.342	3.28
Roadway Characteristic			
Road Class	NC Route [OI]	1.270	3.47
	Secondary [MI]	−0.249	−2.45
Road Surface	Wet [OI]	−1.209	−2.57
	Road Adverse (Watery, Icy, Snowy, Sandy, Muddy, Dirty, Or Graveled) [OI]	−3.246	−2.35
Road Align	Curve Level [MI]	−0.366	−2.43
Speed Limits	36–55 Mph [MI]	0.287	2.35
Development	Residential [MI]	0.240	2.22
	Commercial [OI]	−1.703	−3.39
Weather	Cloudy [OI]	0.725	1.96
Light	Dusk, Down [MI]	0.667	2.83
	Dark [OI]	0.843	2.36
	Dark [MI]	0.654	4.06
Heterogeneity in means of random parameters			
Lap Belt [OI]: Residential		−0.664	−1.83
Heteroscedasticity of random parameters			
Lap Belt [OI]: Male		0.180	1.92

Model statistics: number of observations: 2729, Log-likelihood constant only: −2399.681, Log-likelihood at convergence: −2256.915, Akaike information criterion (AIC): 4563.800.

injury from 2014 to 2016 with 0.0242 in the daytime and 0.0312 in the nighttime. In comparison with pickups, SUVs lead to more serious injury. This is in accord with Adanu et al. (2021) in that risks of crashes could diminish when SUVs are not involved, especially for young drivers. One possible reason for this is that pickup has a more robust structure to absorb more energy in a reverse sideswipe collision.

From Table 5, the SUV factor is a random parameter with a mean of −0.214 and a standard deviation of 1.248 in 2011–2013 daytime. This implies that this factor reduces the probability of moderate injury for 43.19% of reverse sideswipe collisions and enhances the possibility of moderate injury for 56.81% of collisions. It is noted that the mountainous factor shows heteroscedasticity in

Table 10
Marginal effects of factors for reverse sideswipe collisions during time-of-day and three periods.

Period	2008–2010 daytime			2011–2013 daytime			2014–2016 daytime			2008–2010 nighttime			2011–2013 nighttime			2014–2016 nighttime		
Variable	OI	MI	NI	OI	MI	NI	OI	MI	NI	OI	MI	NI	OI	MI	NI	OI	MI	NI
Driver Characteristics																		
Male	0.0045	−0.0293	0.0249													0.0025	−0.0250	0.0225
Old (>50)							−0.0045	0.0008	0.0037									
Lap Belt				−0.1472	0.031	0.1162							−0.1473	0.0341	0.1132	0.0117	−0.0015	−0.0103
Lap and Shoulder Belt	−0.0590	0.0155	0.0436							−0.1768	0.0462	0.1306						
Shoulder Belt				−0.0014	−0.0005	0.0019				−0.2112	0.0552	0.1560	−0.1179	0.0273	0.0906			
Inattention and Disregarded (Sign or Signal)				0.0003	−0.0037	0.0034	0.0005	−0.0025	0.0021									
Exceeded Speed										0.1744	−0.0456	−0.1288						
Improper or Reckless Manner (Turn, Yield, Swerved, Oversteer)										0.1066	−0.0279	−0.0787						
Wrong Lane Use										−0.0144	0.0709	−0.0565				0.0041	0.0026	−0.0067
Alcohol or Drugs	−0.0005	0.0022	−0.0016				0.0018	−0.0004	−0.0014	0.0762	0.0463	−0.1225	−0.0170	0.1008	−0.0838	−0.0006	0.0040	−0.0034
Other (Passing, Follow Closely, Defective Equipment)	−0.0055	−0.0067	0.0122	−0.0059	−0.0069	0.0128	−0.0054	−0.0085	0.0138	−0.0323	−0.0614	0.0937	−0.0474	−0.0360	0.0833	−0.0047	−0.0072	0.0119
Vehicle Characteristics																		
SUV	0.0130	−0.0036	−0.0094	−0.0047	0.0266	−0.0219	−0.0039	0.0242	−0.0203				0.0477	−0.0110	−0.0367	−0.0030	0.0312	−0.0282
Roadway Characteristics																		
Us Route	−0.0011	0.0087	−0.0076	0.0095	−0.0019	−0.0076	0.0086	−0.0015	−0.0071				0.0414	0.02315	−0.06455			
NC Route	0.0058	0.0033	−0.0091	0.0073	0.0027	−0.0100	0.0152	−0.0029	−0.0123				0.0390	0.0297	−0.0686	0.0204	−0.0064	−0.0140
Secondary Route																0.0016	−0.0185	0.0169
Local Street										−0.0966	0.0253	0.0714						
Wet Surface				−0.0053	0.0012	0.0041										−0.0072	0.0021	0.0051
Road Adverse (Watery, Icy, Snowy, Sandy, Muddy, Dirty, or Graveled)	−0.0056	0.0014	0.0042	−0.0015	−0.0010	0.0025	−0.0033	0.0006	0.0027	−0.0912	0.0238	0.0674				−0.0021	0.0006	0.0015
Straight Adverse (Bottom, Grade, Hillcrest)							0.0060	−0.0011	−0.0049									
Curve Level																0.0007	−0.0073	0.0067
Curve Adverse (Bottom, Grade, Hillcrest)	0.0063	−0.0015	−0.0048				0.0045	−0.0007	−0.0037				0.0359	0.0239	−0.0598			
Two-Way, Not Divided				0.0128	−0.0653	0.0525	0.0087	−0.0488	0.0401									
Two-Way, Divided													−0.0129	0.0764	−0.0636			
Smooth Asphalt Pavement	−0.0151	0.0049	0.0103				0.0745	−0.0131	−0.0614									
Coarse Asphalt Pavement	−0.0170	−0.0012	0.0182				0.0471	−0.0084	−0.0387									
Other (Gravel, Sand, Soil)	−0.0007	−0.0004	0.0011															
36–55 Mph							0.0127	0.0121	−0.0248							−0.0038	0.0377	−0.0339
Environment Characteristics																		
Residential				−0.0019	0.0092	−0.0073										−0.0011	0.0121	−0.0110
Commercial	0.0026	−0.0047	0.0022	0.0068	−0.0016	−0.0052										−0.0101	0.0029	0.0071
Mountainous Terrain	−0.0037	−0.0053	0.0091	−0.0042	−0.0049	0.0091	−0.0046	0.0015	0.0032	0.0185	−0.0912	0.0727	−0.0989	0.0229	0.0760			
Cloudy				−0.0012	0.0064	−0.0052										0.0058	−0.0017	−0.0042
Rainy							−0.0008	0.0045	−0.0037									
Dusk, Down light	0.0003	−0.0023	0.0020													−0.0006	0.0083	−0.0077
Dark light				0.0005	−0.0026	0.0021				−0.0742	0.0194	0.0548				0.0149	0.0392	−0.0541

random parameters, indicating a subtle growth (0.752) in the variance that makes moderate injury involving SUVs more likely.

6.3. Roadway characteristics

Compared to the interstate highway, the U.S. and North Carolina routes have similar effects on exacerbating the obvious injury. This distinguishes it from Adanu et al. (2021), who primarily considered sideswipe crashes in the same direction involving commercial vehicles. Collisions that occurred in nighttime (with 0.414 for U.S. route and 0.039 for NC route) have a higher likelihood than in daytime (with 0.0095 for U.S. route and 0.073 for NC route) in 2011–2013, and the same situation can also be found in 2014–2016 for NC route. However, obvious injuries are less likely as the road class downgrades. There is a slight increase in secondary route with merely 0.0016 compared to NC route with 0.0204 in 2014–2016 nighttime, and local street decline the probability of obvious injury by 0.0966.

Compared with the dry road surface, wet surface lessens the drivers' risk of suffering from the obvious injury (with 0.0053 in 2011–2013 daytime and 0.0072 in 2014–2016 nighttime). Actually, hydroplaning on adverse roads such as snowy, icy, sandy, muddy, or graveled might not result in a higher likelihood of more serious injuries. One possible explanation for this counterintuitive outcome is that drivers are usually more cautious under slippery conditions.

For the road alignment that reflects the horizontal curvature, both hillcrest, grade or bottom design, and curve roadway indicate a modest increase of the obvious injury by 0.006 in 2014–2016 daytime and by 0.0007 in 2014–2016 nighttime, respectively. It is even worse when drivers encounter the curve and adverse roadway simultaneously. Meanwhile, two-way without divided configuration is found to be positively associated with obvious injury. However, the probability could decline by 0.0129 when installing the separation facility (e.g., buffer median barriers), which could minimize the impact of opposite-direction crashes (Chitturi, Ooms, Bill, & Noyce, 2011).

In terms of pavement that is classified based on the texture of materials, smooth asphalt is found to be a random parameter with a mean of −1.735 and standard deviation of 2.008 in Table 4. It implies that this factor decreases the probability of obvious injury for 19.38% of collisions and increases the possibility of obvious injury for 56.81% of collisions. Nevertheless, it is insensitive to the possible heterogeneous means and heteroscedastic variances. The reverse effects of smooth and coarse asphalt can be observed between 2008–2010 daytime and 2014–2016 daytime, which reveals the necessity of segmentation across varied periods.

Speed limit with 36–55 mph is statistically significant only in 2014–2016. It is more likely to result in the obvious injury by 0.0127 in the daytime, and a slight growth in moderate injury from 0.0121 in the daytime to 0.0377 in nighttime is identified.

6.4. Environment characteristics

Compared to the land use developed for FWP (farms, woods, and pastures), the residential area slightly decreases the probability of obvious injury by 0.0019 in 2011–2013 daytime and by 0.0011 in 2014–2016 nighttime. However, the commercial area is more likely to result in the obvious injury in the daytime of 2008–2010 and 2011–2013. Moreover, in Tables 4 and 5, the commercial area is a random parameter in that two periods (with a mean of −3.451 and −2.385, a standard deviation of 3.866 and 2.824, respectively), which implies that obvious injury is less likely for 18.60% and 19.92% of the collisions and more likely for 81.40% and 80.08% of crashes, correspondingly. Meanwhile, curve adverse align, adverse road surface, and cloudy are found to produce heterogeneous means of commercial area, increasing its mean that make the crash injury more severe.

For the terrain that reflects the vertical undulation, mountainous terrain has the potential to be risky since steep hillsides make novice drivers harder to maintain the vehicle. However, almost all models except nighttime of 2008–2010 and 2014–2016 present a negative association with obvious injury. One explainable reason for that is that due to the lower speed limits, drivers are more attentive in those graded and hilly roadways.

The weather condition with rain is statistically significant only in 2014–2016 daytime, which has a higher likelihood of sustaining moderate injury with 0.0045. Additionally, distinct temporal variations are observed in the effects of cloudy, dusk, and dark lighting conditions. The cloudy weather modestly increases the probability of moderate injury by 0.0064 and results in relatively less obvious injury by 0.0012 in 2011–2013 daytime; but makes obvious injury more likely by 0.0058 and leads to less moderate injury by 0.0017 in 2014–2016 nighttime. This reverse effect can also be recognized for the dusk light between the 2008–2010 daytime and 2014–2016 nighttime, which again supports the segmentation across different time periods.

7. Conclusions and implications

This study investigates the time-of-day fluctuations and temporal volatility of potential factors that impact the injury severity of reverse sideswipe collisions between daytime and nighttime across three time periods (a segmentation every-three years from 2008 to 2016). With data from the HSIS in North Carolina, an extension of random parameters logit (RPL) models considering the heterogeneous means and heteroscedastic variances are developed and employed to estimate the significant parameters. The transferability tests verify the temporal instability during the time of day and three different time ranges. The variations in the combinations of them have also been explored in the model results. In terms of model structures, the method further reveals the unobserved heterogeneity among the significant variables and makes better fitting performance compared to the conventional RPL model.

A wide variety of contributing factors are identified to have positive associations with the obvious injury, including the male driver, improper or reckless manner (e.g. exceeding speed limit), curve and adverse roadway, and two-way not divided configuration. Nevertheless, factors such as the older driver, residential area, wet road surface, and mountainous terrain could mitigate the injury severity. Meanwhile, significant temporal volatility of the marginal effects over time is observed for belts with various degrees of protection, intoxicated drivers, SUVs at fault, and slippery road surface. Evidence of time-of-day fluctuations is also detected in this study. For instance, restraint with belts is more effective in reducing obvious injury in the nighttime, but it is more likely to result in more severe injury on the higher road class compared to the daytime. Additionally, the reverse effects of asphalt pavement, cloudy weather, and dim lighting conditions can also demonstrate the necessity for the segmentation of the dataset.

Countermeasure for alleviating the injury severity of reverse sideswipe collisions contains the installation of centerline rumble strips. This facility could deliver tactile and acoustic signals when vehicles transgress the centerline with an undulating surface, which can reduce by 44% the fatal or incapacitating injury caused by sideswipe-opposite-direction collisions in rural area (Torbic et al., 2010). Currently, intelligent vehicles equipped with lane departure warning and forward collision avoidance technology could also effectively lessen the risk of conflicts (Adanu et al., 2021).

The findings of the study may provide instructive solutions for engineers and policymakers to enhance sideswipe-related safeguards, establish the safety evaluation system, and clarify the accident liability. Given that this study mainly examines the temporal features of sideswipe injury severity, future work could take the spatiotemporal patterns and hotspot distribution of sideswipe into account. Besides, the socio-economic reasons behind traffic accidents deserve more exploration.

Declaration of 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

The authors want to express their deepest gratitude for the financial support by the United States Department of Transportation, University Transportation Center through the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE) at The University of North Carolina at Charlotte (Grant Number: 69A3551747133).

References

Abdel-Aty, M., Keller, J., & Brady, P. A. (2005). Analysis of types of crashes at signalized intersections by using complete crash data and tree-based regression. *Transportation Research Record, 1908*(1), 37–45.

Adanu, E. K., Lidbe, A., Tedla, E., & Jones, S. (2021a). Factors associated with driver injury severity of lane changing crashes involving younger and older drivers. *Accident Analysis & Prevention, 149*, 105867.

Adanu, E. K., Lidbe, A., Tedla, E., & Jones, S. (2021b). Injury-severity analysis of lane change crashes involving commercial motor vehicles on interstate highways. *Journal of Safety Research, 76*, 30–35.

Administrations, F. H. (2019). KABCO Injury Classification Scale and Definitions. Online at: https://safety. fhwa. dot. gov/hsip/spm/conversion_tbl/pdfs/kabco_ctable_by_state.pdf. Accessed on November, 30.

Al-Bdairi, N. S. S., Behnood, A., & Hernandez, S. (2020). Temporal stability of driver injury severities in animal-vehicle collisions: A random parameters with heterogeneity in means (and variances) approach. *Analytic Methods in Accident Research, 26*, 100120.

Bates, L. J., Davey, J., Watson, B., King, M. J., & Armstrong, K. (2014). Factors contributing to crashes among young drivers. *Sultan Qaboos University Medical Journal, 14*(3), e297.

Behnood, A., & Mannering, F. (2017). Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. *Analytic Methods in Accident Research, 16*, 35–47.

Behnood, A., & Mannering, F. L. (2016). An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. *Analytic Methods in Accident Research, 12*(1–17), 1.

Bham, G. H., Javvadi, B. S., & Manepalli, U. R. (2012). Multinomial logistic regression model for single-vehicle and multivehicle collisions on urban US highways in Arkansas. *Journal of Transportation Engineering, 138*(6), 786–797.

Chen, Z., Qin, X., & Shaon, M. R. R. (2018). Modeling lane-change-related crashes with lane-specific real-time traffic and weather data. *Journal of Intelligent Transportation Systems, 22*(4), 291–300.

Chitturi, M. V., Ooms, A. W., Bill, A. R., & Noyce, D. A. (2011). Injury outcomes and costs for cross-median and median barrier crashes. *Journal of Safety Research, 42*(2), 87–92.

Fitch, G. M., Lee, S. E., Klauer, S., Hankey, J., Sudweeks, J., & Dingus, T. (2009). *Analysis of lane-change crashes and near-crashes*. US Department of Transportation, National Highway Traffic Safety Administration.

Geedipally, S. R., Patil, S., & Lord, D. (2010). Examination of methods to estimate crash counts by collision type. *Transportation Research Record, 2165*(1), 12–20.

Islam, M., & Hernandez, S. (2013). Large truck–involved crashes: Exploratory injury severity analysis. *Journal of Transportation Engineering, 139*(6), 596–604.

Islam, M., Alnawmasi, N., & Mannering, F. (2020). Unobserved heterogeneity and temporal instability in the analysis of work-zone crash-injury severities. *Analytic Methods in Accident Research, 28*, 100130.

Khattak, A. J., Kantor, P., & Council, F. M. (1998). Role of adverse weather in key crash types on limited-access: Roadways implications for advanced weather systems. *Transportation Research Record, 1621*(1), 10–19.

Kim, D. G., Lee, Y., Washington, S., & Choi, K. (2007). Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis & Prevention, 39*(1), 125–134.

Kim, D. G., Washington, S., & Oh, J. (2006). Modeling crash types: New insights into the effects of covariates on crashes at rural intersections. *Journal of Transportation Engineering, 132*(4), 282–292.

Kusano, K. D., & Gabler, H. C. (2013). Characterization of opposite-direction road departure crashes in the united states. *Transportation Research Record, 2377*(1), 14–20.

Lee, C., Abdel-Aty, M., & Hsia, L. (2006). Potential real-time indicators of sideswipe crashes on freeways. *Transportation Research Record, 1953*(1), 41–49.

Li, Z., Wang, W., Chen, R., & Liu, P. (2014). Conditional inference tree-based analysis of hazardous traffic conditions for rear-end and sideswipe collisions with implications for control strategies on freeways. *IET Intelligent Transport Systems, 8*(6), 509–518.

Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research, 11*, 1–16.

Mannering, F. (2018). Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research, 17*, 1–13.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics, 15*(5), 447–470.

Mokhtarimousavi, S. (2019). A time of day analysis of pedestrian-involved crashes in California: Investigation of injury severity, a logistic regression and machine learning approach using HSIS data. *Institute of Transportation Engineers. ITE Journal, 89*(10), 25–33.

Moore, D. N., Schneider, W. H., IV, Savolainen, P. T., & Farzaneh, M. (2011). Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. *Accident Analysis & Prevention, 43*(3), 621–630.

National Center for Statistics and Analysis. Traffic Safety Facts 2018: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812981.

Pande, A., & Abdel-Aty, M. (2006). Assessment of freeway traffic parameters leading to lane-change related collisions. *Accident Analysis & Prevention, 38*(5), 936–948.

Park, S., & Ritchie, S. G. (2004, January). Exploring the relationship between freeway speed variance, lane changing, and vehicle heterogeneity. In *83rd Annual Meeting of Transportation Research Board*.

Qu, X., Wang, W., Wang, W., & Liu, P. (2013). Real-time freeway sideswipe crash prediction by support vector machine. *IET Intelligent Transport Systems, 7*(4), 445–453.

Shawky, M. (2020). Factors affecting lane change crashes. *IATSS Research, 44*(2), 155–161.

Silverstein, C., Schorr, J., & Hamdar, S. H. (2016). Work zones versus nonwork zones: Risk factors leading to rear-end and sideswipe collisions. *Journal of Transportation Safety & Security, 8*(4), 310–326.

Song, L., Fan, W. D., & Li, Y. (2021). Time-of-day variations and the temporal instability of multi-vehicle crash injury severities under the influence of alcohol or drugs after the Great Recession. *Analytic Methods in Accident Research, 32*(670), 100183.

Torbic, D. J., Hutton, J. M., Bokenkroger, C. D., Bauer, K. M., Donnell, E. T., Lyon, C., & Persaud, B. (2010). Guidance on design and application of rumble strips. *Transportation Research Record, 2149*(1), 59–69.

Wang, J., Zhang, Q., Wang, Y., Weng, J., & Yan, X. (2016). Analysis of sideswipe collision precursors considering the spatial-temporal characteristics of freeway traffic. *Journal of Transportation Engineering, 142*(12), 04016064.

Washington, S., Karlaftis, M., & Mannering, F. (2011). *Statistical and Econometric Methods for Transportation Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC.

Ye, X., Pendyala, R. M., Washington, S. P., Konduri, K., & Oh, J. (2009). A simultaneous equations model of crash frequency by collision type for rural intersections. *Safety Science, 47*(3), 443–452.

Zou, W., Wang, X., & Zhang, D. (2017). Truck crash severity in New York city: An investigation of the spatial and the time of day effects. *Accident Analysis & Prevention, 99*, 249–261.

Chengying Hua is a Ph.D. student in the Infrastructure and Environmental Systems (INES) program at the University of North Carolina at Charlotte, Charlotte, NC, U.S.A. He is with the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE). His research topics include transportation safety data analysis and connected and autonomous vehicles.

Wei (David) Fan currently serves as a full professor in the Department of Civil and Environmental Engineering at The University of North Carolina at Charlotte (UNC Charlotte). He is the Director of the USDOT Tier I University Transportation Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE). Dr. Fan holds a Ph.D. (May 2004) in Civil Engineering - Transportation from the University of Texas at Austin (Hook 'em Horns!). His primary research interests include operations research, transportation network modeling & optimization, multimodal transportation systems planning and operations, traffic operations and simulation, transportation safety and crash analysis, intelligent transportation systems (ITS), and computer software development.

Journal of Safety Research 84 (2023) 108–116

Contents lists available at ScienceDirect

Journal of Safety Research

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

Measuring base-rate bias error in workplace safety investigators

Carla L. MacLean [a,*], Itiel E. Dror [b]

[a] Kwantlen Polytechnic University, Department of Psychology, 12666 72 Avenue, Surrey, B.C, Canada
[b] University College London, London, United Kingdom

ARTICLE INFO

Article history:
Received 19 April 2022
Received in revised form 25 July 2022
Accepted 18 October 2022
Available online 28 October 2022

Keywords:
Cognitive bias
Workplace Investigation
Human error bias
Context effects
Expert decision making

ABSTRACT

Introduction: This study explored the magnitude of professional industrial investigators' bias to attribute cause to a person more readily than to situational factors (i.e., human error bias). Such biased opinions may relieve companies from responsibilities and liability, as well as compromise efficacy of suggested preventative measures. *Method:* Professional investigators and undergraduate participants were given a summary of a workplace event and asked to allocate cause to the factors they found causal for the event. The summary was crafted to be objectively balanced in its implication of cause equally between two factors: a worker and a tire. Participants then rated their confidence and the objectivity of their judgment. We then conducted an effect size analysis, which supplemented the findings from our experiment with two previously published research studies that used the same event summary. *Results:* Professionals exhibited a human error bias, but nevertheless believed that they were objective and confident in their conclusions. The lay control group also showed this human error bias. These data, along with previous research data, revealed that, given the equivalent investigative circumstances, this bias was significantly larger with the professional investigators, with an effect size of $d_{unb} = 0.97$, than the control group with an effect size of only $d_{unb} = 0.32$. *Conclusions:* The direction and strength of the human error bias can be quantified, and is shown to be larger in professional investigators compared to lay people. *Practical Applications:* Understanding the strength and direction of bias is a crucial step in mitigating the effects of the bias. The results of the current research demonstrate that mitigation strategies such as proper investigator training, a strong investigation culture, and standardized techniques, are potentially promising interventions to mitigate human error bias.

1. Introduction

Incident investigations are conducted in the wake of workplace events that claim the health and lives of workers. Regulatory and industry-based industrial investigators work to establish how and why events occur so that preventative measures can be implemented to minimize such events from occurring again in the future. Investigations reach direct conclusions regarding who and what was causal with implication to safety, liability, and compensation.

Investigative decisions are often made in evolving, complex and, at times, ambiguous investigative environments. Therefore, the cognitions of workplace investigators play a critical role in guiding the investigative process. Given the importance of investigators' judgments, it is disconcerting that research shows that workplace investigation professionals harbor a human error bias in which they tend to allocate more cause to workers in an event than the evidence dictates (e.g., DeJoy, 1987; LaCroix & DeJoy, 1989; MacLean, Brimacombe, & Lindsay, 2013; MacLean & Read, 2019; c.f., Woodcock, 1995).

The attribution literature demonstrates that all people, not just investigators, are inclined to attribute more cause to individuals involved in an event than the evidence would support (Gawronski, 2004; Nisbett & Ross, 1980). However, additional factors, often present in the workplace investigation, may increase the magnitude of this human error bias for professional investigators. Indeed, research has revealed a significant difference in human error bias between professionals and the lay population (MacLean & Read, 2019; MacLean et al. 2013), but the research does not provide details about the magnitude of this difference.

Kahneman and colleagues, in their book 'Noise' (2021) state that unidirectional and targeted corrective debiasing strategies can be helpful in circumstances where there is a clear likelihood that the bias is present and the direction and magnitude can be anticipated. The difference of human error bias between the professional and lay population indicates a systematic difference in

* Corresponding author.
E-mail addresses: Carla.maclean@kpu.ca (C.L. MacLean), i.dror@ucl.ac.uk (I.E. Dror).

https://doi.org/10.1016/j.jsr.2022.10.012
0022-4375/

the base-rate expectations (the predicted probability) of human error for events in these two populations. However, research that directly demonstrated the effects of biased base-rates on professional investigative judgment is scarce (see Meissner & Kassin, 2002). The goal of the current research is to examine the magnitude of the human error bias in professional investigation and provide concise, empirical evidence of the biasing effect of base-rates on professional judgments.

1.1. Sources of bias

Cognitive bias may undermine honest investigators' good intentions and hard work (e.g., MacLean & Dror, 2016; MacLean, Smith, & Dror, 2020). Bias is not random error, it is the systematic deviation from evidence-based objective judgment (Kahneman, Sibony, & Sunstein, 2021). Dror (2020) provides a taxonomy of eight sources of bias. It shows how sources of bias range from the architecture and constraints of the human brain, to contextual factors that are both environmental and event specific. Inaccurate base-rates are the fourth source of bias presented in the sources of bias hierarchy. Base-rate knowledge is an understanding and expected probability of the event, that is, the rate of occurrence of some feature in a population. The intuitive understanding of base-rates can be helpful and informative when drawing conclusions, however they can also distort decision-making by leading people to believe something is more or less likely to have occurred than is true in their present circumstance (Wickens et al., 2009).

1.2. Base-rate expectation

When determining how an event occurred, it is appropriate to consider both how well the evidence represents the event and also the base-rate frequency of that outcome, that is, its probability. Information in the workplace investigation is often ambiguous and incomplete (Hofmann & Stetzer, 1998). In investigative circumstances that include little diagnostic information, it is logical for investigators to draw on appropriate base-rate information to suggest a cause. For instance, investigators faced with a tire explosion event may think about base-rates of the causes of tire explosions and use this as a data point in the investigation process. Biased judgments can occur when population base-rates (the actual frequency of an occurrence) are inconsistent with the investigators' beliefs (see classic work on the availability and representativeness heuristics; Tversky & Kahneman, 1974).

1.2.1. Human error bias

In the case of workplace investigators, a number of factors may converge to develop biased base-rate expectations of human error. People are not skilled at rationally considering the effect of situational variables on the behavior of others, and instead tend to attribute an actor's actions to his/her/their disposition (e.g., lazy, careless, inattentive, reckless; fundamental attribution error, Nisbett & Ross, 1980; correspondence bias, Gawronski, 2004; Gilbert & Malone, 1995). We use the term human error bias to refer to this cognitive tendency in workplace situations where people overemphasize worker action as causal for an event.

Unique to industrial investigation is that investigators encounter contextual and motivational factors that work to enhance their natural human inclination to overemphasize operator error. For instance, investigators are often employees of the company in which they are conducting the investigation and thus have knowledge of the workers and workplace (Vincoli, 1994). A bias toward a worksite employee can be introduced via an investigator's expectations about a worker's behavior if an investigator is familiar with that worker's unsafe history (MacLean & Read, 2019; MacLean et al., 2013). Human error bias may also be enhanced because industry investigators, who are employees of the organization, may have motivational allegiances to protect the company they work for (Murrie, Boccaccini, Guarnera, & Rufino, 2013; Steensma, den Hartigh, & Lucardie, 1994).

Although the investigation literature endorses the need for unbiased investigations (Sklet, 2002), it also acknowledges that past recommendations have been skewed to be consistent with self interests, focus on those features that were preventable, and/or result in legal liability, rather than inclusive of all investigative findings (Hancock, 2020; Hopkins, 2000). Explicit or implied organizational pressure to find "human error" may guide (perhaps unconsciously) investigators' conclusions about the responsibility of the worker.

Once an investigator has a theory of worker fault, confirmation bias demonstrates how this hypothesis can shape information collection and interpretation to endorse his/her/their predominant perspective (Nickerson, 1998). For investigators, finding additional evidence to endorse their theory of worker cause is not difficult, as workers often deviate from the specific working protocol suggested by the organization and use more common, established practical and pragmatic practices adopted by that worksite's employees (Leveson, 2004). Moreover, the industrial investigation is a feedback impoverished environment that makes it difficult for investigators to hone their decision-making ability. Investigators may never know if their recommended changes, that focused on worker behavior, produced the future "non-event" or if it was simply the extremely low probability of a similar event reoccurring that led to it not occurring again (Woodcock, 1995).

For workplace investigators, the features discussed above should culminate in a rich network of cognitive associations that link different workplace events with operator error. These associations would generate an inflated understanding of the rate of human error in worksite events and ultimately, robust human error base-rate expectations. Thus, the first goal of this research was to investigate the effect that professionals' base-rates may have on their decisions about event cause compared to the undergraduate population.

1.3. Meta-cognitive evaluations of performance

People do not have direct access to their thoughts and because of this, when asked about their accuracy or objectivity, they rely on available cues to make inferences about their performance (Dunlosky & Tauber, 2014). People may rely on the imprecise cue of retrieval fluency to inform them of how confident they should be about their decisions. Information that is deeply learned or encountered recently facilitates access (Benjamin & Bjork, 1996). Thus, the familiarity professional investigators have with industrial events should result in them having relatively higher confidence levels than those in the lay population. People's ability to estimate their objectivity is equally flawed. When asked to estimate how objective they have been, people tend to turn inward and seek residue of bias in their decision making; any residue is largely undetectable (Wilson, Centerbar, & Brekke, 2002). Hence, in addition to being confident, investigators likely see their choices as more objective than they really are. The second goal of this research was to explore the confidence participants had in their decisions and their impressions about their objectivity.

1.4. Research questions

Do individuals have a bias to perceive people involved in events as causal? Would this tendency be demonstrated by participants allocating more cause to a worker than other factors when given industrial event information in which the evidence did not favor the equipment or worker as more causal (Experiment)? We also wanted to know whether professional workplace investigators would have a stronger human error bias than the undergraduate population, given equivalent investigative scenarios (Effect size analysis)? Last, we wanted to know whether participants would believe that they were objective and, also, how confident they were in their assessment of incident information.

To answer our research questions, we first explored the extent to which professional investigators and undergraduate participants allocated cause to a worker, versus an equally salient situational factor, when given a summary of a workplace event. The summary was crafted to be objectively balanced in its implication between two causal factors: a worker and tire (see Appendix A for the incident summary). The information given to the participants contained no diagnostic information that implicated either the worker or the tire as causing the event (Experiment). Participants in our experiment also rated their confidence and objectivity of their judgment.

Second, in addition to the data collection in our experiment, we conducted an effect size analysis that combined the findings from our experiment with two published research studies that used the same event summary. This was done to obtain a more robust measure of the extent to which the undergraduate and professional populations differed in the magnitude of the human bias.

Findings from these analyses reveal, in tangible terms, the magnitude of the human error bias that can be brought to an event scenario by professionals compared to those in the lay population. They will also examine the awareness individuals have for how their tendencies affect their judgments.

2. Method

Participants read a summary of a workplace event and developed an initial hypothesis of what most likely caused the event. They then reported their confidence in their cause allocation judgments and their level of objectively when making the judgment.

2.1. Participants

Participants were undergraduate students, N = 50 (20 males, 28 females, 2 undisclosed) and professional industrial incident investigators from the forestry industry, N = 15 (12 males and 3 females). All participants were fluent English speakers. Undergraduates ranged in age from 18 to 28 years (M = 20.40, SD = 2.62) and professional investigators from 29 to 62 years (M = 43.80, SD = 9.34). Our sample size of 15 professional participants was reached after 9 were removed (three because they had previously taken part in an experiment that used the materials, three because they did not consent to include their data in the data set, and three who misunderstood the industrial event to be the bystander's injury and not the tire explosion). See Table 1 for demographic information of professionals.

2.2. Materials and procedure

Participants completed testing booklets that contained the experimental task. The procedure was as follows: (a) participants provided consent; (b) watched a brief PowerPoint presentation outlining the types of judgments that they would be asked to make in the experiment; (c) read an summary of an industrial event (Appendix A) and were provided with two pieces of non-probative information about the event; (d) reported their demographic information; (e) hypothesized what cause the incident and then allocated a percentage of cause to each causal factor they indicated; (f) rated their level of objectivity and the likelihood that their hypothesis was correct; and (g) were debriefed about the purpose of the research.

2.2.1. Introductory presentation

A PowerPoint presentation showed participants the difference between underlying and direct incident cause. Direct cause(s) were described as worksite factors that immediately contributed to the event's occurrence such as a worker's action, faulty equipment, or environmental conditions. Underlying causes were described as the factors that underlie the direct causes (e.g., poor maintenance, housekeeping, management, funding, supervision or inadequate training). Participants were informed that their task was to provide their most likely hypothesis of what directly caused the event. We sought only direct causes in this research because they illuminate the conditions that were immediately responsible for the incident (see Hollnagel, 2012; Leveson, 2004).

2.2.2. Industrial incident summary

The summary described an event in which a worker was changing a tire on a truck (i.e., the tire man). When the tire man moved to the left to shut off the air flow to the tire the tire violently exploded, hitting a bystander who sustained serious injuries from the incident. The incident summary describes the environment, equipment, and actions of the worker that preceded the tire explosion. Initially developed for MacLean et al. (2013), the summary is a modified incident report from the Forestry industry in British Columbia and was refined with the support of a subject matter expert in forestry investigation. Prior research demonstrates that participants understand details of the event by reading the summary.

Participants were to allocate cause for the tire explosion event, not the injury of the bystander. The bystander's injury was a consequence of the tire explosion and this distinction was made clear to participants at the time of testing. The event description did not provide the necessary detail investigators require to determine tire or tire man cause. For instance, an example of diagnostic information that the tire man caused the tire to explode would have been information indicating that the tire man engaged in actions inconsistent with the manufacturer's instructions for tire and rim servicing prior to mounting and filling the tire. A tire cause for the explosion would have been information that indicated that the structure and function of the tire was compromised. No such diagnostic information was shared in the incident summary.

The summary included an image of the tire man, tire, truck, and bystander. Two different pictures of the tire man and two different pictures of the truck/tire were combined to create four event summary documents. The four event summaries were counterbalanced across participants. See Appendix A for one version of the incident summary.

Table 1
Professional-investigator demographic information.

	N	Mean	Median	Range	SD
Number of Years Investigating	13	15.69	15	2–30	7.39
Number of Investigations a Year	12	16.92	7.75	1–100	27.49
Number of Serious Investigations a Year	12	2.10	1.75	0–5	1.64
Number of Employees in Organization	14	864.07	87.5	25–5000	1401.64

Participants also received two pieces of additional information. This new information did not implicate or exonerate the tire or tire man as causal and was provided to enhance the resolution of the activity. The additional information about the tire man stated that "blood screening done after the incident demonstrated that there was no indication of drugs or alcohol in the Tire Man's blood stream." The additional tire information stated that "the rim of the tire had no visible cracks, dents or signs of metal fatigue." The order of evidence was counterbalanced across participants.

2.2.3. Demographics and filler activity

Participants reported their gender and age. Professionals were also asked if they had previously taken part in an experiment that used these materials and to state their level of experience in investigation. See Table 1 for professional-investigator demographic information. The experience questions we asked of professionals increased the number of items on their demographic questionnaire compared to undergraduates. To maintain a consistent 3–5 minutes interlude between receiving the event information and making their investigative judgments, undergraduates were given three short word games prior to making their judgments.

2.2.4. Ratings of event cause

Participants used an open-ended box to state their hypothesis of what directly caused the workplace event and then allocated the cause of the event between the factors they identified as causal (e.g., tire man 50%; tire 20%; tire cage 30%). Identifying something as mostly versus marginally causal at an initial stage of an investigation has implications for focusing subsequent information collection and decision making (e.g., Carlson & Russo, 2001; Jolley & Douglas, 2017). Using the percentage breakdown informed us about how causal participants perceive each worksite factor they identified. The purpose of the experimental activity was to assess individuals' base-rate tendency to infer worker causality when there is limited information; hence, participants were encouraged to develop an initial theory of event cause rather than respond "I don't know."

The researchers then allocated the percentages associated with participants' responses into three direct cause categories: (1) tire man fault, (2) tire fault, and/or (3) other fault. Any responses that described the root cause of tire man's actions were coded as "tire man cause." Tire man causes were: overinflating the tire, not following procedures, not properly trained, lack of supervision, not inflating the tire in a cage, lack of procedure, inadequate risk assessment, not perceiving tire damage or, the most frequently stated and general cause provided by participants, substandard tire man action. Tire causes were: wear and tear, lack of maintenance, defect in tire, sidewall failure, or the generally category of defective tire. Other causes were: location of bystander, substandard equipment (e.g., pressure gauge, tire cage), environmental factors, or location of the vehicle. Two, independent, raters obtained 82% agreement in categorizing 20% of undergraduate and 40% of professional participants' statements into cause categories. All inconsistent categorizations were resolved through discussion and referencing the categorization criteria.

2.2.5. Meta-cognitive reports

Participants used 9-point Likert scales to make their meta-cognitive judgments. They rated the likelihood that their hypothesis would be proven correct (1 = extremely unlikely to 9 = extremely likely) and the number that best reflected how objective they believe they were in their evaluation of the evidence (1 = not objective to 9 = extremely objective).

3. Results

3.1. Causal attributions

Given equal information about the tire and the tire man, all of which was non-diagnostic, both undergraduates and professionals allocated greater cause to the tire man compared to the tire (see Fig. 1). For undergraduates, the estimated difference between the average amount of cause allocated to the tire and the worker was 44.2% (95% CI [29.37, 59.03]). For professionals, the estimated difference between the average amount of cause allocated to the tire and the worker was larger than the undergraduates, 63% (95% CI [37.07, 88.93]). The magnitude difference in the human error bias between undergraduates and professionals, however, was not explored in this research activity. The lack of exploration was due to our small professional sample size, yet, mean differences suggest a stronger effect in the professional sample. The meta-analyses we present below have adequately robust sample sizes to inform us of the between-group differences on the human error bias.

3.2. Meta-cognitive reports of causal conclusions

Participants reported that it was likely that their hypothesis would be proven correct and that they were somewhat objective in their assessment of the incident information (See Table 2). Professionals did not demonstrate greater confidence in their hypothesis than the control undergraduates.

To explore the relationship between participants' metacognitive reports and their cause allocation a difference score was calculated. This difference score calculated the percentage of cause allocated to the tire subtracted from the percentage of cause allocated to the worker. Undergraduates' difference scores ranged from -90 to 100 ($M = 44.2$, $SD = 52.17$) and professionals difference scores ranged from -60 to 100 ($M = 63$, $SD = 46.82$). There was no correlation between participants difference scores and their (i) confidence that their hypothesis would be proven correct, undergraduates $r(49) = 0.16$, $p = .26$; professionals, $r(15) = 0.26$, $p = .34$ or (ii) level of objectivity, undergraduates $r(49) = 0.22$, $p = .13$; professionals, $r(15) = 0.16$, $p = .56$.

4. Discussion

Our experiment revealed a robust human error bias in both the undergraduate and the professional-investigator samples. Participants provided with the basic details of a serious industrial incident, and no information that indicated cause for the event, tended to *a priori* hypothesize that the worker was more causal than other factors. This human error bias is fueled by the cognitive tendencies discussed in the attribution literature. Investigators' conclusions may also be supported by one or more of the cognitive and/or motivational factors that reinforce causal conclusions of human error in the workplace environment.

Our findings were consistent with the metacognitive literature that has demonstrates that people have difficulty monitoring their cognitive processes (e.g., Benjamin & Bjork, 1996; Ehrlinger, Gilovich, & Ross, 2005). Participants rated themselves as fairly objective and fairly confident that their hypothesis would be proven correct; their ratings were unrelated to their cause allocations. Interestingly, we did not find a difference in the level of confidence demonstrated by professional investigators and the undergraduates. The importance of confidence in the investigation is that it can be used as a cue to third-party observers to assess credibility of a person's information (Weinsheimer, Coburn, Chong, MacLean, & Connolly, 2017). Our findings demonstrate that partic-

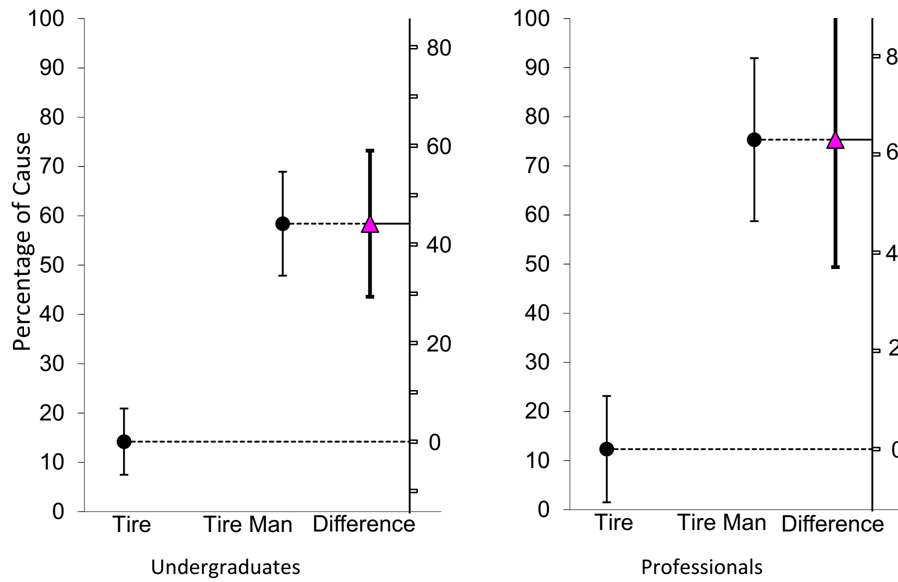


Fig. 1. Cause allocation. Figure from ESCI Data paired showing means and 95% CIs of percentage of cause allocated to the tire and the tire man. Undergraduates: The mean paired difference of 44.2 is shown with its 95% CI against a floating axis, whose zero is lined up with the tire mean. Undergraduate N = 50. Professionals: The mean paired difference of 63 is shown with its 95% CI against a floating axis, whose zero is lined up with the tire mean. Professional N = 15.

Table 2
Metacognitive reporting.

	N	Likert Scale	Mean	SD
Likelihood that your hypothesis will be proven correct.		1 (extremely unlikely) to 9 (extremely likely)		
Undergraduates	49	1 to 9	5.79	1.68
Professionals	15	1 to 9	5.80	2.88
Objectivity in your assessment of the evidence.		1 (not objective) to 9 (extremely objective)		
Undergraduates	49	1 to 9	5.96	1.50
Professionals	15	1 to 9	5.93	1.94

ipants' confidence was not calibrated to their cause allocation pattern. The implications of this miscalibration between judgments and confidence is that safety initiatives could proceed based more on the persuasiveness and confidence of the investigator rather than on the quality of their findings. In interpreting our findings, one must be careful, as professionals differ from undergraduates in a number of ways that make professionals not only susceptible to biases due to wrong or misapplied base-rates, but also due to organizational culture, loyalties, ideology, motivations, age (although individuals who are middle-aged generally score lower on the fundamental attribution error than younger adults; Follett & Hess, 2002), and other factors that distinguish them from non-professionals.

5. Effect size meta-analyses with previous research

The central aims of this paper were to explore if professional investigators are more inclined to allocate cause to the worker, versus an equally salient situational factor, and whether such an effect of professionals is comparable to that of lay population when given equivalent investigative circumstances. We addressed these aims by conducting two small random effects model, meta-analyses. One meta-analysis was conducted for undergraduate participants and the other meta-analysis for professional investigators. For each of the meta-analyses we used the data from three experiments, the current experiment, MacLean et al. (2013), and MacLean and Read (2019). These three experiments were all conducted with both undergraduate and professional populations; used the same tire explosion event stimuli; and asked participants to allocate cause to the factors that they perceived as directly responsible for the incident. In each one of the three experiments, both undergraduate and professional participants experienced the same experimental protocol. Collectively these three studies represent the reporting of 471 undergraduates and 118 professional industrial investigators, which are adequate sample sizes to derive measures of effect size.

These three experiments, however, also differed in three ways. We detail the experimental differences below for transparency and for showing the contribution of the present study above. These differences do not preclude the three studies from being combined in our two meta-analyses. First, in MacLean et al. (2013) and MacLean and Read (2019), prior to learning about the industrial event, half of the participants received information that the tire had a history of unsafe behavior and the other half that the tire man had a history of unsafe behavior. In the experiment presented above we did not provide participants with such additional biasing contextual information prior to reading about the event and allocating cause.

Second, in all three experiments, participants were provided with a balanced amount of information about the causality of both the tire and worker. However, in MacLean et al. (2013) and MacLean and Read (2019) participants were given evidence that indicated that both the tire and the tire man were equally causal for the event prior to allocating cause. In the experiment presented above, participants did not receive any information that indicated cause.

Third, all three studies educated participants about direct cause and asked that they only allocate cause to factors immediately causal for the event. However, in MacLean et al. (2013) and MacLean and Read (2019) participants indicated the category that the percentage of cause should be allocated to (tire man, tire or a number of other factors categorized as "other"). In the present experiment, participants provided percentages with open-ended responses and researchers categorized those responses as "tire man cause," "tire cause," or "other cause." As stated above, despite the differences in experiment format, they are appropriate to be included in our two meta-analyses. All experiments compared the amount of cause allocated to the tire man and tire in participant samples, which were balanced on the diagnosticity and amount of information that they received about the tire and tire man.

The difference in cause allocated to the tire man and the tire was calculated for both undergraduates and professional investigators by computing Cohen's *d* effect size (see Cummings, 2012). For a paired design Cummings recommends computing Cohen's *d* by dividing the mean difference by S_{av} (See Fig. 2 for S_{av} formula). Once computed, the calculated Cohen's *d* values were used in the Single Group Meta-Analysis of *d* page in the ESCI software. We reported the unbiased estimate of δ, d_{unb} , in our analysis, which is calculated from Cohen's *d* (See Fig. 2 for the *d* to d_{unb} conversation formula, and Cummings (2012) for more information on d_{unb}). A positive d_{unb} represent a greater amount of cause being allocated to the worker compared to the tire.

Cohen's *d* effects sizes are considered weak if below 0.20 and strong above 0.80 (Rubin, 2013). The undergraduate data random effects meta-analysis showed a cumulative d_{unb} of 0.32 (CI [−0.23, 0.86]), which reflects that undergraduates demonstrated a relatively weak tendency to allocate cause to a worker (when given information equal in its probative value, about the worker and tire involved in the event). In contrast, the professional random effects meta-analysis showed a cumulative d_{unb} of 0.97 (CI [0.31, 1.63]), demonstrating that professionals had a much stronger bias to report that the worker is causal when given the same information about the tire and worker (see Fig. 2 for the full meta-analyses results).

The heterogeneity calculations for the undergraduate meta-analysis revealed a significant (p <.001) I^2 value of 96.7 (Q = 91.51, T = 0.54), illustrating considerable heterogeneity between studies. The heterogeneity calculations for the professional meta-analysis also revealed heterogeneity with a significant (p <.001) I^2 value of 87.5 (Q = 15.39, T = 0.52). These sizeable heterogeneity results support our use of the random effects model as it is appropriate to assume that the variability demonstrated between studies is larger than what could be reasonably accounted for by sampling variability.

6. Overall discussion

The current paper aimed to empirically examine the biasing effect of base-rates on professional judgments, and reveal the relative strength of the human error bias in the professional investigation population. The cumulative results of the experimental activity and meta-analyses showed that both undergraduates and professionals are inclined to allocate cause to a worker, over other factors, given the same workplace event, but professionals significantly more so. This significantly higher effect in the professional population illuminates the effect of biasing factors, such as experience and expectation in judgments of event cause.

A method suggested to manage the effect of neglected, misapplied or misunderstood base-rates in decision making is to find the most up-to-date population rates for the event individuals are judging and to use those rates as a comparison point (Kahneman, 2011; Neal & Grisso, 2014). Decision makers are encouraged to assess the strength and credibly of the evidence they are evaluating by comparing it to the population base-rate. Unfortunately for professional investigators, human error bias is

Study name	Statistics for each study					
	d_{unb}	N	Variance	Lower limit	Upper limit	p-Value
Undergraduate-Investigators						
*MacLean et al. (2013)	0.45	136	0.008	0.27	0.63	< .001
*MacLean & Read (2019)	-0.15	143	0.007	-0.32	0.02	.08
*MacLean & Read (2019)	-0.33	142	0.007	-0.50	-0.16	< .001
*Current Experiment	1.40	50	0.040	1.02	1.81	< .001
Overall	0.32		0.080	-0.23	0.86	.25
Professional-Investigators						
*MacLean et al. (2013)	0.74	39	0.03	0.39	1.10	< .001
*MacLean & Read (2019)	0.42	64	0.06	0.17	0.69	< .01
* Current Experiment	2.37	15	0.25	1.45	3.55	< .001
Overall	0.97		0.11	0.31	1.63	< .01

Cohen's d_{unb} and 95% CI

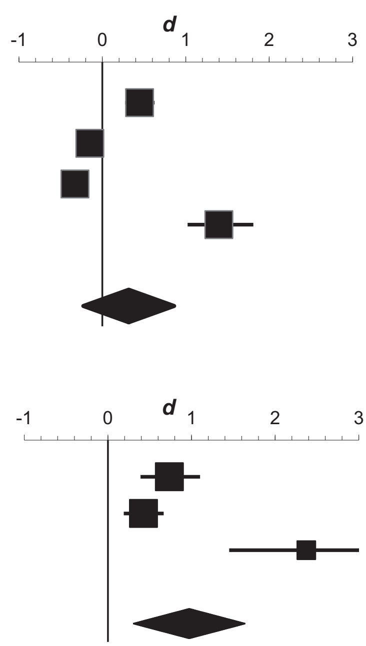


Fig. 2. Two random effects meta-analyses for the strength of the human bias in investigator reporting, with separate meta-analyses according to expertise level (lay undergraduate investigator vs professional investigator). The size of the study marker corresponds to the study weight. 95% confidence interval (CI) is represented by error bars. Cohen's *d* was computed for the paired designs by dividing the mean difference by $S_{av} = \sqrt{\frac{s^2\text{tire man} + s^2\text{tire}}{2}}$. Cohen's d_{unb} was then calculated by ESCI using the following formula $d_{unb} = \left(1 - \frac{3}{4(N-1)-1}\right) x d$ The positive d_{unb}s represent a greater percentage of cause being allocated to the worker than the tire. Two undergraduate samples were included for MacLean and Read (2019) as the total sample size (N = 285) was larger than what ESCI could accommodate for one study.

inherent to the population data investigators would use to establish the "objective" population base-rates. Investigator cognition is the method by which the population data are generated. The safety literature reports that 70%-80% of incidents are caused by operator error (Leveson, 2004). Our results, which were produced by using evidence that implicates a piece of machinery and the worker equally, demonstrates how different investigator conclusions may be from what would be an objective evaluation of the evidence.

A potentially fruitful method of managing the effects of the bias presented in this paper is to focus efforts on minimizing the human error bias; which is the foundation of investigators' base-rate expectations. Because human cognitive tendencies make it quick and efficient for people to conclude human error, investigators must make an effortful second step to thoroughly consider the circumstantial factors that may have contributed to the incident (Gilbert, Pelham, & Krull, 1988). This effortful second step is akin to a strategy discussed in the debiasing literature which is to consider alternatives. Considering alternatives has been moderately successful at dislodging biased patterns of thought and expanding professionals understanding the evidence (Anderson & Sechler, 1986; Chang, Berdini, Mandel, & Tetlock, 2018; Hirt & Markman, 1995; Lord, Lepper, & Preston, 1984).

Debiasing activities like considering alternatives require cognitive investment from the decision maker. Hence, for these strategies to work investigators must have the cognitive resources to engage in a more effortful evaluation of the evidence, as well as the motivation to do so. Professionals under time pressure or fatigued will be less likely to engage the second, more cognitively-involved, step of considering the potential situational variables affecting a worker's behavior and will be at greater risk of over attributing cause to a worker (see Danziger, Levav, & Avnaim-Pessoa, 2011; Fraser-Mackenzie & Dror, 2011 for illustrations of the effect of state on decision making).

Organizational culture also has a meaningful role in reducing this human error bias. The effects of allegiance to protect the company will be reduced with professionals who work in an organization that supports evidence-based decision making regardless of outcome. Also, prioritizing accuracy versus efficiency in the investigation will lead to more comprehensive investigative judgments. An emphasis on accuracy should offset the effect of quick and efficient cognitive tendency to conclude human error and promote deeper consideration of the evidence (see Ask, Granhag, & Rebelius, 2011). Maintaining a balanced investigative practice, however, is important. An emphasis on accuracy should be kept in perspective and not propel investigators to over analyze. Investigators with a compulsion to be flawless would experience increased stress and pressure.

Organizations that support comprehensive, data-driven investigations will likely also support the use of standardized practices and procedures. Standardization is a useful strategy as it reduces the level of subjectivity in decision making. Reducing subjectivity increases the reliability of the judgments and can support bias mitigation (Kahneman et al., 2021). Hence, it is beneficial to employ protocols for information collection, triaging acquired evidence, and employing a third-party independent investigator to have a "fresh look." These organizations will also have a higher likelihood of pursuing investigator training about cognitive bias.

Cognitively informed training, when done properly, is a valuable first step in mitigating bias as it educates individuals on how sources can bias decision making and that bias occurs outside of awareness, which then paves the way to accepting and adopting steps to minimize bias. Our experimental data demonstrated the lack of insight participants had regarding their behavior. Regardless of how they allocated cause, participants were consistently confident in their decisions and believed their choices were based on an objective assessment of the evidence. Investigators with knowledge that bias is affecting and subversive will be motivated to use strategies to mitigate its effects in their decision making.

The current research revealed interesting questions for future study. For instance, research shows that individuals' propensity to demonstrate different biases varies (Gertner, Zaromb, Schneider, & Matthews, 2016). Is it useful to obtain a finely-tuned understanding of an individual investigator's tendency to engage in the human error bias? Perhaps this could be achieved via items imbedded in their investigation tools. Once the magnitude of the bias is detected, how might that information be integrated into the process and procedures of the investigation or future training?

Future research may also consider how an investigator's organizational culture affects their base-rate expectations of human error. Person-centered theories of safety emphasis human traits as a leading factor of event cause and this approach prevails in many organizational settings. However, government and regulatory safety agencies are known to encourage a systems-centered approach to investigation which teaches that events are caused by multiple, interacting human and non-human factors (Holden, 2009). Agencies adopting a more systems-centered approach tend to use information collection techniques that support broad canvassing of information. One such information collection model is SHELL (Liveware, Software, Hardware, Environment; Edwards, 1972; Hawkins, 1993). SHELL encourages information collection about the people involved (Liveware), procedures and resources (Software), physical environment and workplace operations (Environment), as well as physical structures and equipment (Hardware). The professionals in our meta-analysis were largely from industry. Future research should consider if the structure of system-based investigations yields different levels of human error bias at the initial stage of the investigation.

Future research may examine the generalizability of our findings by conducting experiments that use a variety of industrial incident scenarios and include those findings in the meta-analyses shared in this research. We would also welcome research that explores human error bias and investigative decision making beyond the initial hypothesis stage of the industrial investigation. This would expand the understanding of how numerous complexities found in real-world investigations interact with base-line biases (see Woodcock, Drury, Smiley, & Ma, 2005).

7. Conclusion

The current research provided evidence of the biasing effect of professionals' base-rate expectations on their judgments. People have a general tendency to interpret an actor's behavior more as a function of the actor's disposition than the evidence would support. The current research shows that professionals are not immune from this bias –in fact, this human error tendency is actually amplified with professional investigators.

7.1. Practical Applications

The origins of professionals' stronger bias to see the worker as causal compared to the lay population may be multifaceted, but the implication is singular. Allocating a disproportionate amount of cause to a worker involved in an event undermines the safety of work environments as the role of causal factors, other than the worker, may be underreported and thus not given sufficient attention. Investigator cognitive training, a strong investigation culture, and standardized techniques are directions to mitigating human error bias in industrial investigation. Reducing the effect of human error bias will support more complete and accurate understanding of the factors that threaten worker safety and wellbeing.

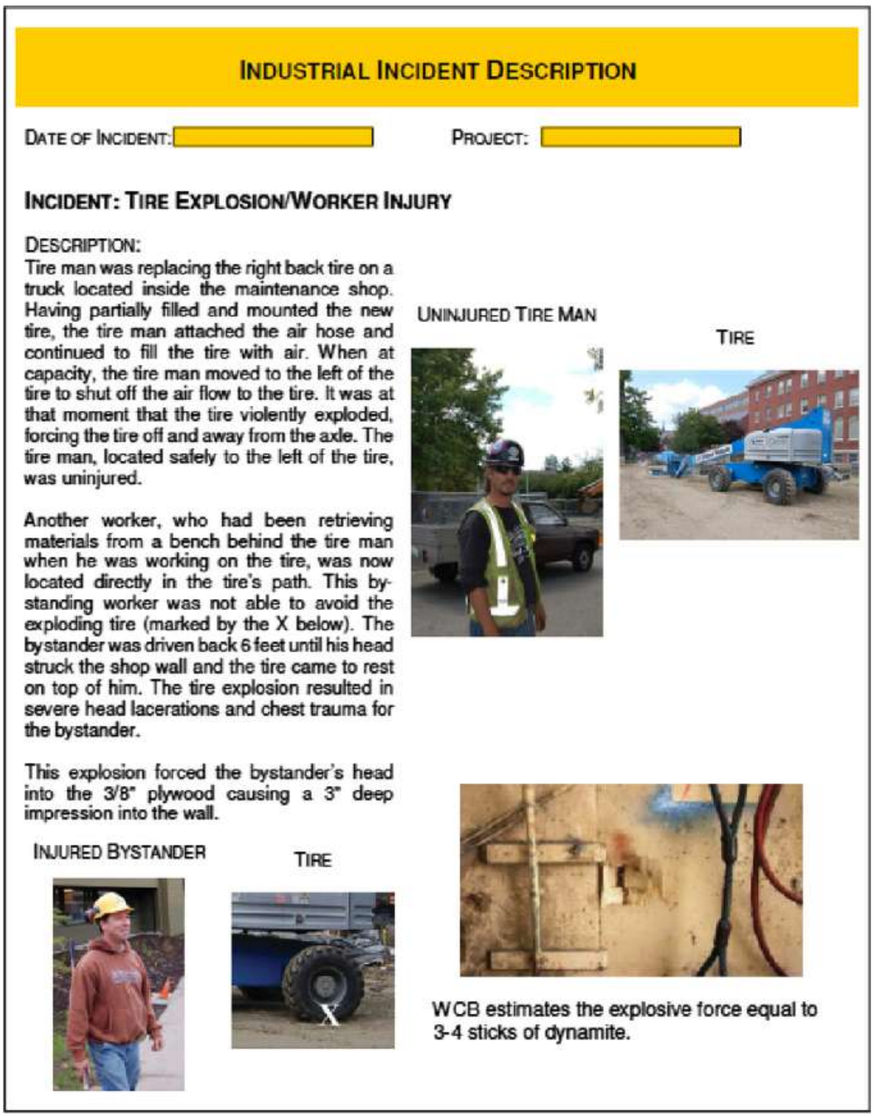
Conflict of Interest

None.

Acknowledgements

We would like to thank the workplace investigators who volunteered their time to participate in this research and the British Columbia Forest Safety Council for their support of this work.

Appendix A



INDUSTRIAL INCIDENT DESCRIPTION

DATE OF INCIDENT: PROJECT:

INCIDENT: TIRE EXPLOSION/WORKER INJURY

DESCRIPTION:
Tire man was replacing the right back tire on a truck located inside the maintenance shop. Having partially filled and mounted the new tire, the tire man attached the air hose and continued to fill the tire with air. When at capacity, the tire man moved to the left of the tire to shut off the air flow to the tire. It was at that moment that the tire violently exploded, forcing the tire off and away from the axle. The tire man, located safely to the left of the tire, was uninjured.

Another worker, who had been retrieving materials from a bench behind the tire man when he was working on the tire, was now located directly in the tire's path. This bystanding worker was not able to avoid the exploding tire (marked by the X below). The bystander was driven back 6 feet until his head struck the shop wall and the tire came to rest on top of him. The tire explosion resulted in severe head lacerations and chest trauma for the bystander.

This explosion forced the bystander's head into the 3/8" plywood causing a 3" deep impression into the wall.

UNINJURED TIRE MAN

TIRE

INJURED BYSTANDER

TIRE

X

WCB estimates the explosive force equal to 3-4 sticks of dynamite.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition: The 27th Carnegie Symposium on Cognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.

Carlson, K. A., & Russo, J. E. (2001). Biased interpretation of evidence by mock jurors. *Journal of Experimental Psychology: Applied, 7*(2), 91–103. https://doi.org/10.1037/1076-898X.7.2.91.

Chang, W., Berdini, E., Mandel, D. R., & Tetlock, P. E. (2018). Restructuring structured analytic techniques in intelligence. *Intelligence and National Security, 33*(3), 337–356. https://doi.org/10.1080/02684527.2017.1400230.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Taylor & Francis Group: Routledge.

Danziger, S., Levav, J., & Avnaim-Pessoa, L. (2011). Extraneous factors in judicial decisions. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 108*(17), 6889–6892.

DeJoy, D. M. (1987). Supervisor attributions and responses for multi-causal workplace accidents. *Journal of Occupational Accidents, 9*, 213–223. https://doi.org/10.1016/0376-6349(87)90013-7.

References

Anderson, C. A., & Sechler, E. S. (1986). Effects of explanation and counterexplanation on the development and use of social theories. *Journal of Personality and Social Psychology, 50*, 24–34. https://doi.org/10.1037/0022-3514.50.1.24.

Ask, K., Granhag, P. A., & Rebelius, A. (2011). Investigators under influence: How social norms activate goal-directed processing of criminal evidence. *Applied Cognitive Psychology, 25*(4), 548–553. https://doi.org/10.1002/acp.1724.

Dror, I. E. (2020). Cognitive and human factors in expert decision making: Six fallacies and the eight sources of bias. *Analytical Chemistry, 92*(12), 7998–8004. https://doi.org/10.1021/acs.analchem.0c00704.

Dunlosky, J., & Tauber, S. K. (2014). Understanding people's metacognitive judgments: An isomechanism framework and its implications for applied and theoretical research. In J. Perfect & D. S. Lindsay (Eds.), *The Sage Handbook of Applied Memory* (pp. 444–464). Thousand Oaks, CA: Sage.

Edwards, E. (1972). Man and machine: Systems for safety. In *Proceedings of British Airline Pilots Associations Technical Symposium* (pp. 21–36). London: British Airline Pilots Associations.

Ehrlinger, J., Gilovich, T., & Ross, L. (2005). Peering into the bias blind spot: People's assessments of bias in themselves and others. *Personality and Social Psychology Bulletin, 31*(5), 680–692.

Follett, K. J., & Hess, T. M. (2002). Aging, cognitive complexity, and the fundamental attribution error. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 57*(4), P312–P323. https://doi.org/10.1093/geronb/57.4.P312.

Fraser-Mackenzie, P. A. F., & Dror, I. E. (2011). Dynamic reasoning and time pressure: Transition from analytical operations to experiential responses. *Theory and Decision, 71*(2), 211–225. https://doi.org/10.1007/s11238-009-9181-z.

Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology, 15*, 183–217. https://doi.org/10.1080/10463280440000026.

Gertner, A., Zaromb, F., Schneider, R. & Matthews, G. (2016). The assessment of biases in cognition: Development and evaluation of an assessment instrument for the measurement of cognitive bias (Publication No. 2015-14120200002-002). MITRE Corporation. https://www.mitre.org/publications/technical-papers/the-assessment-of-biases-in-cognition.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21–38. https://doi.org/10.1037/0033-2909.117.1.21.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal Of Personality And Social Psychology, 54*(5), 733–740. https://doi.org/10.1037/0022-3514.54.5.733.

Hancock, P. A. (2020). Science in court. *Theoretical Issues in Ergonomics Science, 21*, 266–284. https://doi.org/10.1080/1463922X.2019.1640312.

Hawkins, F. H. (1993). *Human factors in flight, second edition.* Ashgate Publishing Company: Aldershot, UK. IATA, 2006. Safety Report.

Hirt, E., & Markman, K. (1995). Multiple Explanation: A Consider- an-Alternative Strategy for Debiasing Judgments. *Journal of Personality and Social Psychology, 69*, 1069–1086. https://doi.org/10.1037/0022-3514.69.6.1069.

Holden, R. J. (2009). People or systems? to blame is human. the fix is to engineer. *Professional Safety, 54*(12), 34–41.

Hofmann, D. A., & Stetzer, A. (1998). The role of safety climate and communication in accident interpretation: Implications for learning from negative events. *Academy of Management Journal, 41*(6), 644–657. https://doi.org/10.2307/256962.

Hollnagel, E. (2012). *FRAM : The functional resonance analysis method : Modelling complex socio-technical systems*. Surrey, Great Britian: Ashgate Publishing Ltd.

Hopkins, A., 2000. *Lessons from Longford*. ISBN 1 86468 422 4, CCH Australia Limited, Australia.

Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology, 47*(8), 459–469. https://doi.org/10.1111/jasp.12453.

Kahneman, D. (2011). *Thinking, fast and slow*. Toronto, Canada: Random House.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little Brown Spark.

LaCroix, D. V., & DeJoy, D. M. (1989). Causal attributions to effort and supervisory response to workplace accidents. *Journal of Occupational Accidents, 11*, 97–109. https://doi.org/10.1016/0376-6349(89)90012-6.

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science, 42*, 237–270. doi: 10.1.1.141.697.

Lord, C., Lepper, M., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*, 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231.

MacLean, C. L., Brimacombe, C. A. E., & Lindsay, D. S. (2013). The role of A Priori knowledge and tunnel vision education. *Law and Human Behavior, 37*, 441–453. https://doi.org/10.1037/lhb0000056.

MacLean, C. L. & Dror, I. E. (2016). A primer on the psychology of cognitive bias. In A. Kesselheim and C. T. Robertson (Eds.), *Blinding as a solution to bias in biomedial science and the courts: A multidisciplinary approach*.

MacLean, C. L., Smith, L., & Dror, I. E. (2020). Experts on trial: Unearthing bias in scientific evidence. *University of British Columbia Law Review, 53*(1), 101–140.

MacLean, C. L., & Read, J. D. (2019). An illusion of objectivity in workplace investigation: The cause analysis chart and consistency, accuracy, and bias in judgments. *Journal of Safety Research, 68*, 139–148. https://doi.org/10.1016/j.jsr.2018.12.008.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior, 26*(5), 469–480. https://doi.org/10.1023/A:1020278620751.

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them?. *Psychological Science, 24*(10), 1889–1897. https://doi.org/10.1177/0956797613481812.

Neal, T. M. S., & Grisso, T. (2014). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law, 20*(2), 200–211. https://doi.org/10.1037/a0035824.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220. https://doi.org/10.1037/1089-2680.2.2.175.

Nisbett, R. E., & Ross, L. D. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Rubin, A. (2013). *Statistics for evidence-based practice and evaluation* (third ed.). Belmont, CA: Cengage Learning.

Sklet, S. (2002). Methods for accident investigation. *Norwegian University of Science and Technology*, 1–75.

Steensma, H., den Hartigh, E., & Lucardie, E. (1994). Social categories, just world belief, locus of control, and causal attributions of occupational accidents. *Social Justice Research, 7*, 281–299. https://doi.org/10.1007/BF02334835.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

Vincoli, J. W. (1994). *Basic guide to accident investigation and loss control*. New York: John Wiley and Sons.

Weinsheimer, C. C., Coburn, P. I., Chong, K., MacLean, C. L., & Connolly, D. A. (2017). Perceptions of credibility for a memory report of a single versus repeated event. *Applied Cognitive Psychology, 31*(4), 414–423. https://doi.org/10.1002/acp.3340.

Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False alerts in air traffic control conflict alerting system: Is there a "Cry Wolf" Effect?. *Human Factors, 51*(4), 446–462. https://doi.org/10.1177/0018720809344720.

Wilson, T. D., Centerbar, D. B., & Brekke, N. (2002). Mental contamination and the debiasing problem. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heurisitics and Biases: The Psychology of Intuitive Judgment*. Cambridge, UK: Cambridge University Press.

Woodcock, K. (1995). Bias in real-world accident cause-finding. In A. C. Bittner & P. Champney (Eds.), *Advances in industrial ergonomics and safety* (pp. 907–914). London, UK: Taylor & Francis.

Woodcock, K., Drury, C. G., Smiley, A., & Ma, J. (2005). Using simulated investigations for accident investigation studies. *Applied Ergonomics, 36*(1), 1–12. https://doi.org/10.1016/j.apergo.2004.10.002.

Dr. Carla L. MacLean is a faculty member in Psychology at Kwantlen Polytechnic University. Her background and publications consider the factors that affect memory, as well as, judgement and decision making in worlds outside of the lab (e.g., industrial incident investigation, forensic investigation, legal decision making, workplace investigation). Her research has been funded by the Social Sciences and Humanities Research Council of Canada, as well as, a number of other funding agencies. Dr. Carla MacLean regularly speaks at both practitioner and academic conferences.

Dr. Itiel Dror's academic work relates to theoretical issues underlying human performance and cognition. His research examines the information processing involved in perception, judgment and decision-making. Dr. Itiel Dror has published dozens of peer-reviewed scientific articles and serves as Associate Editor and on the Editorial Boards of several scientific journals. His specialty is in taking the most theoretical scientific understanding of the human mind, brain and cognition, and translating it into practical and tangible ways to improve human performance in real world domains. This applied research has primarily focused on enhanced cognition through training, decision-making, and use of technology.

Journal of Safety Research 84 (2023) 384–392

Contents lists available at ScienceDirect

Journal of Safety Research

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

Mobile phone penalties and road crashes: Are changes in sanctions effective?

Jane M. Fry *

Demography and Ageing Unit, University of Melbourne, Melbourne, Australia

ARTICLE INFO

Article history:
Received 6 July 2022
Received in revised form 17 August 2022
Accepted 1 December 2022
Available online 12 December 2022

Keywords:
Road crash severity
Fines
Policy intervention
Mobile phone
Britain

ABSTRACT

Introduction: Road crashes are a major, preventable cause of death and serious injury. Being distracted by a mobile phone while driving can increase the risk of a crash by three to four times and increase crash severity. To reduce distracted driving, on 1 March 2017 the penalty for using a hand-held mobile phone while driving in Britain doubled to £200 and six penalty points. *Method:* We examine the effects of this increased penalty on numbers of serious or fatal crashes over 6 weeks either side of the intervention using Regression Discontinuity in Time. *Results:* We find no effect of the intervention, suggesting the increased penalty is not effective in reducing the more serious road crashes. *Conclusions:* We rule out an information problem and an enforcement effect, concluding the increase in fines was insufficient to change behaviour. With very low detection rates of mobile phone use, our result could occur if the perceived certainty of punishment remained very low after the intervention. *Practical application:* Future technology will increase the ability to detect mobile phone usage, and there may be fewer road crashes if the solution is to raise awareness of such technology and publicise numbers of offenders caught. Alternatively, a mobile phone blocking application could avert the problem.

1. Introduction and background

As road crashes are a major, preventable cause of death and serious injury, governments must address this public health problem through evidence-based interventions. One such intervention is the laws and penalties to induce safe driving, as compared to individual level safety programs (Tran, Hyder, Kulanthayan, Singh, & Umar, 2009). Some motor-vehicle offenses can lead to significant costs when individuals are injured or die. Driver behavior affecting traffic violations is more important than other factors in explaining numbers of crashes and their severity (Cardamone, Eboli, Forciniti, & Mazzulla, 2017). It is for these reasons that the link between penalties and road crashes stemming from driver behavior should be explored.

While there is much literature in prominent journals linking specific penalties to driver behavior through reductions in expected utility that lessen the desire to offend (Bates, Soole, & Watson, 2012; Gehrsitz, 2017; Traxler, Westermaier, & Wohlschlegel, 2018) or similarly linking general driving bans to crashes (Anderson & Rees, 2015; De Paola, Scoppa, & Falcone, 2013; DeAngelo & Hansen, 2014; Luca, 2015), there remains considerable uncertainty about the best size and structure of incentives needed to further reduce serious crashes and fatalities (see, for example, Bourgeon & Picard, 2007; De Paola et al., 2013; Hansen, 2015; Kantorowicz-Reznichenko, 2015; Montag, 2014). Moreover, there does not generally seem to be a focus on evaluating the effect of specific changes in penalties on crash outcomes (exceptions include Chang, Chang, & Fan, 2020; Cooper, Gehrsitz, & McIntyre, 2018). This is important as there may be risk-compensating behavior (whereby individuals drive in a more risky manner in terms of non-targeted other behaviors) or a lack of change in perceived risks associated with the targeted behavior that affect this behavior but leave crash risk unchanged (Dionne, Fluet, & Desjardins, 2007; Winston, Maheshri, & Mannering, 2006).

Deterrence theory postulates that to deter individuals from violating the law, punishments should be evaluated in terms of certainty, celerity (swiftness), and severity (Becker, 1968; Tomlinson, 2016). In a reconceptualization of the theory, Stafford and Warr (1993) incorporate punishment avoidance whereby offending may increase if offenders go unpunished as individuals perceive the probability of being caught is low (whether because their own offending or that of others in general has gone unpunished). This is important, as it may explain why increasing severity alone does not always work.

* Address: Demography and Ageing Unit, Melbourne School of Population and Global Health, Level 5, 207 Bouverie Street, The University of Melbourne, VIC 3010, Australia.

E-mail address: Jane.Fry@unimelb.edu.au

https://doi.org/10.1016/j.jsr.2022.12.001
0022-4375/

Experiments have shown that talking on a mobile phone reduces reaction time for drivers (Farmer, Braitman, & Lund, 2010).[1] Distractions take multiple forms, such as cognitive (conversation to task related), visual (looking at the phone), auditory (listening to the phone), and manual (holding the phone or dialing/texting; McEvoy, Stevenson, & Woodward, 2007). Being distracted by a mobile phone while driving can have serious consequences for health, increasing the risk of a crash by three to four times (Elvik, 2011; McEvoy et al., 2007; World Health Organization, 2018), although most of the effect seems to work through use of hand-held mobile phones (Backer-Grøndahl & Sagberg, 2011). Indeed, experiments have shown that talking on a mobile phone reduces reaction time for drivers (Farmer et al., 2010). A link has also been established between smartphone use and crashes in California with effects of about 2.9 % (Hersh, Lang, & Lang, 2022).

Based on survey data, U.S. drivers claimed to spend 6.7 % of driving time talking on mobile phones, although actual rates may be much higher (Farmer et al., 2010). In that study it was estimated that some 19 % of fatal crashes could have been avoided if there were no drivers talking on mobile phones. In a meta-analysis, Elvik (2011) found crash risk to be about three times higher when mobiles were used by drivers and McEvoy et al. (2007) and Redelmeier and Tibshirani (1997) found a fourfold increase in crash risk. Klauer et al. (2014) found a more than twofold increase in crash risk among experienced drivers dialing a mobile phone, but an eightfold increase among novice drivers. Using a mobile phone while driving has also been found to increase crash severity (Donmez & Liu, 2015).

One way of potentially reducing these crash numbers is to control driver behavior, but introducing laws governing mobile phone use does not guarantee compliance (see, for example, Abouk & Adams, 2013). Individuals may use a mobile phone for convenience. A system of penalties for drivers caught using mobile phones is therefore required to induce would-be offenders to obey the laws and to penalize violations. This can be achieved to a greater or lesser extent by changing incentives through introducing harsher penalties for errant driving behaviors.

Evidence on the effectiveness of mobile phone bans is limited to studies of the United States and is mixed. Abouk and Adams (2013) found reductions in fatal crashes involving single car/solo driver, but only immediately following text messaging ban imposition in the United States. Burger, Kaffine, and Yu (2014) found no effect of a ban on hand-held mobile phone use in California. Ferdinand et al. (2014) found fatal crashes were reduced in response to primarily enforced but not secondarily enforced laws. Rocco and Sampaio (2016) found reductions in fatalities in response to both cell phone and texting bans in the United States. Therefore more evidence is required to determine if penalties are effective in Britain.

On 1 December 2003, a law was introduced prohibiting the use of hand-held mobile phones while driving in Britain (UK Department for Transport, 2018). Initially the penalty was a £30 fine. Subsequently there was an increase to £60 and three penalty points introduced in 2007 and the fine increased to £100 in 2013 (UK Department for Transport, 2016a). However, it has been noted that the fine increase alone was not sufficient to significantly affect numbers of drivers using mobile phones and that only the introduction of penalty points in 2007 saw a significant decline in offenses (UK Department for Transport, 2016a). The efficacy of these mobile penalties remains in doubt, as in 2016 there were still 11,961 offenders found guilty of using a hand-held mobile phone while driving (UK Ministry of Justice, 2018).

Rather than focus on numbers of offenses committed, in this paper the effects of a doubling of penalties were evaluated (both fines – from £100 to £200 – and penalty points – from 3 to 6) on 1 March 2017 using data on numbers of serious or fatal crashes reported to police in Britain. Raising these penalties was designed to reduce distracted driving, which has a worse effect on driving than being over the drink-drive limit (UK Department for Transport, 2016a). In addition to these penalty increases, drivers caught twice or accruing 12 penalty points faced going to court, disqualification, and fines of up to £1000. Newly qualified drivers (within two years of gaining a license) could also have their license revoked and truck or bus drivers could be suspended. The doubling of penalties were originally designed to reflect the seriousness of the offense, to treat all drivers equally, and to increase the deterrent effect (UK Department for Transport, 2016a). Fines have a direct link with foregone income, whereas penalty points have an indirect link via license suspension if enough points are accrued. License suspension may increase the cost of transport and in some cases lead to job loss. These are the channels through which changes in penalties should have been effective.

The intervention was well publicized, with some 699 related newspaper articles published between 18 January and 12 April 2017 (Fig. A1). Internet searches related to the increased penalty for mobile phone use also peaked at the time of the intervention but there was sustained activity both before and after the intervention (Fig. A2).

In this paper, it was hypothesized that there will be less use of mobile phones in response to higher penalties, reducing driver distraction and leading to fewer serious or fatal crashes. The focus on these types of crashes is important as in 2017 the cost per fatal crash was £2.1 m and the cost per serious crash was £244,000 (UK Department for Transport, 2018). To the authors' knowledge this is the first attempt to examine the effects of this road safety intervention on number of serious or fatal crashes. In this respect, this research question is unique as health outcomes are focused on rather than general 'road safety' as an outcome of policy. Moreover, 'there remains a dearth of evidence on the effectiveness of interventions to reduce distracted driving' (World Health Organization, 2018, p. 45) and with this analysis of the mobile phone intervention this study seeks to partially fill this gap.

After accounting for longer term trends and seasonal/day of the week effects, Regression Discontinuity in Time (RDiT) analysis was used to see what happened in Britain before and after the intervention and Difference-in-Difference (DiD) analysis was used to compare crashes on different road types that may have been differentially affected to identify 'treatment' effects. This investigation sheds light on the link between penalties and road crashes, which, to date, has been missing from the literature. The results provide critical evidence on the ultimate effectiveness of changes in sanctions designed to improve road safety.

2. Data and summary statistics

British Stats19 data were used on all road crashes involving a serious injury or fatality that are reported to police (UK Department for Transport, 2019). Stats19 is a comprehensive source of data on the number and characteristics of such road crashes. Crash severity is determined by the most serious casualty in the crash and is classified as fatal (death within 30 days), serious (injuries typically requiring hospitalization), and slight (most other injuries). In this analysis the focus is on serious or fatal crashes as they account for over three-quarters of the total costs of injury crashes (UK Department for Transport, 2018) and represent a non-random subset of total crashes (Fry & Farrell, 2022). Moreover, because police-reported crashes are used there will be less under-

[1] Conversely, Papadimitriou, Argyropoulou, Tselentis, and Yannis (2019) find mobile phone use is negatively correlated with speed, indicating there is compensatory behavior at play.

reporting of serious or fatal crashes (UK Department for Transport, 2016b).

Crash numbers are aggregated by day and Local Authority area for six weeks either side of the penalty introduction on 1 Mar 2017, from 18 Jan 2017 to 12 Apr 2017. A six week window was adopted as there is some evidence new habits are formed over that period (Gardner & Lally, 2018; Lally, van Jaarsveld, Potts, & Wardle, 2010) and this allows for the exclusion of any effects of a speeding fine intervention introduced on 24 April in England and Wales. That intervention involved permanently increasing the most serious category of speeding fines from 100 % to 150 % of weekly income.

Any unusual climatic events surrounding the intervention are controlled using UK Met Office Integrated Data Archive System (MIDAS) data on local temperatures and precipitation (UK Met Office, 2018). There is a small but not significant increase in mean numbers of serious or fatal crashes after the intervention (Table 1). However, as there may be other influences on these numbers, an econometric model is used to estimate the true effect of the intervention.

3. Models

To identify the treatment effect of the intervention (which involved a doubling of fines and penalty points for using a mobile phone while driving in England, Wales or Scotland) a two-step Regression Discontinuity in Time (RDiT) procedure was adopted similar to that used by Castriota and Tonin (2019), De Paola et al. (2013) and Hausman and Rapson (2018). RDiT is a variant of Regression Discontinuity Design – a technique that has been used since the 1960 s (Imbens & Wooldridge, 2009) – in which time is the running variable and treatment begins at a particular known point in time, introducing a discontinuity in the series of interest (Hausman & Rapson, 2018).[2] When examining the effects of the intervention, it was assumed that drivers involved in serious or fatal crashes are treated (subject to the higher penalty regime) if the crash occurs on or after 1 March 2017 and untreated if the crash occurs earlier.

In the first step a fixed effects model is estimated for crashes in which someone is killed or seriously injured (KSI) including controls for trends, seasonality, weekdays, public holidays, and weather conditions over a 10-year period to get precise estimates of these effects. It is important to control for these factors in the first step in order to minimize the potential for these factors to result in spurious inferences regarding the treatment effect in the second step. By including this rich set of controls, the first step model is tightly identified (with narrower effects) by variations in crashes by day within small geographic areas. This should allow the authors to estimate the treatment effect among otherwise similar LA-days.

$$ksi_{it} = \beta_1' S_{it} + \beta_2' PH_t + \beta_3' W_{it} + \alpha_i + \varepsilon_{it} \qquad (i = 1, \cdots, 380;\ t = -3035, \cdots, 616) \tag{1}$$

where ksi_{it} is the daily number of crashes in which someone is killed or seriously injured in each LA, S_{it} is a vector of seasonal, weekday and trend variables, PH_t is a vector of public holiday dummies (New Year's day, Good Friday, Easter Monday, May bank holiday, Spring bank holiday, August bank holiday, Christmas day, Boxing day, 2011 Royal Wedding and 2012 Diamond Jubilee), W_{it} is a vector of controls for daily minimum, maximum temperature and precipitation, α_i is an LA fixed effect and ε_{it} is a normal error term.

In step two, the residuals from step one were used (adjusted numbers of crashes, $\tilde{ksi}_{it}$) and estimate a RDiT model using pooled OLS.

$$\tilde{ksi}_{it} = \gamma_0 + \gamma_1 Post_t + u_{it} \qquad (i = 1, \cdots, 380;\ t = -T, \cdots, T) \tag{2}$$

$Post_t$ is a dummy variable identifying days post intervention and γ_1 identifies the treatment effect averaged over the bandwidth T, where T is set at 7, 28 and 42 days to explore how the effect may change over time. The use of 42 days is to control for potential new habit formation as drivers cease to use mobile phones while driving.

This two-step RDiT approach – termed the augmented local linear approach by Hausman and Rapson (2018) – improves on the usual one-step approach (in which trend, seasonal, holiday, weather and treatment effects would be estimated from a single model), as the former accounts for longer term trend, seasonal/holiday and weather effects over 10 years rather than short term effects estimated over a much shorter period under the one-step approach. This allows for more precise estimates of these controls.

Causal effects of the treatment on the outcome can be estimated by assuming observations close to each other but either side of the treatment 'threshold' are otherwise the same. By taking a relatively short window either side of the threshold and controlling for other factors (public holidays, weather and Local Authority fixed effects) in the first step of this procedure, there is very little scope for other factors to have changed and affected the results, so the estimate identifies the causal effect, assuming the treatment is independent of confounding factors. The result is interpreted as the average treatment effect. Robust standard errors are estimated as there is only one level in the treatment (all LAs under consideration are subject to the treatment) and this precludes clustering the standard errors.

4. Results

Crash numbers vary around trend, seasonal, day of week, and public holiday effects associated with traffic volumes, vehicle, and environmental safety features and some aspects of driver behavior (resulting from say fatigue or time pressure on different types of days). Numbers also vary with weather conditions such as temperature and rainfall (affecting driving speeds and stopping distances and general road conditions). These factors are likely to be long term phenomena, so their effects over a 10 year period are estimated. For illustrative purposes, Fig. 1 shows serious or fatal crash numbers for Britain, with predicted numbers of crashes (conditional on long term (national) trend, monthly seasonal and public holiday effects).[3] Based on this picture, numbers of serious or fatal crashes are expected to increase significantly immediately after the policy intervention due to seasonality/public holidays. It is therefore essential that these effects (together with day of the week effects) be removed from the data in order to identify the effects of the penalty increase. Numbers of more serious crashes can also vary with weather conditions, so this study wants to ensure any estimated effects of the intervention are not confounded by particularly good or bad weather conditions immediately after the intervention.

Focusing on the period six weeks either side of the intervention, Fig. 2 shows predicted numbers of crashes across Britain based on trend, seasonal, weekday, and public holiday effects, as well as actual number of crashes. From the chart there does not appear to be an appreciable drop in actual number of serious or fatal

[2] The two step RDiT procedure is similar to the Interrupted Time Series (ITS) approach, although ITS uses one step. Both procedures rely on the assumption of smoothness of all confounders across the threshold.

[3] For simplicity, weekday effects are not shown.

Table 1
Summary statistics on daily serious or fatal crashes at Local Authority level.

Time	Obs	Mean	Std. Dev.	Min	Max
18 Jan 2017–28 Feb 2017	15,960	0.1623	0.4276	0	5
1 Mar 2017–12 Apr 2017	16,340	0.1679	0.4318	0	4

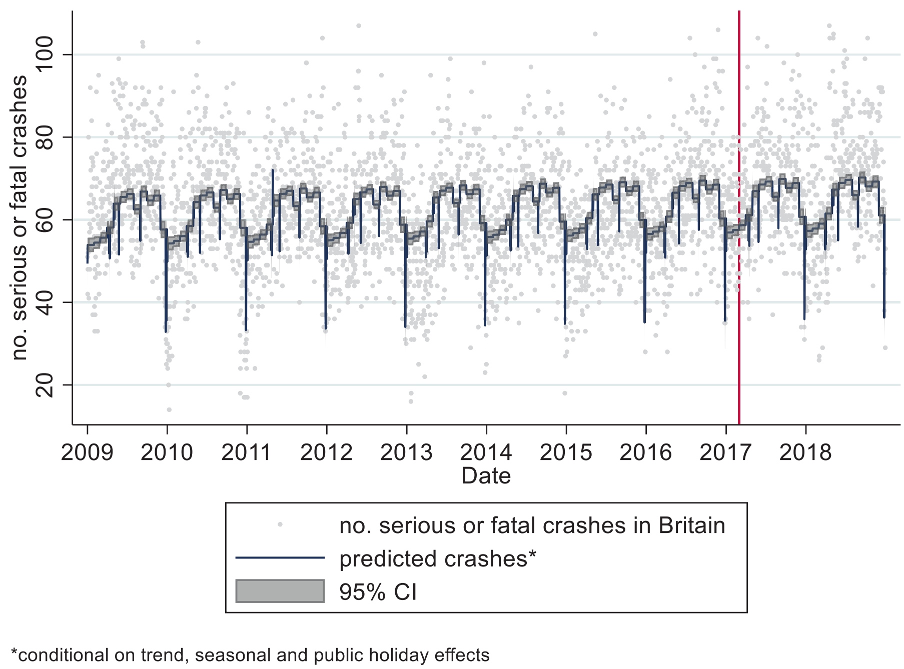


Fig. 1. Serious or fatal crashes and predictions over 10 years, Britain. Predictions are based on a regression model incorporating a national trend, seasonal and public holiday dummies over the period 1/1/2009 to 31/12/2018. The date of the intervention is indicated by the red line. For presentational simplicity, day of week effects are not included.

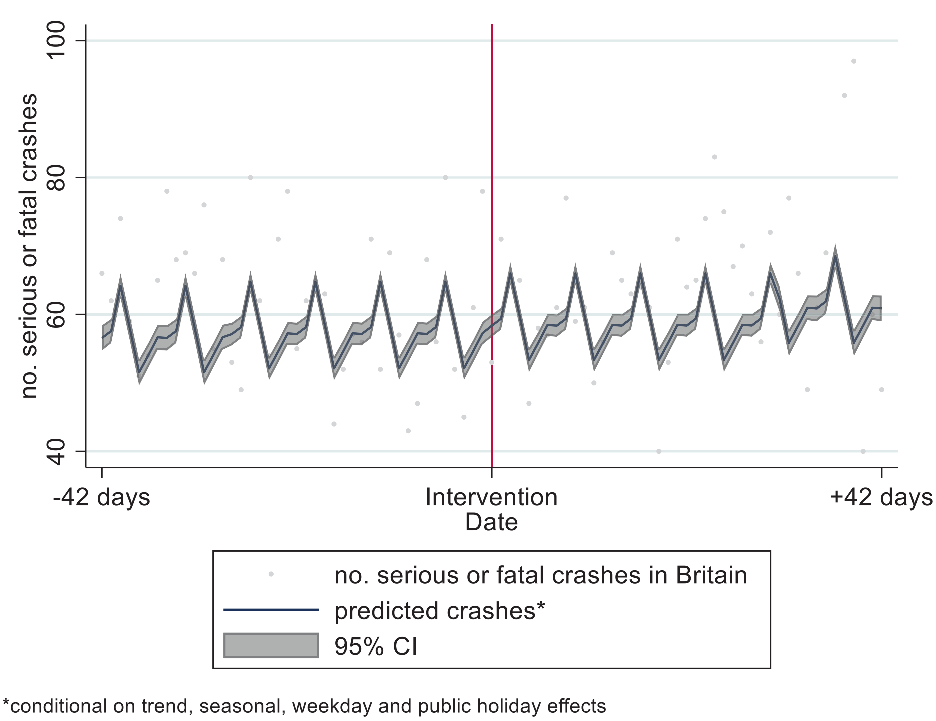


Fig. 2. Serious or fatal crashes and predictions for the main observation window, Britain. Predictions are based on a regression model incorporating a national trend, seasonal, weekday and public holiday dummies over the period 1/1/2009 to 31/12/2018. The chart focuses on 42 days either side of the intervention.

Table 2
RDiT (pooled) modelling results, step 2[a]

	Coefficient (SE)		
Bandwidth	7 days	28 days	42 days
Post	−0.0029 (0.0111)	0.0040 (0.0058)	−0.0029 (0.0048)
Constant	0.0107 (0.0081)	0.0057 (0.0041)	0.0113*** (0.0034)
Adjusted R-squared	−1.64e-04	−2.41e-05	−1.96e-05
n	5,700	21,660	32,300

[a] Dependent variable is the residuals from the first step model. Robust standard errors are shown in parentheses. Step one modelling results are available on request. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

crashes after the intervention, so attention is now turned to the results from the statistical analysis.

The RDiT results in Table 2 show the effects of the mobile phone reform on serious or fatal crashes in Britain (coefficient *Post*) are statistically insignificant and very small over 7, 28, and 42 days post intervention, with longer periods allowing for habit formation over time (Gardner & Lally, 2018; Lally et al., 2010). Results for step one are available on request.[4]

Fig. 3 shows the adjusted numbers of serious or fatal crashes in Britain, with predictions from the RDiT and indicates the effect of the intervention is not statistically significant.

Effects of the intervention may change over time as individuals adapt their driving behavior and crashes respond. Thus far a single average treatment effect has been investigated. To investigate whether there is an adjustment to the intervention, step two of the RDiT is re-run over the 42-day window and the effect of the intervention to vary by each of the 6 weeks post intervention is allowed (Post week 1 – 6, Table 3). By comparing the effects over time, the authors can gauge what type of behavioral change, if any, has occurred. The intervention may have several effects on driver behavior and therefore crashes. After the intervention, when penalties increase, there might be a one-off change in behavior for every-one that permanently lowers crashes but leaves them on the same trend. Alternatively, there might be an increasing adjustment to the new levels of fines as information about the policy change spreads throughout the population and behavior adjusts permanently. Yet again, there might only be a temporary change in behavior, whereby crashes decline immediately following the policy change but then revert to pre-intervention levels, or there may be no change in behavior. These results show no significant effect on serious or fatal crashes for any of the six weeks post-intervention.

With daily serious or fatal crashes by local authority being relatively rare (ranging from 0 to 5), there may be a concern that there is zero effect because at the daily level the crash count is zero and the RDiT model is predominantly estimated across a range of zeros. To address this concern, the authors re-estimate Eqs. (1) and (2) aggregating over space (daily crashes from local authority level to GOR and country level) and time (daily to weekly crashes at local authority level). Table 4 shows results from the step 2 model.

The intervention may have different effects on motorways versus more minor roads if individuals have different expectations about the probability of detection and therefore conviction on such roads. If individuals perceive there is less chance of being caught on minor roads (as there may be less police activity and poorer lighting conditions), there may be a smaller effect on serious or fatal crashes on B and C roads relative to motorways.

To gauge these effects, serious or fatal crashes by road type are considered. While non-motorways are spread across Britain, motorways are few and (often) far between. Table 5 reports results of a DiD analysis comparing road types in England and Wales (motorways and B/C roads) before and after the intervention. Only England and Wales are considered to keep the analyses comparable geographically. Crashes on B and C roads are used as a control group relative to those on motorways as the treatment group, hypothesizing there is greater certainty of detection and therefore punishment for offenses on motorways than on B/C roads. Results show serious or fatal crashes are significantly fewer on motorways than on B/C roads, but that the trends remain parallel after the intervention (no significant difference in the DiD interaction term). Thus the authors cannot say there is a significant effect of the intervention on one road type relative to the other.

5. Discussion

The aim of this paper is to examine the effects of the road safety intervention (comprising doubling of penalties and drivers caught twice or accruing 12 penalty points faced going to court, disqualification, and fines of up to £1000) on number of serious or fatal crashes. Allowing for immediate, intermediate, and longer-term responses to the increased penalties by analyzing results for 7, 28, and 42 days post-intervention, no effect was found on serious or fatal crashes. Since these first results estimate an average effect over the window, the authors also allow for effects to differ within the window by estimating weekly effects. However, no significant effect of the intervention was found. There may be some concern at the number of zeros in the data, so the authors aggregate over space and time but again find no significant effects. Finally, differences by road type are investigated and no effect on the difference-in-difference term was found.

Despite a lack of statistical significance, it is important to consider 'practical significance' in terms of the magnitude of these results. The data show there were 0.16 serious or fatal crashes per LA per day before the intervention. The main results for the 7, 28, and 42 days post intervention periods showed effects on numbers of crashes were in the region of just 2 % for a doubling of penalties (and this effect was *positive* for the 28-day period). To put this in context, with about 62 serious or fatal crashes across Britain on an average day during the 42 days prior to the intervention, the estimated effect would be equivalent to just 1.12 fewer of these types of crashes.

Such a lack of effect was found by Burger et al. (2014) in relation to a ban on mobile use for California. Additionally, no long term effect was found between crashes and texting bans in the United States (Abouk & Adams, 2013). The results of our study are consistent with annual data on serious or fatal crashes where mobile phone use was identified as a contributory factor, which show only a small reduction and the share of crashes involving a mobile phone has remained relatively high over the 2016/2017 period compared to earlier (UK Department for Transport, 2017). A lack of effect on behavior could result if individuals do not consider the £100 increase in fine to be significant (low income elasticity), if they deem the probability of conviction to be low, if the time between violation and conviction is long (Li, Hu, Zhang, Ren, & Liu, 2022), or if they are largely unaware of the change.

There is unlikely to be an information problem, as, at the time of the intervention, 29 million people saw the promotional advertisements and 12 million people saw related content on social media. As a result, 90 % of people were likely aware of the increase in

[4] To test the robustness of our results to the observation window for step one, we re-estimated the 42-day effects based on a 5-year window and found no significant effects (coef. 0.002; SE 0.005; p 0.706). We also estimated a traditional Regression Discontinuity Design (i.e. two steps combined) with numbers of crashes as the dependent variable and still found no significant effect of the intervention (coef. −0.032; SE 2.92; p 0.991).

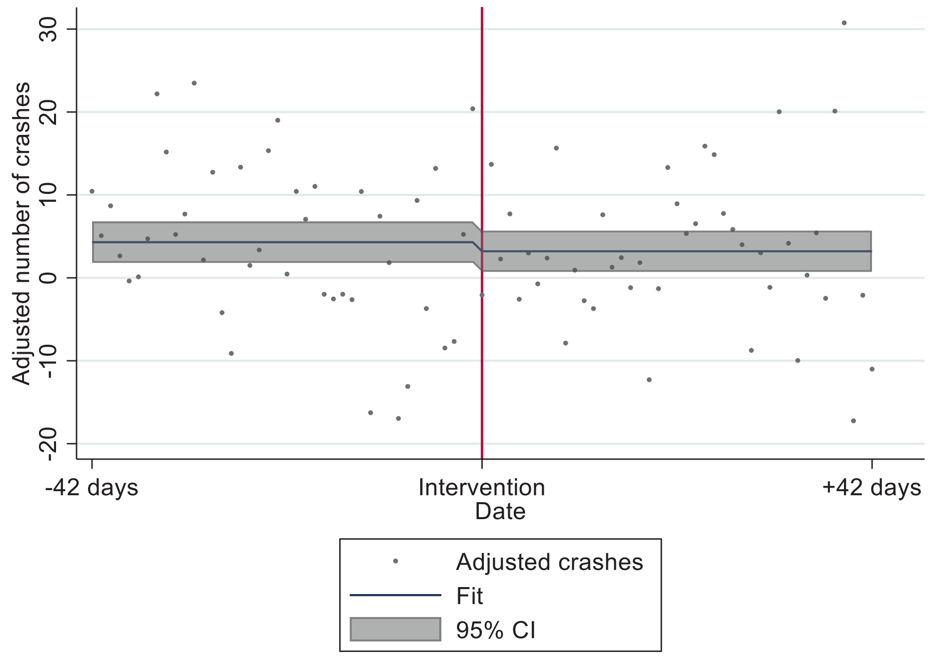


Fig. 3. Adjusted numbers of serious or fatal crashes and RDiT fitted values, Britain. The RDiT sample covers 42 days either side of the intervention (March 1, 2017). Dependent variables are the residuals from the first step models.

Table 3
RDiT (pooled) modelling results allowing intervention effects to vary by week over 42 day bandwidth, step 2[a.]

	Coefficient (SE)
Post week 1	−0.004 (0.008)
Post week 2	−0.007 (0.009)
Post week 3	−0.007 (0.008)
Post week 4	0.011 (0.009)
Post week 5	−0.008 (0.009)
Post week 6	−0.003 (0.009)
Constant	0.011*** (0.003)
Adjusted R-squared	−5.43e-05
n	32,300

[a] Dependent variable is the residuals from the first step model. Post week X is a dummy variable equal to 1 for all days in week X post-intervention and zero otherwise. Robust standard errors are shown in parentheses. Step one modelling results are available on request. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 4
RDiT (pooled) modelling results, aggregated data step 2[a.]

	Coefficient (SE)		
Aggregation	GOR by day	Country by day	LA by week
Post	−0.126 (0.205)	−0.463 (2.828)	−0.008 (0.040)
Constant	0.403*** (0.146)	1.423 (1.991)	0.070** (0.029)
Adjusted R-squared	−6.66E-04	−3.85E-03	−2.11E-04
n	935	255	4560

[a] Dependent variable is the residuals from the first step model, aggregated to the same level as step 2. Robust standard errors are shown in parentheses. Step one modelling results are available on request. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 5
Step 2 DiD (pooled) modelling results, by road type, England and Wales[a.]

Variables	Coefficient (SE)
T (1 = post)	−0.001 (0.002)
G (1 = motorway)	−0.004** (0.002)
DiD	0.002 (0.003)
Constant	0.003** (0.002)

[a] The DiD sample covers 42 days either side of the intervention (March 1, 2017). Dependent variable is the residuals from the first step models. Robust standard errors are shown in parentheses. Adjusted R-squared = 0.000. n = 45,390. Step 1 modelling results are available on request. *** p < 0.01, ** p < 0.05, * p < 0.1.

penalties (https://www.gov.uk/government/news/tens-of-thousands-of-drivers-get-increased-fines-for-using-mobiles-at-wheel). Survey data indicate there was a significant decline in stated mobile use from 2016 to 2017 (UK Royal Automobile Club, 2017). A search of Newsbank newspaper articles and Google trends data also showed there were some articles/searches about the change published prior to the intervention, but that peak interest occurred on the day.

Another factor that may contribute to the effectiveness of the intervention is enforcement activity, which changes the expected penalty that individuals respond to. Data from UK Home Office (2019) suggests overall numbers of constables remained steady about the time of the intervention, so we might expect drivers to also consider enforcement activity to have remained fairly constant and this would lower the expected penalty compared to a situation in which policing activity increased, thus increasing incentives to offend. For England and Wales, fixed penalty notices for using a hand-held mobile while driving were only 53,000 in

2017: 34 % lower than 2016 (UK Department for Transport, 2020). No significant change in serious or fatal crashes would be consistent with deterrence theory in which certainty (related to enforcement), celerity, and severity (linked to penalties) combine to deter individuals from committing crimes (Becker, 1968; Tomlinson, 2016) if there was reduced enforcement offsetting the effects of the change in penalties to leave the expected penalty unchanged. This lack of result would also be consistent with the idea that certainty of punishment is more important than celerity or severity (Tomlinson, 2016).

Moreover, for some individuals, risk preferences may be such that the chances of detection remain too small to change behavior. Enforcement may become easier and less costly in the future with advances in technology substituting for additional police. For example, extensive use of specialized cameras may be able to constantly detect mobile phone usage across Britain. Were this to occur, behavior may change in response to publicity about the technology and associated infringements, as cameras have been shown to be effective in reducing road crashes (Pineda-Jaramillo, Barrera-Jiménez, & Mesa-Arango, 2022). Mobile phone and seat belt cameras were introduced in New South Wales, Australia in 2020 and in Queensland in 2021. These cameras are a mix of fixed and portable and were designed to take images of registration plates and the front seats of the vehicle, using artificial intelligence to detect mobile phone use by the driver (Queensland Government, 2021; Transport for NSW, 2020).

Another potential explanation proposed by Bar-Ilan and Sacerdote (2004) is that an individual's response to a financial penalty might depend on how likely he or she is to comply with the penalty imposed. For some individuals, the expected penalty may be insufficient to deter offending behavior. For example, these individuals may have a very high value of time and therefore be prepared to pay large fines for, say, the convenience of using a mobile phone while driving. Perhaps individuals who offend the most are also those who disregard penalties. Bourgeon and Picard (2007) develop a theoretical model that shows fines are less effective in deterring most drivers from committing driving offenses than are penalty points (leading to license suspension/withdrawal). In particular, some drivers are 'chronically reckless' and do not respond to fines, so incapacitation strategies such as revoking their license (or imprisonment) is the only way to stop such behavior – by keeping them off the roads (Bourgeon & Picard, 2007). This is partly the idea behind increasing penalty points for mobile phone use.

A final explanation is that there is compensatory behavior at play and drivers reduce their speed when using a mobile phone (Papadimitriou et al., 2019) or adopt other self-regulatory behaviors (Kaviani, Young, Robards, & Koppel, 2021), thereby reducing numbers of fatal or serious crashes.

While punishments for mobile phone use may not be effective for various reasons, one solution to the problem could be blocking mobile phone use through an application deployed on smartphones that blocks use after detecting the vehicle is moving. Use of such an application has been shown to be effective among teen drivers in the United States (Creaser, Edwards, Morris, & Donath, 2015).

Despite the rigorous methodology and robust analysis of the best available data, there remain limitations of this study. Firstly, data on enforcement activity (ideally by day and local authority, but at least as a before/after intervention metric) were not available and could have been used to explain our lack of treatment effect if enforcement remained unchanged or declined. Such data could have been in the form of policing levels and/or fines issued. Secondly, the authors were unable to determine if the amount of the fine was too low to have an impact on driving behaviors and survey data on 'willingness to pay' could have been used to assess the likely response of drivers to the level of the penalty.

6. Conclusions

In analyzing the economics of crime, penalties are designed to modify behavior according to the socially optimum outcome. One type of crime that can have significant health impacts via road crashes is behavioral driving offenses, such as using a mobile phone while driving. Using data on crashes reported to police in Britain, this study examined the effects of a penalty intervention on numbers of serious or fatal crashes. By analyzing the link between specific penalties and crashes, the contribution is unique. As the literature considers the effects of penalties on infringement notices or general bans in mobile phone use on crashes, there is no point of comparison for these results.

Although economic theory suggests there should be an impact on driver behavior and road safety literature indicates this should reduce serious crashes, this analysis of serious or fatal crashes in Britain reveals no significant effect of the doubling of fines and penalty points for mobile phone use while driving. The authors rule out an information problem and an enforcement effect. Thus the remaining explanation is that the increase in fines was insufficient to change behavior. This could occur if the perceived certainty of punishment remained very low as mobile phone use is currently relatively easy to do without detection. Future technology will increase the ability to detect mobile phone usage. If so, the solution may be to raise awareness of such technology and publicize numbers of offenders caught, increasing certainty of punishment in line with deterrence theory. This may also be a lower cost option than increasing enforcement through use of additional police. Alternatively, a mobile phone blocking application could avert the problem.

Conflict of interest

None.

CRediT authorship contribution statement

Jane M. Fry: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization.

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

The author would like to thank Lisa Farrell, Michael Shields, David Johnston and two anonymous referees for helpful comments. This research was supported by an Australian Government Research Training Program Scholarship. The Australian Government had no role in the conduct of the research and preparation of the paper. This research has benefitted from feedback received at the Australian Health Economics Society Conference. All remaining errors and omissions are solely the responsibility of the author.

Appendix

See Figs. A1 and A2.

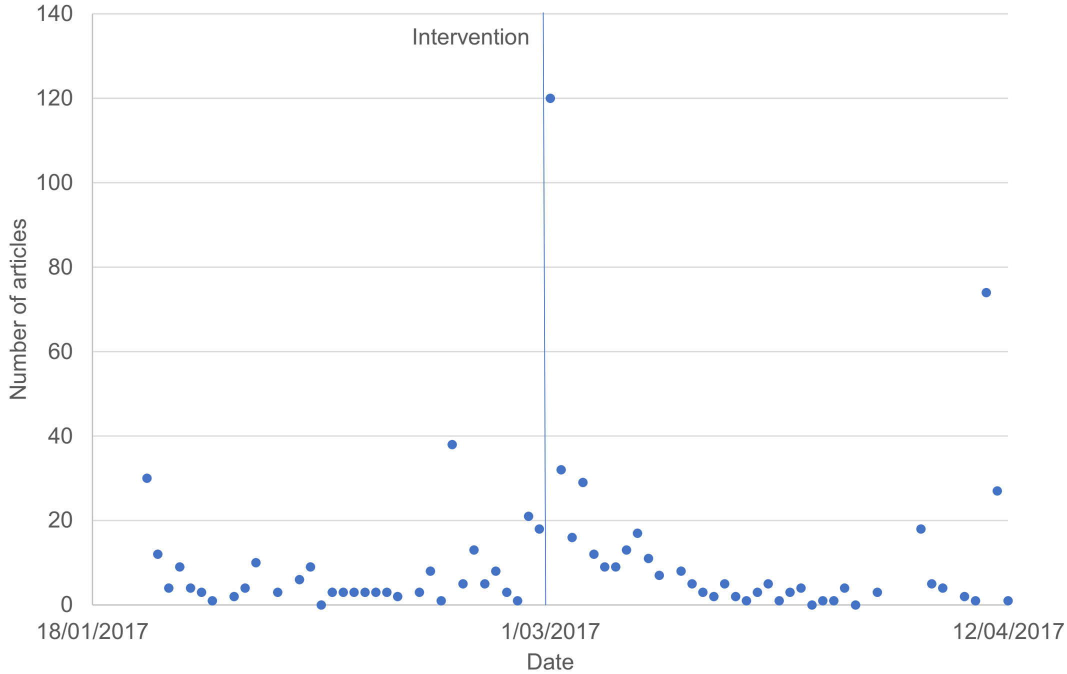


Fig. A1. Newspaper articles mentioning the mobile phone penalty increase. Source: Authors' Newsbank newspapers search.

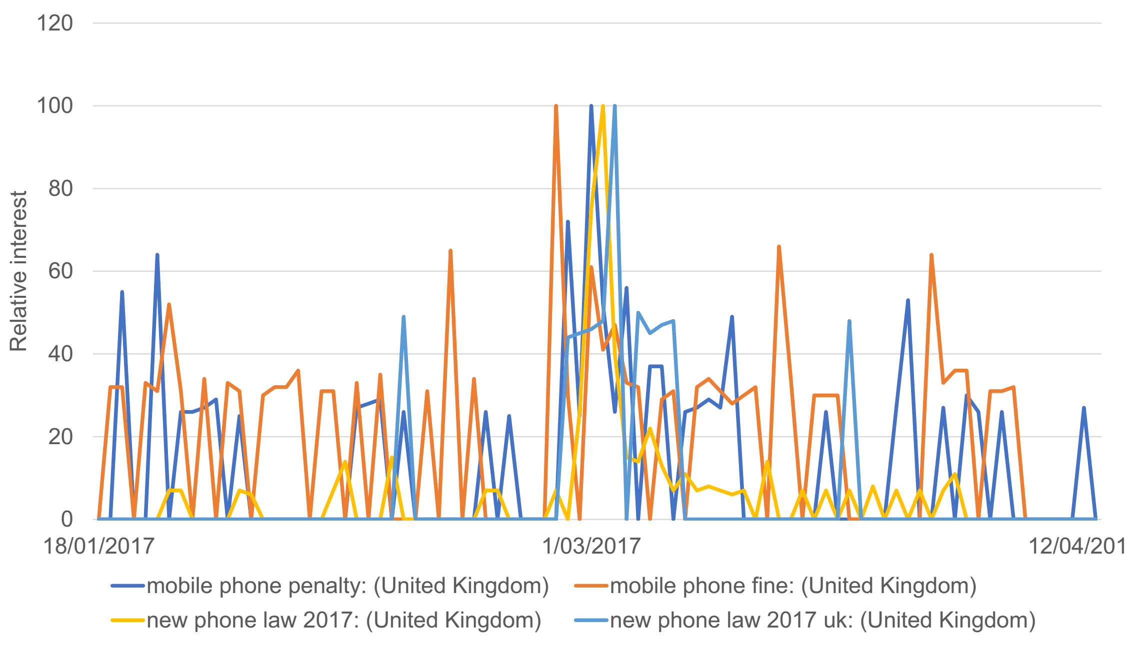


Fig. A2. Google searches relating to increased mobile phone use penalties. Source: Authors' Google trends searches.

References

Abouk, R., & Adams, S. (2013). Texting Bans and Fatal Accidents on Roadways: Do They Work? Or Do Drivers Just React to Announcements of Bans? *American Economic Journal: Applied Economics, 5*(2), 179–199. https://doi.org/10.1257/app.5.2.179.

Anderson, D. M., & Rees, D. I. (2015). Per se drugged driving laws and traffic fatalities. *International Review of Law and Economics, 42*, 122–134. https://doi.org/10.1016/j.irle.2015.02.004.

Backer-Grøndahl, A., & Sagberg, F. (2011). Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. *Safety Science, 49*(2), 324–330. https://doi.org/10.1016/j.ssci.2010.09.009.

Bar-Ilan, A., & Sacerdote, B. (2004). The response of criminals and noncriminals to fines. *The Journal of Law and Economics, 47*(1), 1–17. https://doi.org/10.1086/380471.

Bates, L., Soole, D., & Watson, B. (2012). The Effectiveness of Traffic Policing in Reducing Traffic Crashes. In T. Prenzler (Ed.), *Policing and Security in Practice: Challenges and Achievements* (pp. 90–109). London: Palgrave Macmillan UK.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76*(2), 169–217. https://doi.org/10.1086/259394.

Bourgeon, J.-M., & Picard, P. (2007). Point-record driving licence and road safety: An economic approach. *Journal of Public Economics, 91*(1), 235–258. https://doi.org/10.1016/j.jpubeco.2006.05.007.

Burger, N. E., Kaffine, D. T., & Yu, B. (2014). Did California's hand-held cell phone ban reduce accidents? *Transportation Research Part A: Policy and Practice, 66*, 162–172. https://doi.org/10.1016/j.tra.2014.05.008.

Cardamone, A. S., Eboli, L., Forciniti, C., & Mazzulla, G. (2017). How usual behavior can affect perceived drivers' psychological state while driving. *Transport, 32*(1), 13–22. https://doi.org/10.3846/16484142.2015.1059885.

Castriota, S., & Tonin, M. (2019). *Stay or flee? Probability versus severity of punishment in hit-and-run accidents*. IZA Discussion Paper No. 12693. IZA. Bonn.

Chang, H., Chang, K., & Fan, E. (2020). The Intended and Unintended Effects of Drunk Driving Policies. *Oxford Bulletin of Economics and Statistics, 82*(1), 23–49. https://doi.org/10.1111/obes.12326.

Cooper, B., Gehrsitz, M., & McIntyre, S. G. (2018). *Drink, death and driving: Do BAC limit reductions improve road safety?* Glasgow: Strathclyde Discussion Papers in Economics. University of Strathclyde.

Creaser, J. I., Edwards, C. J., Morris, N. L., & Donath, M. (2015). Are cellular phone blocking applications effective for novice teen drivers? *Journal of Safety Research, 54*, 75.e29-78. doi:https://doi.org/10.1016/j.jsr.2015.06.014.

De Paola, M., Scoppa, V., & Falcone, M. (2013). The deterrent effects of the penalty points system for driving offences: A regression discontinuity approach. *Empirical Economics, 45*(2), 965–985. https://doi.org/10.1007/s00181-012-0642-9.

DeAngelo, G., & Hansen, B. (2014). Life and death in the fast lane: Police enforcement and traffic fatalities. *American Economic Journal: Economic Policy, 6*(2), 231–257. https://doi.org/10.1257/pol.6.2.231.

Dionne, G., Fluet, C., & Desjardins, D. (2007). Predicted risk perception and risk-taking behavior: The case of impaired driving. *Journal of Risk and Uncertainty, 35* (3), 237–264. https://doi.org/10.1007/s11166-007-9023-8.

Donmez, B., & Liu, Z. (2015). Associations of distraction involvement and age with driver injury severities. *Journal of Safety Research, 52*, 23–28. https://doi.org/10.1016/j.jsr.2014.12.001.

Elvik, R. (2011). Effects of mobile phone use on accident risk: Problems of meta-analysis when studies are few and bad. *Transportation Research Record, 2236*(1), 20–26. https://doi.org/10.3141/2236-03.

Farmer, C. M., Braitman, K. A., & Lund, A. K. (2010). Cell phone use while driving and attributable crash risk. *Traffic Injury Prevention, 11*(5), 466–470. https://doi.org/10.1080/15389588.2010.494191.

Ferdinand, A. O., Menachemi, N., Sen, B., Blackburn, J. L., Morrisey, M., & Nelson, L. (2014). Impact of texting laws on motor vehicular fatalities in the United States. *American Journal of Public Health, 104*(8), 1370–1377. https://doi.org/10.2105/AJPH.2014.301894.

Fry, J. M., & Farrell, L. (2022). Road accidents: Unexpected costs of stock market movements. *Oxford Economic Papers, gpac016*. https://doi.org/10.1093/oep/gpac016.

Gardner, B., & Lally, P. (2018). Modelling Habit Formation and Its Determinants. In B. Verplanken (Ed.), *The Psychology of Habit: Theory, Mechanisms, Change, and Contexts* (pp. 207–229). Cham: Springer International Publishing.

Gehrsitz, M. (2017). Speeding, punishment, and recidivism: Evidence from a Regression Discontinuity Design. *The Journal of Law and Economics, 60*(3), 497–528. https://doi.org/10.1086/694844.

Hansen, B. (2015). Punishment and deterrence: Evidence from drunk driving. *The American Economic Review, 105*(4), 1581–1617. https://doi.org/10.1257/aer.20130189.

Hausman, C., & Rapson, D. S. (2018). Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics, 10*(1), 533–552. https://doi.org/10.1146/annurev-resource-121517-033306.

Hersh, J., Lang, B. J., & Lang, M. (2022). Car accidents, smartphone adoption and 3G coverage. *Journal of Economic Behavior & Organization, 196*, 278–293. https://doi.org/10.1016/j.jebo.2022.01.033.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5.

Kantorowicz-Reznichenko, E. (2015). Day-fines: Should the rich pay more? *Review of Law & Economics, 11*(3), 481–501. https://doi.org/10.1515/rle-2014-0045.

Kaviani, F., Young, K. L., Robards, B., & Koppel, S. (2021). "Like it's wrong, but it's not that wrong:" Exploring the normalization of risk-compensatory strategies among young drivers engaging in illegal smartphone use. *Journal of Safety Research, 78*, 292–302. https://doi.org/10.1016/j.jsr.2021.06.010.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine, 370*(1), 54–59. https://doi.org/10.1056/NEJMsa1204142.

Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology, 40*(6), 998–1009. https://doi.org/10.1002/ejsp.674.

Li, H., Hu, H., Zhang, Z., Ren, G., & Liu, X. (2022). Impacts of enforcement cameras on pedestrians' risk perception and drivers' behaviors at non-signalized crosswalks. *Journal of Safety Research, 81*, 313–325. https://doi.org/10.1016/j.jsr.2022.03.008.

Luca, D. L. (2015). Do traffic tickets reduce motor vehicle accidents? Evidence from a natural experiment. *Journal of Policy Analysis & Management, 34*(1), 85–106. https://doi.org/10.1002/pam.21798.

McEvoy, S. P., Stevenson, M. R., & Woodward, M. (2007). The contribution of passengers versus mobile phone use to motor vehicle crashes resulting in hospital attendance by the driver. *Accident Analysis & Prevention, 39*(6), 1170–1176. https://doi.org/10.1016/j.aap.2007.03.004.

Montag, J. (2014). A radical change in traffic law: Effects on fatalities in the Czech Republic. *Journal of Public Health, 36*(4), 539–545. https://doi.org/10.1093/pubmed/fdu005.

Papadimitriou, E., Argyropoulou, A., Tselentis, D. I., & Yannis, G. (2019). Analysis of driver behavior through smartphone data: The case of mobile phone use while driving. *Safety Science, 119*, 91–97. https://doi.org/10.1016/j.ssci.2019.05.059.

Pineda-Jaramillo, J., Barrera-Jiménez, H., & Mesa-Arango, R. (2022). Unveiling the relevance of traffic enforcement cameras on the severity of vehicle–pedestrian collisions in an urban environment with machine learning models. *Journal of Safety Research, 81*, 225–238. https://doi.org/10.1016/j.jsr.2022.02.014.

Queensland Government. (2021). Mobile phone and seatbelt cameras. Retrieved from https://www.qld.gov.au/transport/safety/fines/cameras.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine, 336*(7), 453–458. https://doi.org/10.1056/NEJM199702133360701.

Rocco, L., & Sampaio, B. (2016). Are handheld cell phone and texting bans really effective in reducing fatalities? *Empirical Economics, 51*(2), 853–876. https://doi.org/10.1007/s00181-015-1018-8.

Stafford, M. C., & Warr, M. (1993). A Reconceptualization of General and Specific Deterrence. *Journal of Research in Crime and Delinquency, 30*(2), 123–135. https://doi.org/10.1177/0022427893030002001.

Tomlinson, K. D. (2016). An Examination of Deterrence Theory: Where Do We Stand Retrieved from. *Federal Probation, 80*(3), 33–38 https://heinonline.org/HOL/P?h=hein.journals/fedpro80&i=1.

Tran, N. T., Hyder, A. A., Kulanthayan, S., Singh, S., & Umar, R. S. R. (2009). Engaging policy makers in road safety research in Malaysia: A theoretical and contextual analysis. *Health Policy, 90*(1), 58–65. https://doi.org/10.1016/j.healthpol.2008.08.009.

Transport for NSW. (2020). Mobile phone detection cameras. Retrieved from https://roadsafety.transport.nsw.gov.au/stayingsafe/mobilephones/technology.html.

Traxler, C., Westermaier, F. G., & Wohlschlegel, A. (2018). Bunching on the Autobahn? Speeding responses to a 'notched' penalty scheme. *Journal of Public Economics, 157*, 78–94. https://doi.org/10.1016/j.jpubeco.2017.11.006.

UK Department for Transport (2016a). *A consultation on changes to the Fixed Penalty Notice and penalty points for the use of a hand-held mobile phone whilst driving*. London: Department for Transport.

UK Department for Transport (2016b). *Reported road casualties Great Britain 2015: Annual report*. London: Department for Transport.

UK Department for Transport. (2017). *Table RAS50001 Contributory factors in reported accidents by severity, Great Britain, 2016*. Retrieved from: https://www.gov.uk/government/statistical-data-sets/ras50-contributory-factors.

UK Department for Transport (2018). *Reported road casualties Great Britain: 2017 annual report*. London: Department for Transport.

UK Department for Transport. (2019). *Road Safety Data*. Retrieved from: https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data.

UK Department for Transport. (2020). *Police powers and procedures, England and Wales, year ending 31 March 2020 second edition*. Retrieved from: https://www.gov.uk/government/statistics/police-powers-and-procedures-england-and-wales-year-ending-31-march-2020.

UK Home Office. (2019). *Table S1 Police workforce, excluding those on career breaks or maternity/paternity leave (comparable with previously published figures)*, 31 March 1996 to 31 March 2019. Retrieved from: https://www.gov.uk/government/statistics/police-workforce-england-and-wales-31-march-2019.

UK Met Office. (2018). *Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current)*. . Retrieved from: http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0.

UK Ministry of Justice. (2018). *Table A6.2 - Offenders found guilty at all courts, by motoring offence, 2008 to 2018*. Retrieved from: https://www.gov.uk/government/statistics/criminal-justice-system-statistics-quarterly-december-2018.

UK Royal Automobile Club (2017). *RAC Report on Motoring 2017*. Walsall: RAC.

Winston, C., Maheshri, V., & Mannering, F. (2006). An exploration of the offset hypothesis using disaggregate data: The case of airbags and antilock brakes. *Journal of Risk and Uncertainty, 32*(2), 83–99. https://doi.org/10.1007/s11166-006-8288-7.

World Health Organization (2018). *Global status report on road safety 2018*. Geneva: WHO.

Jane M. Fry Jane is a postdoctoral research fellow in the School of Population and Global Health at the University of Melbourne. She is a researcher with over 30 years of experience in the private sector, public sector and academia in Australia. Jane has conducted applied research in various areas of economics, econometrics and marketing, writing reports and publishing in several high ranking academic journals. Her research interests lie broadly in ageing, health, applied microeconomics and data analysis. Jane has a Bachelor of Economics (Hons.), Masters of Economics and a PhD (Health Economics) from Monash University.

Journal of Safety Research 84 (2023) 41–60

Contents lists available at ScienceDirect

Journal of Safety Research

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

One year of COVID-19: Impacts on safe driving behavior and policy recommendations

Eva Michelaraki *, Marios Sekadakis, Christos Katrakazas, Apostolos Ziakopoulos, George Yannis

National Technical University of Athens, Department of Transportation Planning and Engineering, 5 Heroon Polytechniou str, GR-15773 Athens, Greece

ARTICLE INFO

Article history:
Received 23 November 2021
Received in revised form 4 July 2022
Accepted 17 October 2022
Available online 24 October 2022

Keywords:
COVID-19
Response measures
Driving behavior
Road crashes
k-Means clustering

ABSTRACT

Introduction: In the unprecedented year of 2020, the rapid spread of COVID-19 disrupted everyday activities worldwide, leading the majority of countries to impose lockdowns and confine citizens in order to minimize the exponential increase in cases and casualties. To date, very few studies have been concerned with the effect of the pandemic on driving behavior and road safety, and usually explore data from a limited time span. *Method:* This study presents a descriptive overview of several driving behavior indicators as well as road crash data in correlation with the strictness of response measures in Greece and the Kingdom of Saudi Arabia (KSA). A k-means clustering approach was also employed to detect meaningful patterns. *Results:* Results indicated that during the lockdown periods, speeds were increased by up to 6%, while harsh events were increased by about 35% in the two countries, compared to the period after the confinement. However, the imposition of another lockdown did not cause radical changes in Greek driving behavior during the late months of 2020. Finally, the clustering algorithm identified a "baseline," a "restrictions," and a "lockdown" driving behavior cluster, and it was shown that harsh braking frequency was the most distinctive factor. *Policy recommendations:* Based on these findings, policymakers should focus on the reduction and enforcement of speed limits, especially within urban areas, as well as the incorporation of active travelers in the current transport infrastructure.

1. Introduction

In 2020, the COVID-19 pandemic dominated every aspect of life globally by infecting around 100 million individuals, leading to more than 2 million casualties (Dong et al., 2020). When the spread of COVID-19 started increasing around the world, the majority of governments imposed lockdowns as means of restricting non-essential civilian movements, while all recreational, religious, cultural, dining, and entertainment establishments were instructed to cease operations.

With typical social activities interrupted, the transportation ecosystem was disturbed as well. Recent studies have been mainly focused on the effect of COVID-19 on travel behavior (e.g., Barbieri et al., 2020; De Vos, 2020) but also on air travel operations (Hotle & Mumbower, 2021) and shared mobility (Padmanabhan et al., 2021). As can be anticipated, road traffic volumes were heavily reduced (De Vos, 2020; Vingilis et al., 2020). This decline in traffic volumes has led to higher speeds and more frequent harsh events (Katrakazas et al., 2020) while large reductions in crashes have been recorded (Aloi et al., 2020; Katrakazas et al., 2020; Saladié et al., 2020).

During the year 2020, COVID-19 response measures varied from country to country according to the fluctuation of the number of cases and patients in the available Intensive Care Units (ICUs). The harshness of the response measures has been captured by the stringency index introduced by Hale et al. (2020), which explores information on 19 indicators of COVID-19 government responses and corresponds to the strictness of government policies on the matter. Nevertheless, to date the strictness of government policy on COVID-19 response measures has not yet been correlated with driving behavior during the pandemic. Furthermore, an overall limited number of studies have been concerned with road safety and driving behavior during the pandemic, and the majority of those studies explore a limited timespan for data collection.

The aforementioned reasons form the motivation for the current paper, which aims at providing a detailed overview of how COVID-19 affected road safety indicators in Greece and the Kingdom of Saudi Arabia (KSA), while accounting for the strictness of COVID-19 countermeasures. Although this paper does not employ advanced statistical approaches to fulfil its aim, it is the first of its kind to provide a detailed overview of highly disaggregated naturalistic driving behavior data and provides result-based policy rec-

* Corresponding author.
E-mail address: evamich@mail.ntua.gr (E. Michelaraki).

https://doi.org/10.1016/j.jsr.2022.10.007
0022-4375/

ommendations. The rationale behind the two countries comes up to the fact of naturalistic driving data provision.

In order to fulfill this aim, a descriptive exploration of 12 months of data regarding several driving behavior indicators (i.e., average speed, speeding percentage, average driving speed, harsh accelerations/100 km, harsh brakings/100 km, total duration, total driven distance, and mobile phone usage duration/driving duration) is initially presented to understand the effect of the pandemic on driving during 2020. It is worth noting that harsh acceleration refers to a driver event where more force than normal is applied to the vehicle's accelerator system. The term acceleration used in the manuscript does not refer to lateral acceleration (hard cornering), as the latter variable acts transversely to the direction of travel of a car. Thus, particular emphasis was given to the harsh accelerator variable as it can be an indicator of aggressive or unsafe driving behavior. The exploratory analysis is supplemented by an unsupervised pattern recognition algorithm, aiming at identifying clusters of weeks according to driving behavior and the strictness of COVID-19 response measures. Following the effect of the pandemic on driving behavior and road safety, policy recommendations are discussed in order to pave the way for post-pandemic safer roads.

The paper is structured as follows: initially, the literature with regards to driving behavior and road safety during the pandemic is reviewed. This is followed by an overview of the data needed for the exploratory analysis. The main part of this paper is dedicated to depicting the changes in driving behavior during 2020 and is followed by a section on weekly pattern identification. Finally, the results are discussed and helpful conclusions for researchers and policymakers are provided.

2. Literature review

To be able to provide an overview of the effect of COVID-19 in Greece and the KSA, the literature was reviewed for studies correlating the pandemic and driving behavior or road safety. The search took place in the databases Google Scholar and Scopus using the Boolean terms {"COVID-19" or "Pandemic" and "driving behavior" or "driving behavior" or "road safety"}.

From the list of examined papers, it was observed that the majority of studies on the effect of COVID-19 on transportation were concerned with changes in travel behavior and mode choice (e.g., Bhaduri et al., 2020; De Vos, 2020; Jenelius & Cebecauer, 2020; Parady et al., 2020; Shamshiripour et al., 2020). Only 10 of the studies were concerned with the road safety effects of the pandemic and were chosen to be further reviewed for the purposes of this research. The retrieved papers can be divided into three categories: (a) the ones focusing on epidemiological models and analyzing road safety as yet another health consequence of the pandemic; (b) the ones providing descriptive evidence of the effect of the pandemic; and (c) the ones utilizing advances statistical tools to investigate crucial indicators and explain the impact of COVID-19 on injuries, road crashes, and driving behavior.

With regards to epidemiological studies, a compensation effect between damage from epidemic deaths and road crashes-related deaths in Italy was examined (Colonna & Intini, 2020). It was demonstrated that damage from loss of human capital and health care costs could have been fulfilled if a lockdown was imposed 10 days earlier. Similarly, Lemke et al. (2020) promote syndemic (i.e., population-level clustering of social and health problems, as per Singer et al., 2017) frameworks for the evaluation of commercial driver stress, health, and safety; but their work is limited to a theoretical discussion on advantages of such frameworks and potential enhancements these may offer in safety assessment during the pandemic.

Descriptive results are presented in Saladié et al. (2020), where the reduction in road crashes in the province of Tarragona was presented by comparing the frequency of crashes and checking statistical significance using a chi-square test on weekdays and weekends as well as different road types. A large reduction in crashes (74% compared to February of 2020; 76% compared to 2019) was observed and was associated with the overall reduction of traffic volumes. Likewise, Katrakazas et al. (2020) provided descriptive evidence from Greece and the KSA with regards to COVID-19 and driving behavior. It was observed that when a lockdown was imposed, a slight increase by 6–11% in average driving speed was observed, while harsh accelerations and brakings per 100 km were more frequent by up to 12% when compared to normal operations. Nevertheless, the results presented in the aforementioned studies were purely descriptive, without significant statistical analyses.

To date, only a few studies have conducted statistical analyses with regards to the effect of COVID-19 on driving behavior both using simple as well as more sophisticated models. These can be further distinguished between simple modeling and hypotheses testing and time-series regression modeling of the effect of the pandemic. For instance, Prasetijo et al. (2021) used a simple linear fit speed model and underlined the importance of road design to incorporate sudden changes in traffic volumes with regards to safety. On the same principle, using crowdsourced cycling data from July 2019 to March 2020, Hong and Mcarthur (2020) employed a simple linear regression model, but mixed results with regards to the safety of cyclists are presented.

A more sophisticated approach is presented in Qureshi et al. (2020), where a two-sample t-test was utilized to identify differences in road crashes before and after a lockdown in the United States, as well as ARIMA modeling for autocorrelation and trend analysis was implemented. The reduction of road crashes compared to non-serious or injuries was found significant, but more complex analyses would shed light on the influencing factors. Stavrinos et al. (2020), using multi-level modeling, demonstrated that after the appearance of COVID-19 driving days per week decreased by 37%, while vehicle miles driven dropped by 35. Similar results were presented by Roe et al. (2020) using within-subjects general linear models on a sample of elderly drivers. It was demonstrated that driving days as well as frequency of speeding were reduced. Finally, a full time-series modeling approach was employed by Inada et al. (2020). Using a seasonal ARIMA model and data from January to May 2020, the authors concluded that the lockdown was the crucial factor for speed-related traffic violations, which consequently led to an increase of fatal road crashes. It was further revealed that speeding increased by 52% in March 2020 compared to March 2019.

Moreover, Sekadakis et al. (2021) analyzed the impact of COVID-19 on the total number of road safety figures using time-series forecasting in Greece. It was found that road collisions, fatalities, and slightly injuries were decreased, mainly due to the remarkable decrease of traffic volumes. Similarly, Seasonal AutoRegressive Integrated Moving Average (SARIMA) and XGBoost algorithms were implemented in order to identify the impact of the COVID-19 on driving performance (Katrakazas et al., 2021). Results revealed that average speed increased by 2.27 km/h on average compared to the forecasted evolution, while harsh brakings per distance (i.e. 100 km) increased to almost 1.51 on average. Interestingly, road crashes in Greece were reduced by 49% during the months of COVID-19 in comparison to the non-COVID-19 period. Another study was conducted aiming to provide a comparative overview of the impact of COVID-19 on traffic safety behavior (Michelaraki et al., 2021). It was revealed that speeding percentage, average speed, and harsh accelerations or brakings increased during the lockdown period. Lastly, a reduction in traffic volumes (i.e., people driving and walking) was also observed.

From the aforementioned literature findings, it is evident that no study has yet presented an overview of naturalistic driving data throughout 2020, and the impact of the strictness of response measures as well as machine learning approaches are yet to be utilized to investigate patterns correlating the pandemic with driving behavior. As a result, the current paper is an attempt to fill this particular research gap.

3. Data overview

To present a descriptive overview of the impact of the pandemic within 2020, four types of data are utilized both for Greece and the KSA:

- COVID-19 data on cases and casualties
- Governmental response measures
- Naturalistic driving data captured from novel smartphone apps by OSeven Telematics
- Traffic exposure data

These data are further overviewed in the following sections. It should be noted that Greece and the KSA were chosen compared to others, since only these countries had the appropriate sample size for further investigation. In particular, trip data were collected from a specific subset of the population of Greece and the KSA (i.e., users of OSeven mobile phone application) and additional information (e.g., gender, age, educational level) was not provided due to the anonymity of the drivers. No examination or analysis based on any demographic or personal characteristics of the examined sample was possible due to standing Greek and European data protection legislation (GDPR). As a consequence, this study retains a scope of macroscopic examination of driver behavior, considering the trips produced by the drivers collectively.

3.1. COVID-19 cases and casualties

Data on COVID-19 confirmed cases and casualties were retrieved from the corresponding ministries of health and were cross-checked with press releases and popular websites (e.g., Worldometer, 2020). The evolution of COVID-19 cases and casualties in the two countries are presented in Fig. A1 in the Appendix.

3.2. COVID-19 response measures

Regarding COVID-19 response measures apart from governmental press releases, the government response tracker for COVID-19, put together by the University of Oxford (Hale et al., 2020) was reviewed in order to obtain a homogenized set of validated response measures for Greece and the KSA. Table A1 gives an overview of the response measures milestones for the two countries, while Fig. A2 provides a timeline of the evolution of the response measures stringency index (i.e., the strictness of the measures), as shown in the Appendix.

With respect to Greece, the first lockdown of non-essential movements refers to the period between 23/03/2020 and 04/05/2020. Then, after a 42-day lockdown, Greece began to gradually lift restrictions on movement and restart business activities. The second lockdown of restrictions refers to the period between 07/11/2020 and 31/12/2020, when Greece put in place new measures on movements. Thus, the 6-month framework, between 04/05/20 and 07/11/20, refers to the period between the first and the second lockdown. KSA announced a lockdown of non-essential movements along with the suspension of all domestic and international travel on March 26. After the aforementioned restrictions took place, the number of daily confirmed cases shrunk dramatically and by June 21, all curfews were lifted. By the end of 2020, the KSA was seeing more daily recoveries than cases.

3.3. Naturalistic driving behavior data

For the purpose of the current research, OSeven has provided trip data from its database for Greece and the KSA for a 12-month timeframe from 01/01/2020 to 31/12/2020.The provided dataset corresponds to the same set of random users so that the data before and after the COVID-19 crisis are fully comparable. It should be noted that OSeven Telematics (oseven.io) uses a smartphone application and a platform in order to explore data from smartphone sensors (e.g., GPS, accelerometer, and gyroscope data). State-of-the-art technology and algorithms, reliable metrics, and novel gamification schemes are used in order to help drivers understand their weak points and improve themselves. For each trip, a vast quantity of data were collected and communicated through Wi-Fi or cellular network, and valuable critical information such as features, highlights, and driving scores was generated in order to assess the driver's profile and performance. Data were then transferred to the OSeven backend infrastructure, where it was analyzed with filtering, signal processing, machine learning algorithms, and safety/eco rating models. The final outcome of the analysis is risk-related driving events such as speeding, mobile use, harsh accelerations, and harsh brakings, as well as safety/and eco scores.

A standard procedure is followed every time a new trip is retrieved by the application: the application collects in real-time the data from the sensors of the mobile phone and then data processing takes place. All the variables in the analyzed data were derived from a combination of machine learning methods (data fusion, clustering & classification). Since OSeven has strict data sharing policies, further information cannot be provided at the moment. Nevertheless, additional details for data extraction regarding the OSeven application can be found in Papadimitriou et al. (2019) and Kontaxi et al. (2021).

It should be clearly mentioned that the OSeven platform is able to detect different driving patterns as well as recognize whether the user is a driver or a passenger. Undoubtedly, drivers have their own driving patterns, strongly affected by their personality and daily routine and the recorded driving behavior is totally different when they are passengers instead of drivers. To that end, OSeven developed a set of machine learning algorithms that can reliably determine if the user is the driver or a passenger, taking into consideration all the above parameters. It is worth noting that this specific driver/passenger recognition achieves over 92% accuracy. However, in case of false alarms, users are able to confirm through the app if there were drivers or passengers. For instance, if a person who had the app on their phone was a passenger, driving data from another driver during the specific trip were not collected by the app and thus, these data were not included in the analyses. This is also explicitly stated in the OSeven terms of use, and the drivers understand that their smartphone becomes a driving recording device when driving, thus becoming more mindful of it and limiting false recording.

The OSeven application has been utilized for road safety research, as described in several studies (Stavrakaki et al., 2019; Tselentis et al., 2019; Yannis et al., 2017). A similar approach is followed in the present paper. A large amount of data were recorded using the aforementioned state-of-the-art platform, as described in recent research utilized this specific scheme (Papadimitriou et al., 2019). For instance, harsh events (i.e., harsh accelerations and harsh brakings) are calculated via machine learning algorithms and data fusion. There is not a rule-based approach, using as input

Table 2
Descriptive statistics for the available indicators of the entire database in Greece and the KSA.

Indicator	Greece (268,549 Trips)					KSA (448,736 Trips)				
	Median	St. dev.	Max	Min	IQR	Median	St. dev.	Max	Min	IQR
Average speeding (km/h)	13.22	11.3	154.47	0.00	20.27	16.48	13.02	128.16	0.00	23.23
Speeding percentage	1.34%	11.06%	98.76%	0%	8.71%	2.14%	13.16%	100%	0%	10.79%
Average driving speed (km/h)	41.02	18.20	178.36	5.57	22.89	53.04	20.68	178.93	1.92	26.42
Harsh accelerations/100 km	0.00	22.04	299.40	0.00	14.44	0.00	16.79	282.92	0.00	12.27
Harsh brakings/100 km	6.33	22.89	293.08	0.00	23.05	11.94	23.83	299.04	0.00	27.20
Total duration (sec)	870	1153.17	25,549	300	908	1020	1580.27	36,138	300	1194
Total driven distance (km)	7.60	24.81	648.69	0.50	11.39	12.60	36.09	1006.69	0.61	19.65
Mobile phone usage duration/driving duration	0%	14.61%	99.86%	0%	1.74%	1.59%	19.76%	99.83%	0%	15.76%
Stringency index	53.3	27.3	84.3	0	32.4	55.9	27.6	94.4	0.0	21.3

the values of the accelerometer or values from additional sensors (e.g., GPS, orientation, gyroscope). In addition, the outputs of the OSeven algorithms have been evaluated both in the published studies and used by major insurance companies in several countries (e.g., Greece, UK, Brazil, Qatar); this serves as evidence with regards the acceptance of the proposed algorithms implemented.

It is worth mentioning that OSeven follows strict information security procedures and privacy policies, which comply with the General Data Protection Regulation (GDPR) and related European Union directives. Therefore, all data have been provided by OSeven in a completely anonymized format and no geolocation information for the trips (apart from the related country) have been included in the dataset. A similar dataset was utilized in previous analysis by the authors (Katrakazas et al., 2020; 2021). Privacy policy statements cover the type of data that are collected, the reason they were collected, the time that they are stored, and the measures that they have been taken to protect them based on encryption standards for data in transit and at rest. OSeven technology has already been accepted and approved by several national authorities and compliance officers of multinational brands and it complies with the national regulation in EU and around the world.

What is more, details on the data collected, the purpose of the collection as well as information on the storage and retention of data, are explicitly stated in compliance with the GDPR. In addition, OSeven is also audited for ISO 27001 by TÜV Hellas. The ISO certification verifies the focus on meeting the highest security and privacy standards by auditing and constantly improving their policies, systems and procedures. Thus, customers can be reassured that their data is treated within OSeven Platform with integrity and confidentiality. It should be also mentioned that OSeven is dedicated into maintaining high-security standards in the design, implementation and delivery of its services and products and this is an iterative approach subject to annual assessment.

It becomes evident that OSeven is compliant with international and European privacy and security standards. As it was previously mentioned, the data that have been used are fully anonymized and their recording has been approved by users of the app through the terms and conditions of the company. Moreover, since data from the OSeven platform are high-level and aggregated, it was assumed that they do not violate ethical concerns, since their use has already been approved for use in several peer-reviewed publications (e.g., Papadimitriou et al., 2019; Yannis et al., 2017; Tselentis et al., 2019; Stavrakaki et al., 2019; Kontaxi et al., 2021).

Overall, the authors state that the current study complied with the Declaration of Helsinki's ethical principles because no one was harmed, physically or emotionally, during the driving measurements, and because all of the drivers participated voluntarily. The OSeven application is open-access and has no impact on drivers when driving. The OSeven application aims to improve eco-driving and road safety while tracking and evaluating the driver's performance. The director of the Department of Transportation Planning and Engineering of the School of Civil Engineering at the National Technical University of Athens also gave his approval to the ethics rules.

The driving indicators included in the analysis are presented in Table 1.

Table 2 presents the descriptive statistics (i.e., median, standard deviation, max, min, interquartile range-IQR) with regards to the entire database in both countries. The subset of trips provided by OSeven for the aforementioned time framework included approximately 268,549 trips in Greece and 448,736 trips in the KSA. It was revealed that driving performance indicators (i.e., speeding percentage, harsh brakings per 100 km, total/driving duration) for the KSA appeared to have higher values compared to the corresponding parameters for Greece.

In order to have an initial depiction of changes happening to these indicators during the evolution of the COVID-19 pandemic, Table 3 provides descriptive statistics for the lockdown periods in both countries. It is evident that during the lockdown periods, the total number of trips was much lower compared to the period under normal circumstances. With regards to Greece, it was demonstrated that during the first lockdown period, values for the majority of available indicators were higher compared to the

Table 1
Description of the driving indicators of the analysis *(Source: OSeven)*.

Indicator	Unit	Description
Total duration	sec	Total trip duration
Total distance	km	Total trip distance
Harsh accelerations/100 km	-	Number of harsh accelerations per distance (i.e. 100 km)
Harsh brakings/100 km	-	Number of harsh brakings per distance (i.e. 100 km)
Speeding duration	sec	Total duration of speeding in a trip
Average speeding	km/h	Average speed over the speed limit
Average driving speed	km/h	Average speed during driving with stops been excluded from the duration of the trip
Mobile phone usage duration	sec	Total duration of mobile usage
Speeding percentage	%	Ratio of speeding duration in a trip per total duration of driving
Mobile phone usage duration /driving duration	%	Ratio of total duration of mobile usage per total duration of driving

Table 3
Descriptive statistics for the available indicators of the first and second lockdown in Greece and the KSA.

Indicator	Greece – 1st lockdown (14,649 Trips)					Greece – 2nd lockdown (36,390 Trips)					KSA (86,967 Trips)				
	Median	Stdev	Max	Min	IQR	Median	Stdev	Max	Min	IQR	Median	Stdev	Max	Min	IQR
Average speeding (km/h)	14.06	12.09	103.47	0.00	21.39	13.51	11.02	119.17	0.00	20.08	17.02	13.60	128.16	0.00	24.16
Speeding percentage	1.55%	12.92%	90.56%	0%	10.7%	1.74%	11.70%	96.72%	0%	10.24%	2.76%	15.04%	100%	0%	13.56%
Average driving speed (km/h)	45.30	19.16	153.79	11.19	25.04	41.90	17.12	164.25	6.29	22.25	54.44	21.48	176.94	9.21	28.79
Harsh accelerations/100 km	0.00	23.12	299.4	0.00	16.26	0.00	22.58	298.86	0.00	12.77	1.55	18.79	282.80	0.00	14.89
Harsh brakings/100 km	10.10	26.36	259.34	0.00	28.60	4.49	23.21	275.23	0.00	21.72	13.55	25.76	286.77	0.00	30.53
Total duration (sec)	801	967.6	14,513	300	766	786	971.85	20,380	300	741	934	1361.4	36,138	300	1016
Total distance driven per trip (km)	7.75	20.68	428.72	1.03	11.73	7.20	19.30	504.73	0.54	9.99	11.5	34.46	1006.6	0.61	17.27
Mobile phone usage duration/driving duration	0%	15.91%	99.66%	0%	2.03%	0%	15.23%	99.86%	0%	1.93%	1.32%	19.77%	99.80%	0%	13.89%
Stringency index	84.2	0.4	84.3	81.5	0.00	81.4	2.7	84.3	78.7	5.56	86.2	8.6	94.4	69.9	5.45

second lockdown period. This is due to the fact that drivers observing empty roads were willing to undertake more risks and appeared to have worse driving behavior in comparison to the second wave of the COVID-19 pandemic. Furthermore, it should be noted that during the lockdown period in the KSA, an overall increase in average speeding, average total/driving speed, and harsh events per distance was observed compared to the entire database.

3.4. Traffic exposure data

In order to be able to provide a holistic overview of the COVID-19 impact, exposure data were also extracted. As usually traffic data acquisition from national authorities requires additional time, it was chosen to use the Apple mobility report data as a proxy of traffic exposure in the study areas. Similar data have been utilized in previous work with regards to driving behavior and COVID-19 (Katrakazas et al., 2020). The aggregated data collected from Apple showed the mobility trends for major cities and several countries or regions. The information was generated by counting the number of requests made to Apple for directions. Data availability in a particular city, country, or region was subject to a number of factors, including minimum thresholds for direction requests made per day.

3.4.1. Driving traffic volumes

The COVID-19 outbreak as well as government responses of each country had a much more volatile effect on travel patterns. To begin with, during the first lockdown period of COVID-19, a great reduction in the volume of people driving was identified in Greece. However, during the second lockdown period, driving traffic volumes were much higher, roughly by 91% than the first one. Actually, during the first lockdown, traffic virtually disappeared, but ahead of a second one, there was a noticeable rise in people traveling by road. Interestingly, Greece performed high congestion levels throughout the summer. Traffic started to pick up significantly during August 2020 when the peak-time congestion on Greek roads hit a 400% increase, compared to the first lockdown. According to the available data for November-December 2020, driving traffic volumes decreased and a 62% reduction was identified compared to the period between the first and the second lockdown. With regards to the KSA, when the curfew was lifted throughout the country, driving traffic volumes started to increase significantly and an 82% rise was observed compared to the period of the first lockdown. Figs. 1 and 2 illustrate the volume of driving sessions of Apple users from January 2020 to December 2020 along with the number of COVID-19 cases in Greece and the KSA, respectively.

4. Methodology

In order to provide a yearly overview of the impact of COVID-19, the aforementioned data are presented descriptively in order to identify critical changes throughout 2020. Comparisons were made with regards to the lockdown periods as well as periods with restrictions between the lockdown states, and explanatory figures are provided to depict the status of driving behavior and road safety indicators in relation to COVID-19 cases.

In order to identify patterns with regards to driving behavior indicators and the strictness of response measures (stringency index), clustering was utilized on weekly aggregated OSeven trip data. Within the present study, clustering is a useful technique in order to divide the trip sample into several distinct categories. The evaluation of the cluster centroids describing these categories can provide insights as to whether driving behavior differs system-

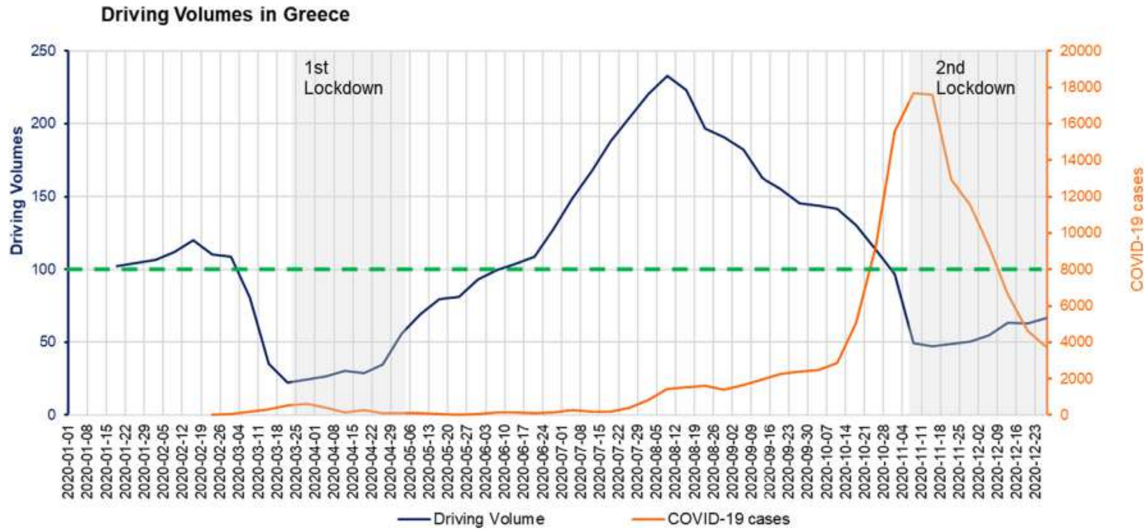


Fig. 1. Driving volumes per week along with the evolution of COVID-19 cases in Greece (Source: Apple, Data Processing: NTUA).

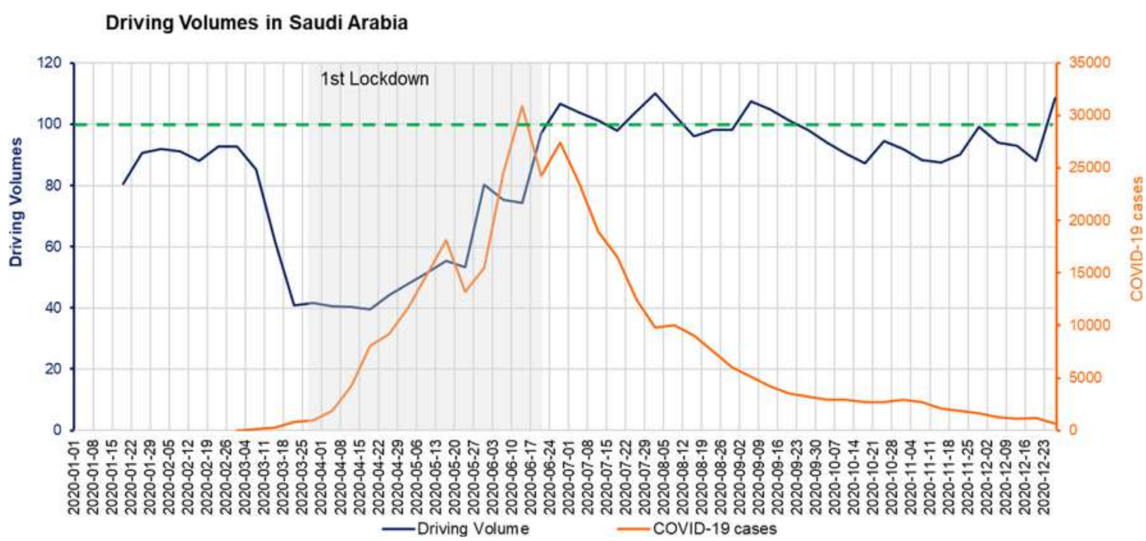


Fig. 2. Driving volumes per week along with the evolution of COVID-19 cases in the KSA (Source: Apple, Data Processing: NTUA).

atically on a macroscopic scale during the periods of shifting measures due to the pandemic. A well-known and straightforward technique is k-Means clustering, an algorithm used to divide datasets into clusters of similar magnitudes.

The k-means algorithm searches for a specific number of clusters (k) in a given dataset. The algorithm first initiates by randomly selecting centroids in the data. Each data-point is then assigned to the nearest centroid, forming the requested k clusters. Centroids are re-computed for the formed clusters, and thus their location changes. Calculations are then performed to re-assign each data-point to their new centroid. Afterwards, iterative calculations are conducted until no reassignments are made and thus the centroids have stabilized. The popularity of k-means algorithms presented in the past (e.g., Hartigan & Wong, 1979) has led to several customized approaches in the literature (e.g., Kanungo et al., 2002; Likas et al., 2003). K-means has been used widely for clustering purposes in several transport/road safety studies as well (e.g., Yannis et al., 2007; Mantouka et al., 2019).

5. Descriptive overview of the COVID-19 impact on driving behavior

5.1. Trip characteristics

In this section, the impacts of COVID-19 on trip characteristics and, more specifically, on total trip duration and distance are discussed.

5.1.1. Total duration

From Fig. 3, it is evident that total duration during the periods of lockdown is similar (i.e., a 1% reduction in total duration was observed in Greece during the second lockdown compared to the first one). When the restrictions on movement and business activities were gradually lifted, total duration increased by 22%, compared to the first lockdown period. At the same time, total duration dropped again roughly by 19% during the second lockdown compared to the period between the first and second lock-

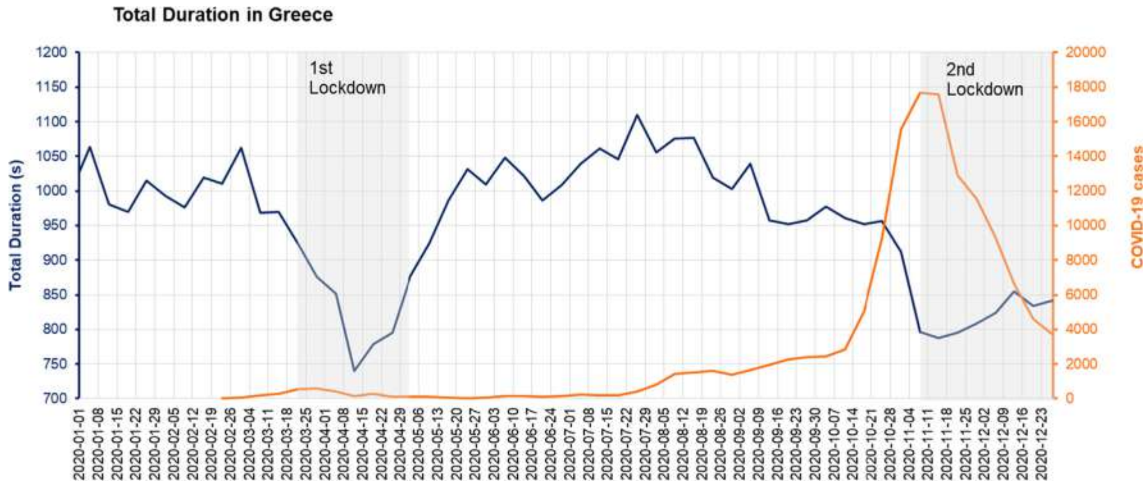


Fig. 3. Total duration per week along with the evolution of COVID-19 cases in Greece (Source: OSeven, Data Processing: NTUA).

down. In the KSA, after the end of the lockdown, more vehicles on city streets were observed; thus, a 10% rise in total duration was identified. Overall, the total number of driving trips in the KSA was significantly reduced due to the lockdown period, as depicted in Fig. 4.

5.1.2. Total distance driven per trip

In accordance with the total distance driven, the COVID-19 pandemic also had a direct effect on active drivers on the roads. Specifically, the second wave of COVID-19 pandemic led to a 10% reduction in total distance driven in Greece, compared to the first one. After the end of the first lockdown period, Greek drivers started to increase weekly mileage, reaching an 18% increase in total distance monitored in March and April compared to the period between the first and the second lockdown (i.e., from May to early-November 2020). Total driving distance per week dropped again by around 23% in November and December after the second lockdown compared to the period from May to early November. Similarly, the total distance driven per trip was also reduced during the lockdown period in the KSA. After the end of the lockdown period, a 13% increase in miles driven was observed when comparing data from March to June (i.e., COVID-19 lockdown period) with data from the end of June to December (i.e., after the end of lockdown of non-essential movements). Figs. 5 and 6 illustrate the changes in the total distance driven per trip in Greece and the KSA, respectively.

5.2. Driving behavior

5.2.1. Average driving speed

It is worth mentioning that during the first and second lockdown periods, an overall increase in average driving speed was identified compared to the period between the first and the second lockdown (i.e., from May to early-November 2020). When a decrease in driving traffic volumes was observed, drivers tended to increase their average driving speed. In particular, after the end of the first lockdown period, Greek drivers started to gradually increase their average driving speed, while a 5% drop in average driving speed monitored in March and April was identified compared to the period between the first and second lockdown (i.e., from May to early-November 2020). Additionally, the second wave of COVID-19 pandemic led to a 5% decrease in average driving speed in Greece, compared to the first one. Interestingly, average driving speed had not changed in November and December (i.e., after the second lockdown had been announced) compared to the period before (i.e., between the first and the second lockdown from May to early-November). Regarding the KSA, no change in average driving speed was identified, when comparing data from March to June (i.e., COVID-19 lockdown period) with data from end-June to December (i.e., after the end of lockdown of non-essential movement). Figs. 7 and 8 illustrate the changes in the average driving speed in Greece and the KSA, respectively.

5.2.2. Average speeding

As shown in Figs. 9 and 10, both in Greece and the KSA, average speeding was reduced due to the lockdown restrictions. In particular, during the second lockdown period, a negligible 1% reduction in average speeding was identified in Greece compared to the first one. Additionally, when Greece began to gradually lift restrictions on movement and restart business activities, average speeding decreased by 1% compared to the first lockdown period (i.e., March and April 2020), while there was no change in the average speeding during the second lockdown period compared to the period between the first and the second lockdown. Regarding the KSA, after the end of the lockdown period, a 9% decrease in average speeding was observed.

5.2.3. Speeding percentage

The second wave of COVID-19 pandemic led to a 2% reduction in the ratio of speeding duration/driving duration in Greece, compared to the first one. After the end of the first lockdown period, a 9% drop in speeding percentage compared to the period between the first and the second lockdown was identified. Furthermore, the ratio of speeding duration/driving duration was increased by around 8% in November and December after the second lockdown had been announced compared to the period before (i.e., between the first and the second lockdown from May to early-November). Interestingly, the speeding percentage was also reduced after the end of the lockdown period in the KSA. In particular, a 23% drop was observed when comparing data from March to June (i.e., COVID-19 lockdown period) with data from end-June to December (i.e., after the end of lockdown of non-essential movements). Figs. 11 and 12 illustrate the changes in the ratio of speeding duration/driving duration in Greece and the KSA, respectively.

5.2.4. Harsh accelerations per 100 km

With regards to harsh accelerations per 100 km, during the first phase of the lockdown and especially in April 2020, these were

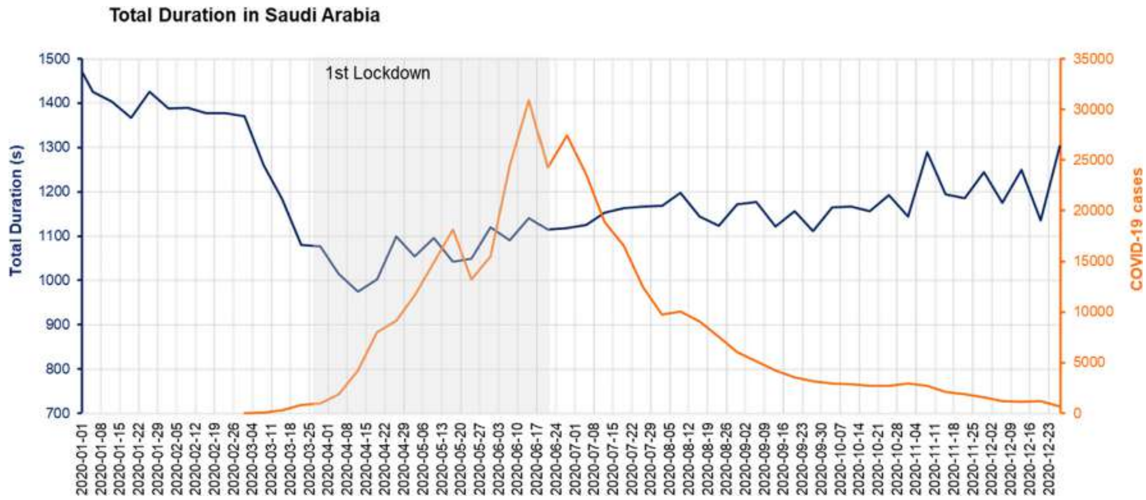


Fig. 4. Total duration per week along with the evolution of COVID-19 cases in the KSA (Source: OSeven, Data Processing: NTUA).

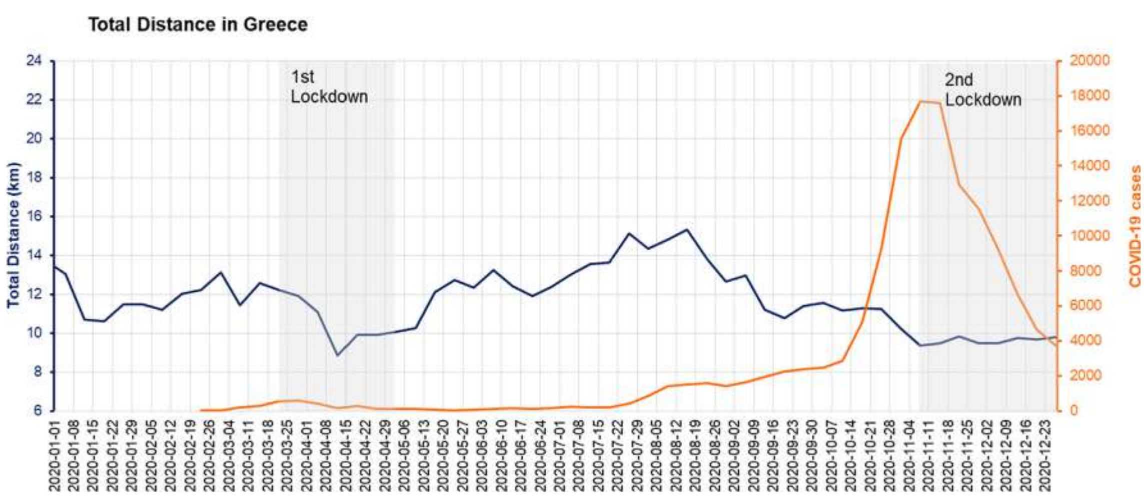


Fig. 5. Total distance per week along with the evolution of COVID-19 cases in Greece (Source: OSeven, Data Processing: NTUA).

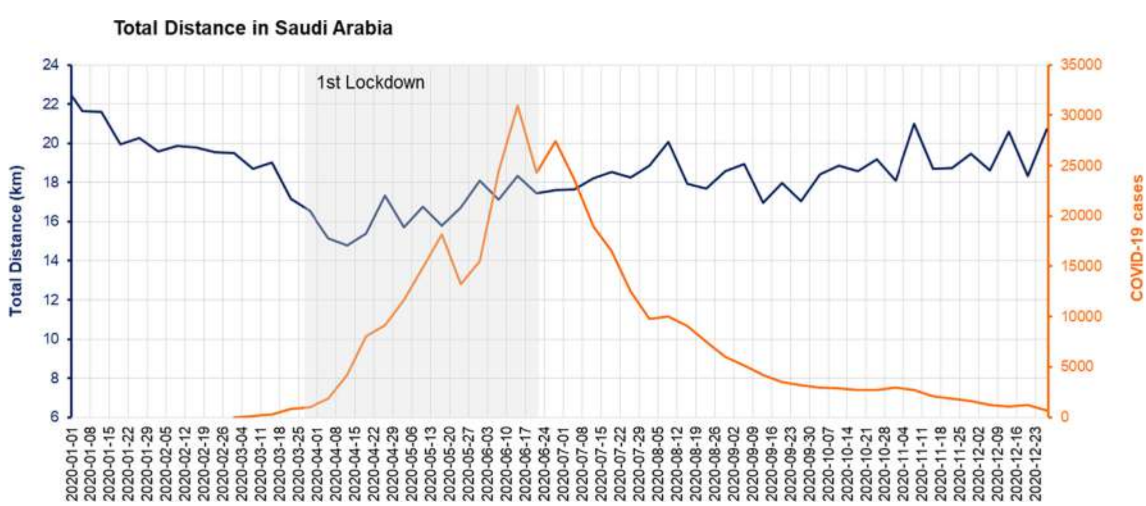


Fig. 6. Total distance per week along with the evolution of COVID-19 cases in the KSA (Source: OSeven, Data Processing: NTUA).

increased compared to February (i.e., before the appearance of COVID-19 pandemic), as shown in Fig. 13. It should be noticed that during the second lockdown period, a 17% decrease in harsh accelerations per 100 km was identified in Greece compared to the first one. When the restrictions on movement and business activities were gradually lifted, harsh acceleration events per distance again

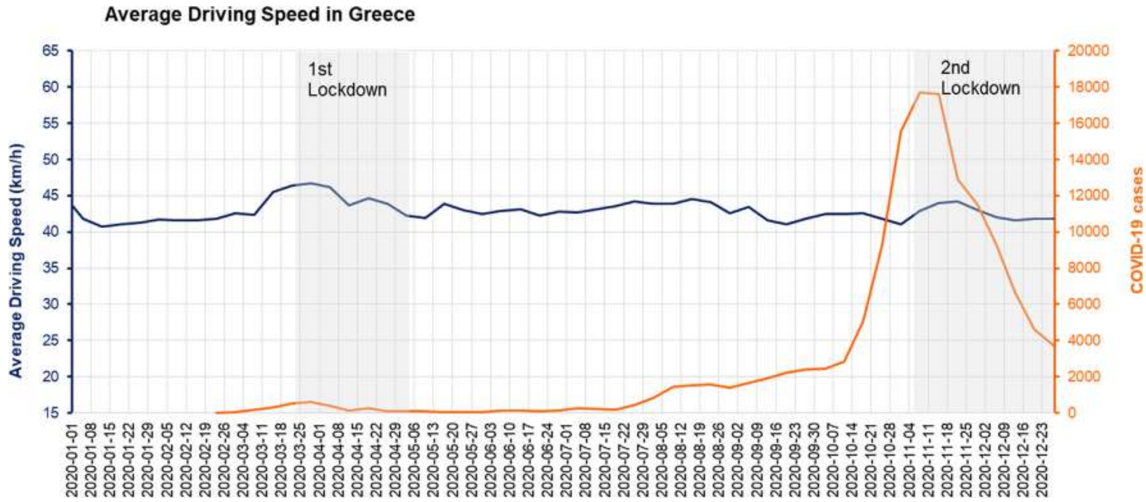


Fig. 7. Average driving speed per week along with the evolution of COVID-19 cases in Greece (Source: OSeven, Data Processing: NTUA).

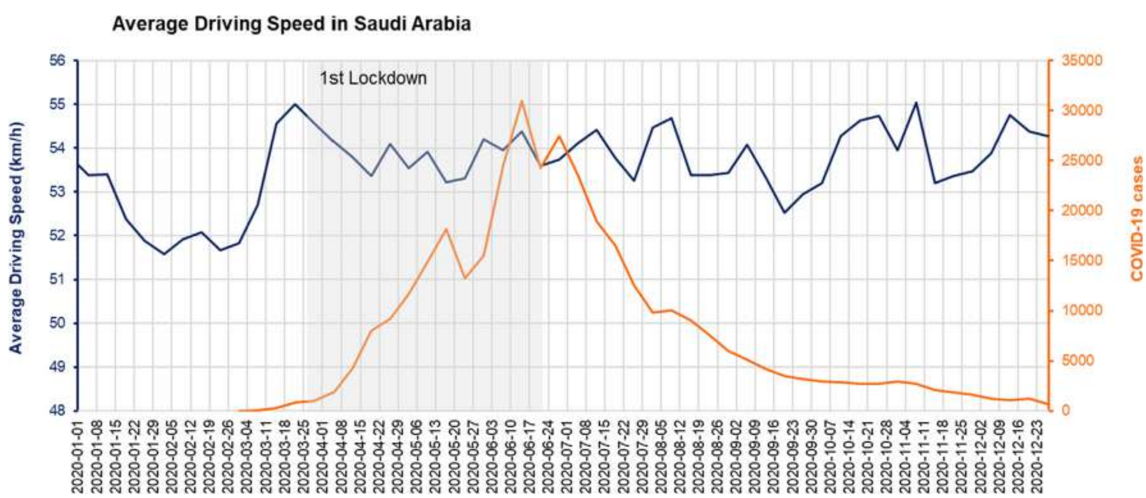


Fig. 8. Average driving speed per week along with the evolution of COVID-19 cases in the KSA (Source: OSeven, Data Processing: NTUA).

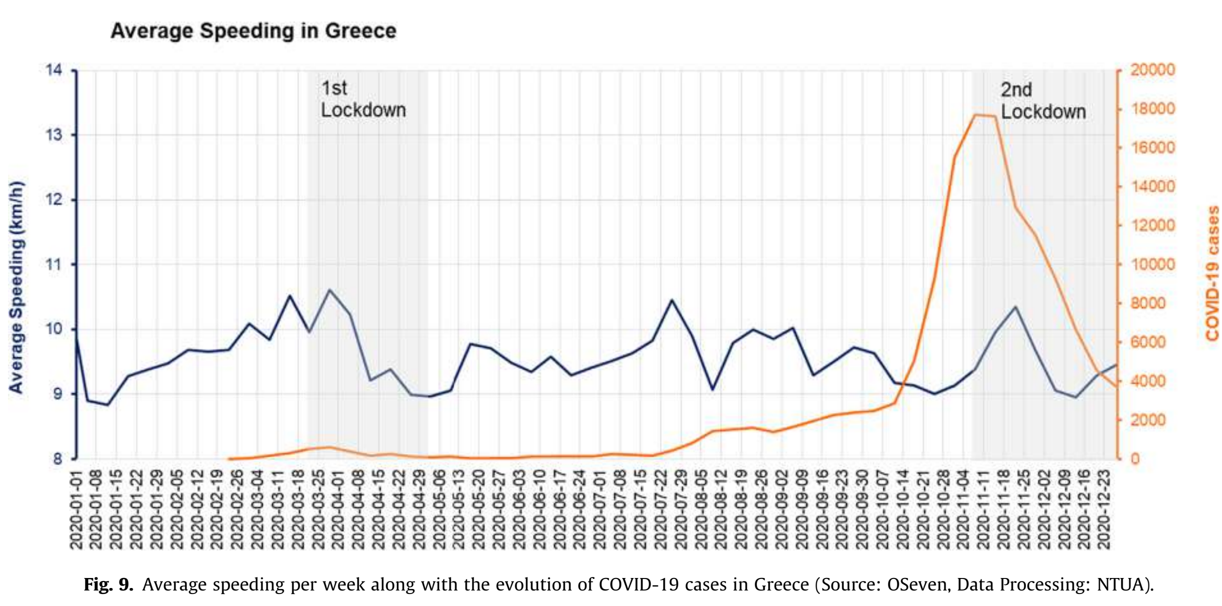


Fig. 9. Average speeding per week along with the evolution of COVID-19 cases in Greece (Source: OSeven, Data Processing: NTUA).

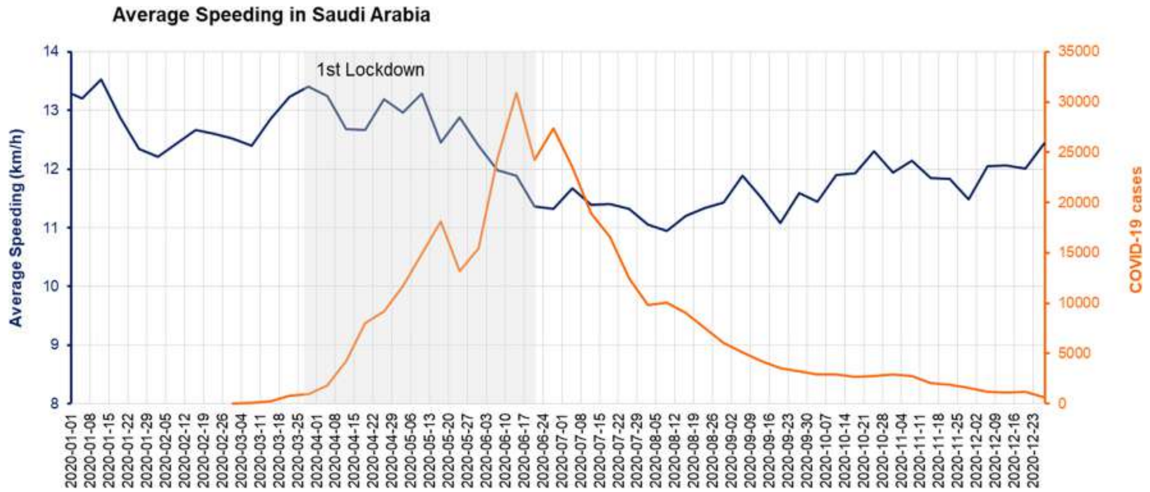


Fig. 10. Average speeding per week along with the evolution of COVID-19 cases in the KSA (Source: OSeven, Data Processing: NTUA).

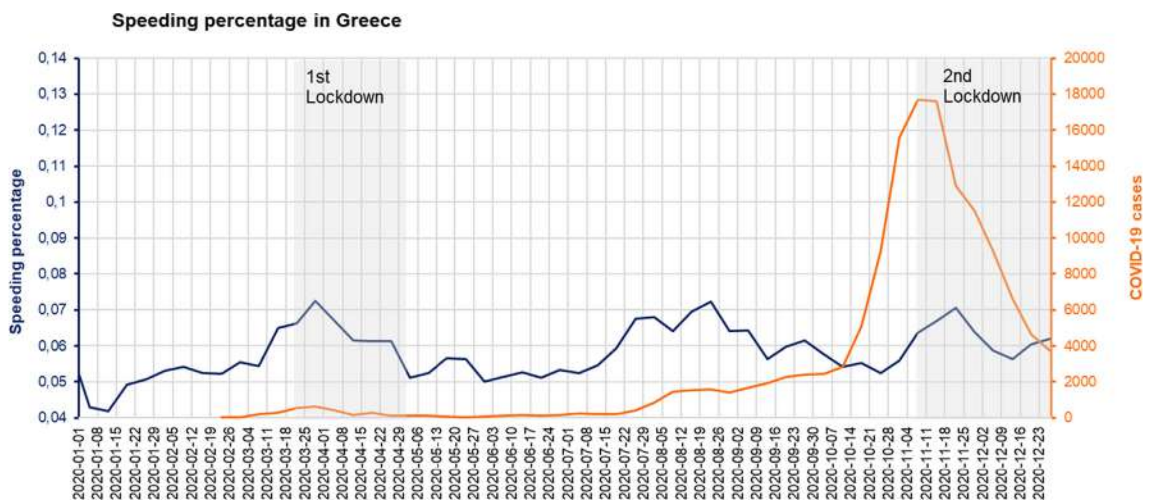


Fig. 11. Speeding percentage per week along with the evolution of COVID-19 cases in Greece (Source: OSeven, Data Processing: NTUA).

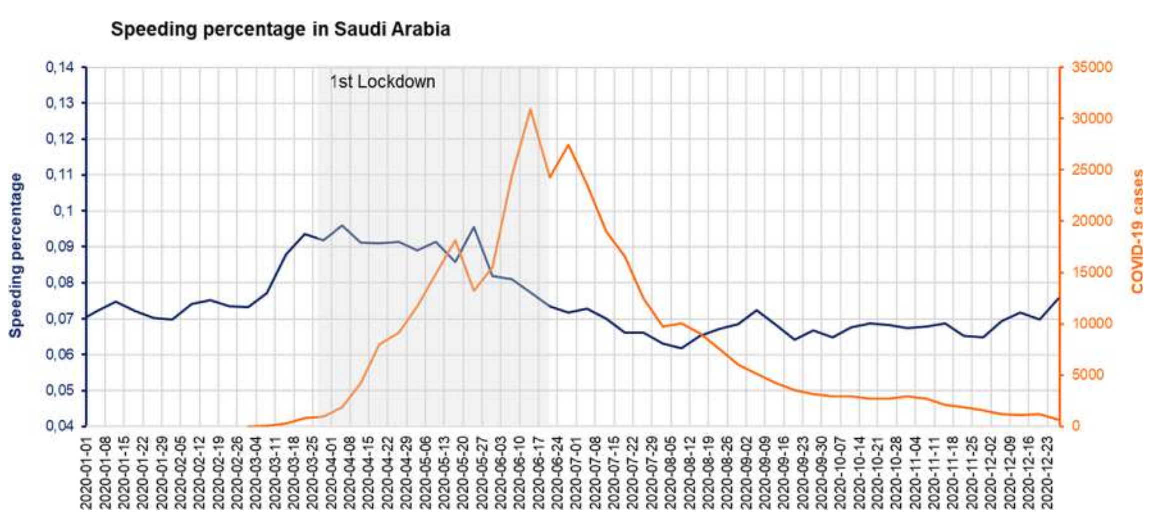


Fig. 12. Speeding percentage per week along with the evolution of COVID-19 cases in the KSA (Source: OSeven, Data Processing: NTUA).

dropped by 18% compared to the first lockdown period. Interestingly, a negligible 2% increase in harsh accelerations per 100 km was observed during the second lockdown compared to the period between the first and the second lockdown. With regards to the KSA, it was revealed that drivers accelerated harshly during the months of COVID-19. Overall, after the end of the lockdown period,

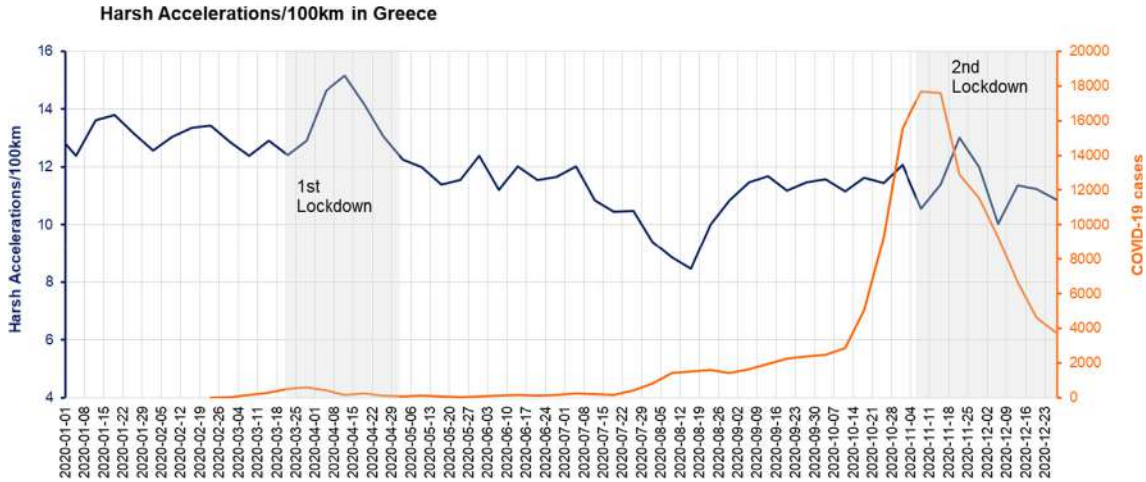


Fig. 13. Harsh accelerations/100 km per week along with the evolution of COVID-19 cases in Greece (Source: OSeven, Data Processing: NTUA).

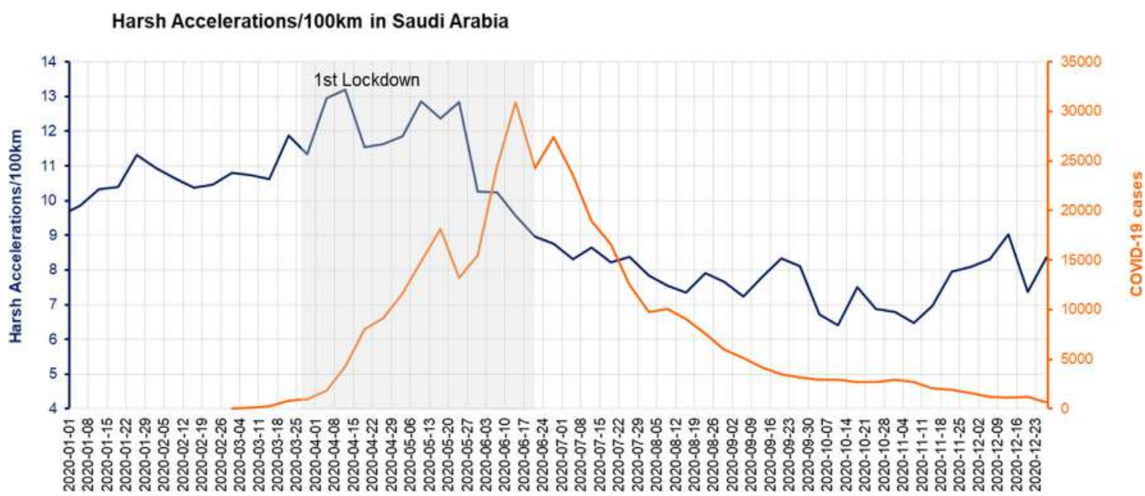


Fig. 14. Harsh accelerations/100 km per week along with the evolution of COVID-19 cases in the KSA (Source OSeven, Data Processing: NTUA).

a 34% decrease in harsh accelerations per 100 km and less harsh accelerations per distance were observed, which indicates that drivers improved their driving behavior after the COVID-19 pandemic, as depicted in Fig. 14.

5.2.5. Harsh brakings per 100 km

Similar to harsh acceleration patterns, harsh brakings per 100 km were decreased during the second lockdown period compared to the first one. Additionally, when Greece began to gradually lift restrictions on movement and restart business activities, harsh brakings per distance dropped by 33% compared to the first lockdown period (i.e., March and April 2020), while there was a 10% increase in harsh brakings per 100 km during the second lockdown period compared to the period between the first and the second lockdown. With regards to the KSA, after the end of the lockdown period, a 23% reduction in harsh brakings per distance was identified. Figs. 15 and 16 depict the changes in harsh brakings per 100 km in Greece and the KSA, respectively.

5.2.6. Mobile phone usage duration/driving duration

With regards to mobile phone use, a general increase in the ratio of mobile phone usage duration per driving duration during the lockdown periods, compared to the period between the first and the second lockdown in Greece and the KSA, respectively, is presented in Fig. A3 in the Appendix. In more detail, the second wave of COVID-19 pandemic led to a 6% decrease in the ratio of mobile phone usage duration per driving duration in Greece, compared to the first one. After the end of the first lockdown period, a 9% reduction in the ratio of mobile phone usage duration per driving duration was identified compared to the period between the first and the second lockdown. Interestingly, mobile phone usage duration per driving duration increased by 4% in November and December after the second lockdown had been announced compared to the period between the first and the second (i.e., from May to early-November 2020). Similarly, the ratio of mobile phone usage duration per driving duration was also raised during the lockdown period in the KSA. Afterwards, a 5% increase was observed when comparing data from March to June (i.e., COVID-19 lockdown period) with data from end-June to December (i.e., after the end of lockdown of non-essential movement).

5.3. Road crashes

A more comprehensive picture of the effects of COVID-19 pandemic on road safety can be drawn from the high-quality data on the total number of road crashes. Fig. 17 illustrates the difference

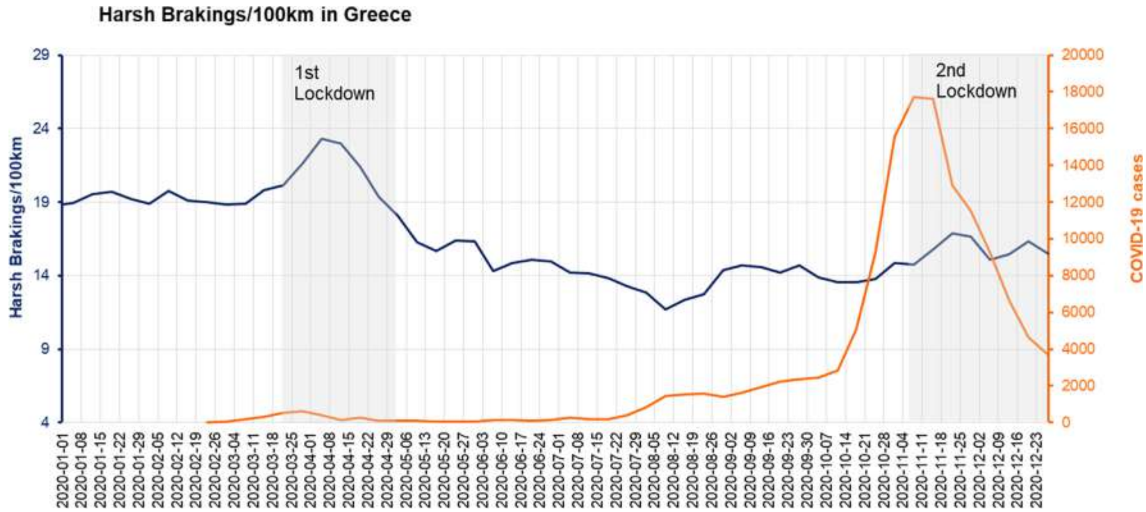


Fig. 15. Harsh brakings/100 km per week along with the evolution of COVID-19 cases in Greece (Source: OSeven, Data Processing: NTUA).

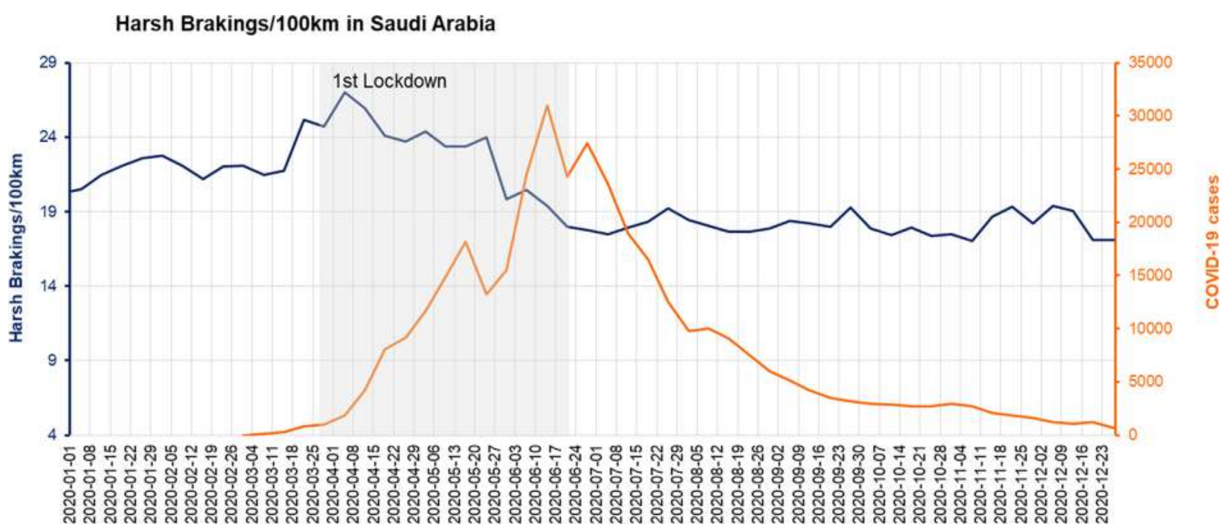


Fig. 16. Harsh brakings/100 km per week along with the evolution of COVID-19 cases in the KSA (Source: OSeven, Data Processing: NTUA).

in the total number of road crashes from January to December 2020 in Greece. Specifically, during the second lockdown period, a 46% increase in the total number of road crashes was observed compared to the first one. After the end of the first lockdown period, driving volumes were gradually increased and a 116% rise in the total number of road crashes was identified in the period between the first and the second lockdown compared to the first lockdown period. Interestingly, a 32% reduction in road crashes was observed in November-December 2020 (i.e., during the second lockdown) compared to the period before (i.e., between the first and the second lockdown from May to early-November 2020). Lastly, it should be noted that monthly data for road crashes are available only for Greece, while there is no evidence for road crashes in the KSA during 2020.

5.4. Overview

Table 4 summarizes the changes in exposure, driving behavior, road crashes, and response measures strictness during and after the lockdown periods for each country. It should be clarified that the first lockdown period in Greece refers to the period from April to May 2020, the second lockdown period refers to the period from early-November to December 2020, while the period between the first and the second lockdown refers to the period from May to early-November 2020. With regards to KSA, the first lockdown period (i.e., from end-March to June 2020) is compared to the post-lockdown period (i.e., from June to December 2020).

The total number of trips and traffic volumes in Greece reduced by 70% for people driving, respectively, during the first lockdown compared to the period before the appearance of COVID-19 pandemic. However, increased driving volumes, roughly by 100%, during the second COVID-19 lockdown compared to the first one. Exposure indicators (i.e., distance traveled, total duration, and driving duration) were also decreased during the first lockdown period compared to the period before. Similarly, the aforementioned indicators were also reduced during the second COVID-19 lockdown compared to the first one and the period between the lockdowns. A dramatic increase in total/driving duration and total distance during the second lockdown period compared to the period between the first and the second lockdown was also identified.

In Greece, during the first lockdown period, driving behavior indicators (i.e., average speeds, speeding percentage, harsh accelerations/ brakings per 100 km, mobile phone usage duration/driving duration) increased to a great extent compared to the period

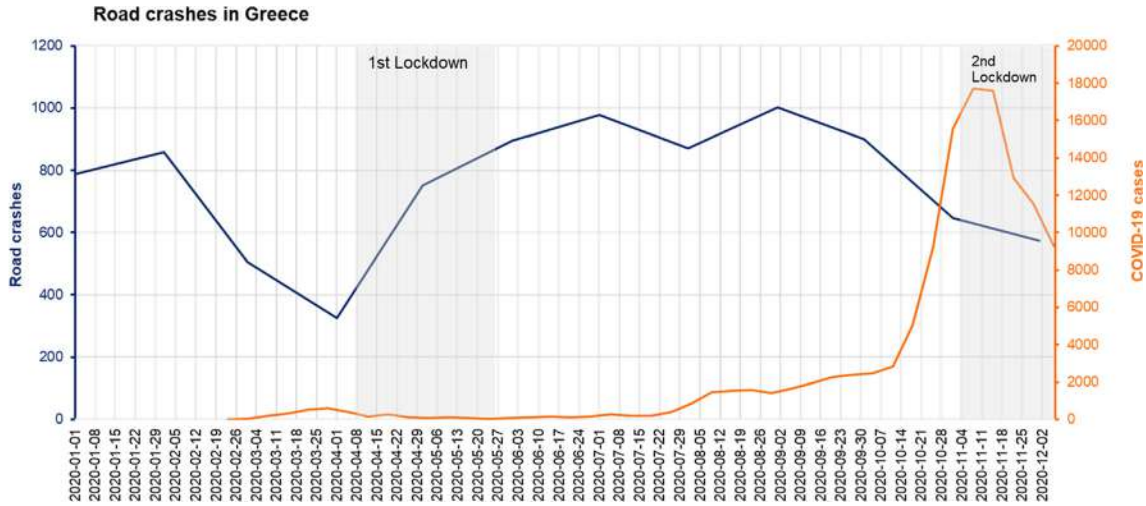


Fig. 17. Road crashes per month along with the evolution of COVID-19 cases in Greece **(Source: ELSTAT, Data Processing: NTUA)**.

Table 4
Changes in traffic volumes, driving behavior and road safety during and after the COVID-19 lockdown period in Greece and the KSA.

		Greece				KSA	
		Percentage change comparison between *period before and 1st lockdown*	Percentage change comparison between *1st lockdown and- 2nd lockdown*	Percentage change comparison between *1st lockdown and period between 1st and 2nd lockdown*	Percentage change comparison between *period between 1st and 2nd lockdown and 2nd lockdown*	Percentage change comparison between *period before and 1st lockdown*	Percentage change comparison between *1st lockdown and period after*
Exposure	Total duration	-18%	-1%	22%	-19%	-21%	10%
	Total distance	-11%	-10%	18%	-23%	-18%	13%
	Driving traffic volumes	-70%	91%	400%	-62%	-36%	82%
Driving behavior	Average driving speed	7%	-5%	-5%	0%	2%	0%
	Average speeding	1%	-1%	-1%	0%	0%	-9%
	Speeding percentage	22%	-2%	-9%	8%	17%	-23%
	Harsh accelerations per 100km	5%	-17%	-18%	2%	11%	-34%
	Harsh brakings per 100km	11%	-26%	-33%	10%	7%	-23%
	Mobile phone usage duration/driving duration	21%	-6%	-9%	4%	-8%	5%
Road Crashes	Crashes*	-49%	46%	116%	-26%	-	-
Response measures strictness	Stringency index	NA	-3%	-33%	45%	NA	-31%

*Road crashes data for the KSA are not available.

before. For instance, during the first lockdown, average speed, speeding percentage, and mobile phone usage duration/driving duration increased by 10%, 22%, and 21%, respectively, compared to the period before and 1% (second lockdown) compared to the pre-pandemic period. This indicates that with fewer vehicles on city streets, slightly more drivers are blowing the speed limit. After the end of lockdown periods, a significant drop in speeding percentage was identified. Harsh accelerations/100 km and harsh brakings/100 km increased by 5% and 11% during the first lockdown compared to the period before. Interestingly, during the second lockdown, harsh events reduced by up to 17%.

It is worth mentioning that during the first lockdown period in Greece, an overall 50% reduction in road traffic crashes was observed compared to the period before the appearance of COVID-19 pandemic. In addition, during the second lockdown period, a 26% decrease in the total number of road traffic crashes was identified compared to the period between.

With regards to the KSA, similar patterns to Greece were also observed for both exposure and driving behavior indicators. A 36% and 27% reduction of people driving was identified during the lockdown period compared to the period before. After the lockdown, people driving adapted immediately to baseline frequencies and traffic volumes increased. In addition, a 2% spike in average speed was identified during the lockdown period compared to the period before. The speeding percentage increased by 17%, while the number of harsh accelerations and brakings per 100 km

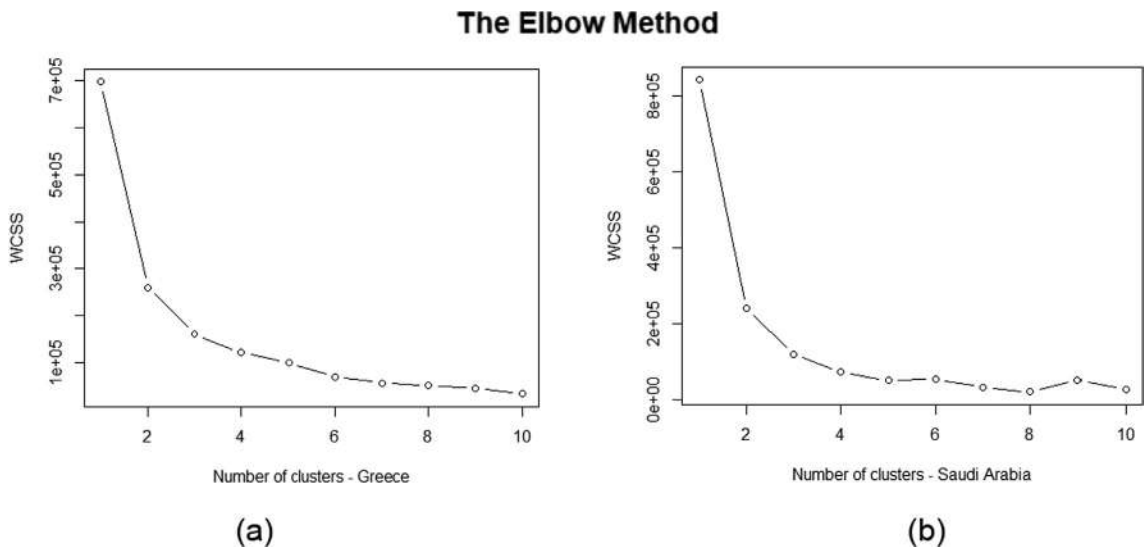


Fig. 18. Optimal number of clusters according to Elbow method for **(a) Greece and (b) the KSA.**

increased by 11% and 7%, respectively. Lastly, it should be noted that monthly data for road traffic crashes were not available for KSA.

6. Pattern identification according to driving characteristics and stringency index

This section presents the results of cluster analysis conducted on weekly aggregated data for Greece and the KSA. The process of selection of the final number of clusters is initially explained, followed by the presentation and interpretation of the centroid values.

In the current study, the elbow method is followed, which defines clusters by minimizing the total intra-cluster variation, expressed by the Within Cluster Sum of Squares (WCSS) (Kodinariya & Makwana, 2013). For the purpose of choosing the optimal number of clusters, the elbow method using WCSS was applied as shown in Fig. 18 for Greece (left) and the KSA (right). These figures depict the optimal number of clusters (i.e., the *k* value to be applied for the examined data set. The optimal value is extracted by the "elbow" or "knee" of the depicted curve (Kodinariya & Makwana, 2013). Consequently, the optimal number of clusters is three for both countries. In addition, the average silhouette method was tested in order to validate the outcomes of the elbow method. The average silhouette score for three clusters was equal to 0.47 in the Greek dataset and 0.58 in the KSA dataset, denoting acceptable k-means algorithm fits.

The added value of the k-means clustering approach is the confirmation of systematic differences in the data. In other words, after applying this unsupervised method, it is observed that the weeks can be consistently organized in distinct categories with different centroid values. Centroid differences are observed across all driver performance metrics and for the Stringency Index as well. Therefore, this machine learning algorithm provides added validation of the differences already perceived by descriptive statistics for the examined data.

6.1. Greece

For Greece, the clustering results of weekly aggregated data for 2020 are shown on Table 5. The practical meaning of each category is to provide a grouping for similar week types of 2020. In the clustering analysis, seven variables (i.e., total duration, total distance, harsh brakings/100 km, driving volume, driving speed, speeding percentage, and stringency index of lockdown) were included. The presented values of the variables were the mean value of each centroid along with the corresponding standard deviation enclosed in parentheses.

Based on centroid values, the weekly data of 2020 can be classified into the following three distinct categories:

- Cluster 1 – 'Baseline': Comprises 29 weeks with a centroid stringency index of 36.99. These can be considered as the "baseline" weeks, before the existence of lockdown measures (January and February) and time periods with the loosest restrictions (June, the second half of September and October). The baseline cluster is depicted in green in Fig. 19.
- Cluster 2 – 'Restrictions': Comprises 10 weeks with a centroid stringency index of 54.33. This group of weeks belongs to the summer of 2020 (specifically in July, August and half of September) when modest restriction measures were in place. As restrictive measures were not as strict or extensive as a full lockdown, this category is distinct from the lockdown category. These months formed a separate category probably due to the increased stringency index compared to June and October. At that time, all the examined variables (driving behavior and driving volume indicators) present peaks according to the descriptive statistics. The descriptive average stringency index of these two months is 55, which is quite close to the stringency index presented by the clustering analysis. The restrictions cluster is depicted in red in Fig. 19.
- Cluster 3 – 'Lockdown': Comprises 14 weeks with a centroid stringency index of 82.13. This group of weeks represents the periods where full lockdown measures were in place. The descriptive average stringency index for both lockdowns is 85, which is quite similar to the calculated centroid value. The lockdown cluster is depicted in blue in Fig. 19.

When examining the centroid values of these categories, the differences in driver behavior parameters become apparent. The restrictions cluster shows increased distance and trip duration from the baseline cluster, indicating that trips are longer and farther after the lockdown period. Driving volume appears to be considerably increased, probably due to two factors: (a) the anxiousness of people to travel after the lockdown period and (b) the large tourism-related traffic volumes generated during the

Table 5
Clustering centroid mean values per week type for Greece.

Number	Category	Total Duration	Total Distance	Harsh Brakings /100 km	Driving Volume	Driving Speed	Speeding percentage	Stringency Index
1	Baseline	989.23 (38.35)	11.77 (0.90)	16.71 (1.40)	97.59 (43.74)	42.25 (1.83)	5.36% (0.72%)	36.99 (27.02)
2	Restrictions	1052.78 (30.84)	13.93 (0.94)	13.42 (0.75)	195.43 (26.07)	43.61 (1.17)	6.36% (1.01%)	54.33(5.01)
3	Lockdown	825.72 (46.92)	10.04 (0.94)	18.23 (1.74)	44.47 (15.26)	43.68 (2.69)	6.29% (0.79%)	82.13(3.00)

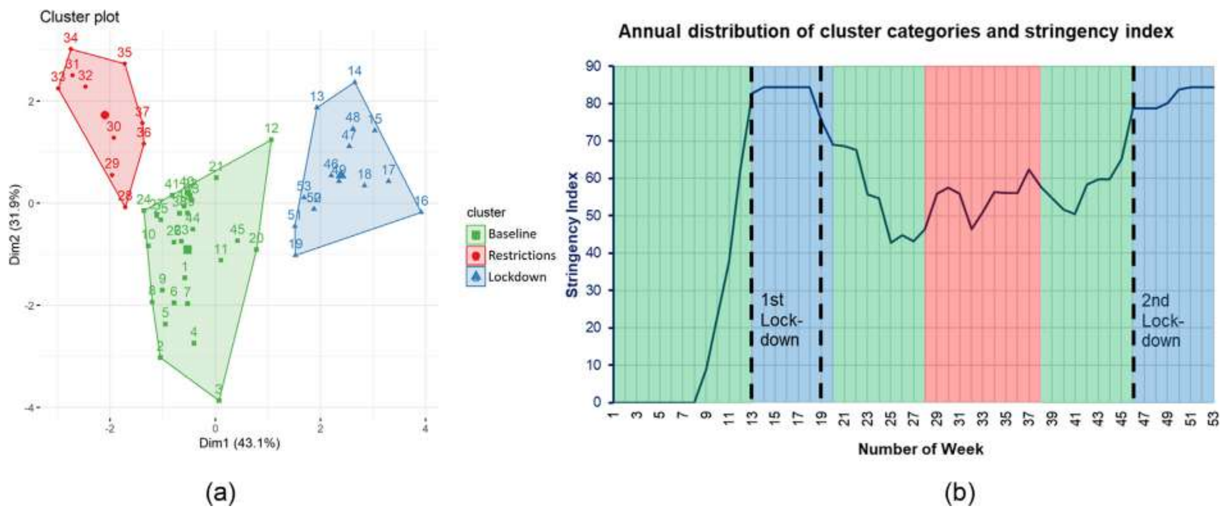


Fig. 19. (a) Cluster plot and (b) Annual distribution of cluster categories and stringency index (Greece).

summer in Greece. Harsh brakings/100 km appear to be reduced from the baseline, possibly due to more careful driving or disproportionately increased travel distances. Since the temporal percentage of speeding duration seems elevated, the second explanation appears more credible.

As expected, the lockdown cluster is the one with the most limited travel, with reductions in trip duration, total distance, and driving volume. Interestingly, a peak of harsh brakings/100 km is also observed, indicating that this category comprises weeks with increased road safety risk. Drivers may have been more aggressive overall due to encountering less traffic and more available space on the roads, and taking advantage of this new situation. The ratio of speeding to driving duration appears similar to that of the restrictions cluster.

Despite these observed differences, driving speed appears mostly unaffected between clusters; this is perhaps due to the large number of trips, which leads to the absorption of fluctuations for this parameter. It is worth highlighting that in R-studio, clusters are represented by conducting an internal Principal Component Analysis (PCA) and then visualizing the results with the two most prominent components serving as × and y axis. In other words, PCA is conducted primarily in order to supply distinguishable clusters for easier visualization.

The obtained cluster categories are also illustrated in the plots of Fig. 19. In the upper part of Fig. 19, a 2-dimensional cluster plot is used to depict all the clusters and centroids of the analysis. It should be noted that week numbering starts by considering the first lockdown week as week 1, the following week as week 2, and so on until week 53. As per standard process, cluster plot axes are described by the two most prominent PCA components of the sample (Dimension 1 and 2). In the lower part of Fig. 19, an explanatory plot with all the clusters was created in order to contrast the clustering results with the observed stringency index, plotted with a separate curve. The outcome of the clustering can be regarded as satisfactory based on the fact that both lockdowns were defined precisely enough by the k-means algorithm. Only one week (week 19; from 2020-05-03 until 2020–05-10) appears to have been erroneously classified.

For the purpose of confirming the clustering analysis, a linear discriminant analysis (LDA) was conducted with a train/test dataset ratio of 80%/20%. Results indicated correct classification in 100% of the test dataset cases for Greece.

6.2. KSA

Regarding the KSA, the clustering centroids of weekly aggregated data for 2020 are presented in Table 6, with standard deviations reported in parentheses. The process and interpretation of the centroids follows the same logic as the one followed for Greece:

- Cluster 1 – 'Baseline': Comprises 12 weeks with a centroid stringency index of 13.59. These can be considered as the "baseline" weeks, referring to the time period before the enforcement of lockdown measures (January, February, and three weeks of March). The baseline cluster is depicted in green in Fig. 20.
- Cluster 2 – 'Restrictions': Comprises 30 weeks with a centroid stringency index of 59.89. This group of weeks refers to the time period after the end of lockdown restrictions, when certain looser restrictions remained with an average stringency index. The restrictions cluster is depicted in red in Fig. 20.
- Cluster 3 – 'Lockdown': Comprises 11 weeks with a centroid stringency index of 88.92. These can be considered as the weeks with full lockdown measures in effect. The lockdown cluster is depicted in blue in Fig. 20.

As previously stated, the examination of clusters reveals the differences in driver behavior parameters. The 'restrictions' cluster shows decreased distance and trip duration from the baseline cluster, indicating that trips are shorter in distance and duration after the lockdown period, indicating a different reaction compared to Greece. Driving volume appears to be considerably increased, pos-

Table 6
Clustering centroid mean values per week type for the KSA.

Number	Category	Total Duration	Total Distance	Harsh Brakings /100 km	Driving Volume	Driving Speed	Speeding percentage	Stringency Index
1	Baseline	1382.19 (49.66)	20.49 (0.98)	20.94 (1.74)	68.71 (41.91)	52.76 (2.53)	7.24% (0.45%)	13.59 (18.36)
2	Restrictions	1164.09 (38.72)	18.44 (0.79)	18.48 (0.9)	94.67 (10.27)	53.83 (1.20)	6.96% (0.86%)	59.89 (9.63)
3	Lockdown	1052.90 (41.48)	16.22 (0.88)	24.20 (1.08)	48.25 (10.63)	53.90 (1.07)	9.07% (0.69%)	88.92 (9.63)

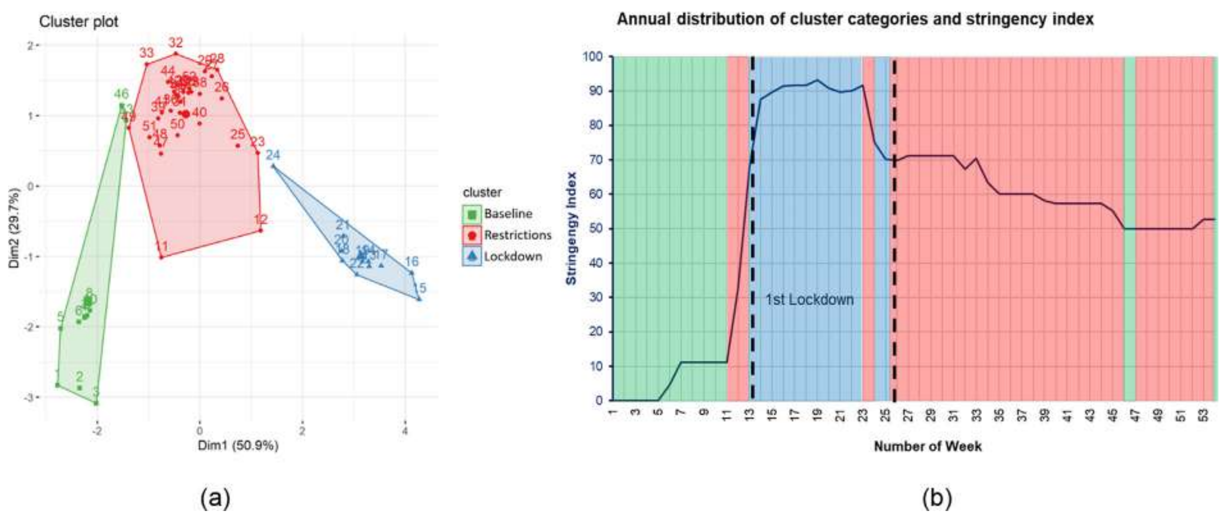


Fig. 20. (a) Cluster plot and (b) Annual distribution of cluster categories and stringency index (KSA).

sibly due to the same reasons for Greece, anxiousness of people to travel and tourism during these months. Harsh brakings/100 km appear to be once again reduced from the baseline. Since the temporal percentage of speeding duration appears close to the baseline, and total distance is reduced, innate more careful driving of drivers is a likely explanation.

As expected, once again the lockdown cluster is the one with the most limited travel, with the largest reductions in trip duration, total distance, and driving volume. Interestingly, a peak of harsh brakings/100 km is also observed, indicating that this category comprises weeks with increased road safety risk. Once again, drivers may have been more aggressive and taking advantage of emptier roads. The ratio of speeding to driving duration appears to be the highest between clusters. Consistently with Greece, driving speed appears mostly unaffected between clusters; this is perhaps due to the large number of trips, which leads to the absorption of fluctuations for this parameter.

The obtained cluster categories are also illustrated in the upper and lower plots of Fig. 20. The quality of the clustering can be considered adequate, as the lockdown period was defined sufficiently by the clustering analysis. Contrary to Greece, the KSA clusters were separated into three continuous time periods (before, during, and after the lockdown); it should be noted, however, that there were five erroneously grouped weeks this time. For both countries, the sample size utilized as input for the k-means algorithm, comprising of the 52 weeks of 2020, can be considered marginally adequate. While there is no precise definition in the literature, values of 2n are cited as appropriate (Formann, 1984), with n being the number of included variables. With the present data, five variables would be the maximum permissible, requiring $2^5 = 32$ data points. Since it was not possible to obtain more weekly average data within the framework of this study, the k-means algorithm was fitted for five and six variables only in addition to the previous analyses. The clustering results and categories were not affected. The cohesion as well as the reasonable centroid value interpretations of three clusters across two different countries allowed for the seven variable variants to be retained as the selected analyses.

The linear discriminant analysis (i.e., LDA) on the KSA clustering indicated correct classification in 90% of the test dataset cases for the KSA.

7. Discussion on the impact of the pandemic on road safety

The paper presents evidence of the impact of COVID-19 and constitutes a first attempt to understand the relationship between the strictness of corresponding response measures and the effect on driving behavior and crash frequency. The paper builds upon previous work (Katrakazas et al., 2020), which provided an overview of the effect on driving behavior and road safety during the first wave of the pandemic. The overview is completed with the current paper, which for the first time describes the effect of the pandemic using data from the entire 2020 year.

With "normal" everyday life disrupted, it was evident that driving behavior would be significantly affected. The first wave of the pandemic took governments and citizens by surprise, but at present, one year into the pandemic, signs of adjustment to the new reality are becoming apparent. For example, data showed that when lockdown states were lifted, a dramatic change in traffic by up to 400% in Greece was observed. In the KSA the corresponding change rose to 82%, an increase which is still considerably high. Following the same pattern, traffic volumes were substantially increased when comparing the first and second lockdowns. This finding is interesting when put into the perspective of the stringency of response measures and COVID-19 cases and casualties, since the lockdown periods (i.e., the first and second lockdown in Greece as well as the single lockdown in the KSA) are characterized by strict measures and a high number of COVID-19 cases and casualties. As a result, it can be tentatively assumed that drivers adjusted to these new conditions and behaved as if no restrictions were applied.

For Greece, driving after the shock of the first lockdown, all driving behavior indicators pointed towards safer attitudes. Harsh accelerations and brakings were reduced by up to 33% after lifting the response measures, while speeding and mobile phone use per driving duration were reduced by up to 9%. This positive outcome with regards to driving behavior, however, lasted until the imposition of the second lockdown in Greece, where it was demonstrated that drivers speeded more in relation with the average driving duration and also increased their harsh acceleration, harsh braking, and mobile phone usage frequency. On the contrary, in the KSA all driving indicators except mobile phone use, demonstrated a reduction after the end of the lockdown and thus hinted toward safer driving.

Unfortunately, the positive attitudes of Greek drivers immediately after the first lockdown and until the imposition of a second one, were not reflected in the frequency of road crashes. For example, crashes increased in Greece after the first lockdown by 116% when comparing the period between the first and second lockdown with time spent under confinement in March and April 2020. This is in contrast with other studies regarding the effect of COVID-19 on crash frequency (Inada et al., 2020; Muley et al., 2021; Saladié et al., 2020) although previous studies used a more limited time span for their data collection. The fluctuation in crashes and injuries during the period where lockdowns are imposed or resolved is something that is expected, however, due to the rapid changes in social and behavioral patterns during the pandemic (Calderon-Anyosa & Kaufman, 2021). Nevertheless, further in-depth crash research during the pandemic in Greece is needed in order to discover contributing and causal factors that led to the crashes in question.

Taking into account the stringency index of the response measures, it was demonstrated that driving behavior can be clustered into three categories: (a) baseline, which represents the period before restrictions or lockdowns, (b) restriction, which comprises periods where response measures start to get stricter, and (c) lockdown, which represents the strictest measures taken in form of a lockdown. From Table 6, it is evident that harsh brakings are one of the most representative indicators between clusters for Greece (together with driving volume and the stringency index), as large differences between the three groups are observed. Moreover, the most interesting finding from Table 6 is that in periods where restrictions are tightened in Greece, traffic volume is significantly higher, probably due to the good weather, as the "Restrictions" cluster includes the summer months and because of a potential fear of more strict measures that might come, which led people to more outdoor activities such as shopping when it was still allowed (2020). Similarly, with regards to the KSA, distinctions between clusters can be more clearly observed for harsh brakings and driving volume as well as the ratio between speed and driving duration, probably due to the fact that during strict response measures, streets were emptier. It should also be mentioned that the stress imposed by confinement conditions around the world (Lee & You, 2020; Lemke et al., 2020; Singh & Tech, 2020) might have also affected the increase of traffic volumes after the lift of the lockdown, as well as more dangerous behavior when streets were emptier due to stay-at-house governmental instructions.

One of the added contributions of this study was the continuous monitoring of smartphone metrics, which provide naturalistic driving data in a seamless and non-intrusive manner. These metrics disclose valuable information on the safety profiles of drivers and their fluctuations during times of the pandemic. The one-year span provides an opportunity to witness the effects of different lockdown or other restriction policies as they are not instantaneous in their inception, application, enforcement, and removal. The combination of these two features is also an advantage of the study; few naturalistic experiments are conducted in new-found conditions for such a duration.

One limitation of the present study is that the evidence presented by the authors majorly consists of descriptive statistics and explanatory figures, whereas a small part is dedicated to cluster analysis with regards to driving behavior and the contingency of response measures. The sample limitations with regards to the k-means application have also been mentioned and therefore an extended dataset should be explored in the future to confirm the present results. Building upon the previous limitation, syndemic or homeostatic compensations of the impacts of different pandemic periods or different types of travel (e.g., "familiar" and "non-familiar" travel) are not taken into account in the current approach. With respect to the distraction variable, a limitation of the analyzed data is that the application cannot recognize if a passenger uses the mobile phone of the driver. Nevertheless, this is typically not the case as the drivers understand that their phones become driving monitoring devices when using the application. Additionally, no data were available for the exposure indicator for crashes per kilometer driven. Moreover, it was not possible to include road type, time of day, and driving under the influence of alcohol or drugs in the analysis due to personal data protection legislation. Finally, regarding the driving behavior, data collection data were provided in their final format by OSeven Telematics, but the actual algorithms of obtaining the indicators (e.g., speeding or harsh events) from smartphone sensors are intellectually protected and unknown to the authors; therefore a "black-box" effect exists.

As a final remark, the unconventional truth is that, in Greece, the period after the first lockdown was imposed presents more similarities than differences with a "normal" situation regarding driving behavior. Differences between the period between the lockdowns in terms of driving behavior showed that speeds were similar between the two periods and the greatest effect was an increase by 10% of harsh brakings per 100 km of driving, during a period where stricter measures by 45% were imposed. On the contrary, in the KSA, although exposure increased in terms of traffic volumes and distance traveled, drivers were more careful while on the road.

8. Policy recommendations

The COVID-19 pandemic showcased the fragility that mobility patterns face in cases of unpredicted health or societal emergencies. Policymakers in the road safety domain should act proactively in the years to come to incorporate safety lessons from the pandemic period. As it was demonstrated in this paper, driving speed was significantly increased during the periods when lockdowns were imposed, due to the heavily reduced traffic volumes for motorized traffic. Towards that end, the paradigm of reducing speed limits inside urban areas to 20 or 30 km/h, as declared by the Stockholm Declaration of the 3rd global ministerial conference on road safety (Trafikverket | Swedish Transport Administration, 2021) should be extended. With lower speeds, crash risk, severe injuries, and harsh events will be apprehended. Such policies have already been applied in Paris, Brussels, and Bilbao[1] and could be extended to major metropolitan areas worldwide.

Although active traveling increased during the pandemic period, due to the avoidance of crowding and public transportation, mortality of pedestrians and cyclists was generally increased, when considering the increase in exposure (ONISR - French Road Safety Observatory, 2021). As a result, measures to incorporate or increase VRUs and active travelers in the present 'car-dominated'

[1] https://etsc.eu/30km-h-limits-set-to-spread-in-2021/.

infrastructure are needed. For example, new recovery and resilience funds directed to road safety (both infrastructure and policy) need to be created in order to timely adapt to potential infrastructure changes for more active traveling. Finally, state-wide policies that enforce social responsibility, as well as boost smart speed and traffic safety enforcement are also encouraged. In that principle, data with regards to the pandemic (e.g., countermeasures in effect along with COVID-19 cases and casualties) could be integrated with safety data (e.g., speed limits and current traffic conditions) within mobile applications to inform drivers to proactively take care of their speed and driving behavior while driving on different urban and suburban areas.

9. Conclusions

The present research aimed at presenting descriptive evidence of the impact of COVID-19 and the corresponding response measures on driving behavior and road safety for the entire year of 2020. For the first time to date, data from two countries (i.e., Greece and the KSA) were explored and were correlated with the stringency of COVID-19 response measures. To fulfil that aim, following the presentation of the representative figures and overview tables, a clustering approach (i.e., k-Means) was utilized in order to identify patterns correlating the stringency of government measures with driving behavior indicators. By examining the statistics of the entire year, it was evident that the dissolution of imposed lockdown led to an increase in traffic volumes, but also to smoother driving behavior. This difference was found to be more prominent when the first lockdown in Greece and the KSA was lifted, where speeds were reduced by up to 6% in Greece and by 9% in the KSA.

Using the k-means clustering technique, it was revealed that 2020 can be split into three clusters of driving behavior: (a) baseline, one depicting driving behavior when no or light response measures apply, (b) restrictions, when COVID-19-related cases and casualties increase and thus stringency of measures increases, and (c) lockdown. The clustering results validated that the most significant differences in driving behavior of Greek drivers were found between the "Restrictions" and "Lockdown" phases in terms of exposure (total distance, total duration, and driving volume) as well as harsh braking frequency. On the contrary, negligible differences were found for speeds and speeding for all three clusters.

With regards to the policy implications of the findings, the relevant stakeholders should focus on the reduction of driving speeds, as these are indicated in the Stockholm declaration as well as the safe incorporation of active traveling modes in the current infrastructure by utilizing resilience funds and social responsibility measures.

Finally, additional crash research is needed to analyze the composition of traffic exposure, psychology, and COVID-19 as contributing factors for road safety during the pandemic period in Greece and other countries. Data for crashes per kilometer driven were inaccessible in the current research; hence, a future study could take the aforementioned indicator into consideration. As a result, future research should be directed toward analyzing crash frequency and the captured driving behavior indicators from a time-series perspective. Furthermore, a future study could expand the scope of the findings by classifying the distraction types (e.g., texting, calls, navigation) and providing insights into this topic. Psychological factors (e.g., fatigue, sadness) during the pandemic, socio-demographic characteristics (e.g., gender, age, educational level), as well as the overall effect on human psychology of the unknown future could also be taken into account. This direction will assist in better understanding the influence of the response measures as exogenous factors as well as the inter-relationship between the evolution of COVID-19 in terms of cases or casualties and the progression of driving behavior or road safety indicators.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to thank OSeven Telematics, London, UK for providing all necessary data explored to accomplish this study.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Christos Katrakazas, Apostolos Ziakopoulos; data collection: Eva Michelaraki, Marios Sekadakis; analysis and interpretation of results: Christos Katrakazas, Apostolos Ziakopoulos, Eva Michelaraki, Marios Sekadakis; draft manuscript preparation: Christos Katrakazas, Apostolos Ziakopoulos, Eva Michelaraki, Marios Sekadakis, George Yannis. All authors reviewed the results and approved the final version of the manuscript.

Appendix A

See Table A1 and see Figs. A1–A3.

Table A1
Timeline of COVID-19 response measures in Greece in Greece and KSA.

Greece	
1st Lockdown of non-essential movements	23/03/2020
End of the 1st lockdown	04/05/2020
Closure of bars, cafes, restaurants, theatres and concert halls	02/11/2020
travel Lockdown of non-essential movements	07/11/2020
Closure of primary schools and kindergartens	14/11/2020
Opening of shops and hairdressers	14/12/2020
KSA	
Closure of educational institutions	09/03/2020
Lockdown of non-essential movements in Qatif	09/03/2020
Closure of shops, restaurants, coffee shops and public parks	15/03/2020
Lockdown of non-essential movements in Mecca, Medina, Riyadh	26/03/2020
End of Lockdown	21/06/2020

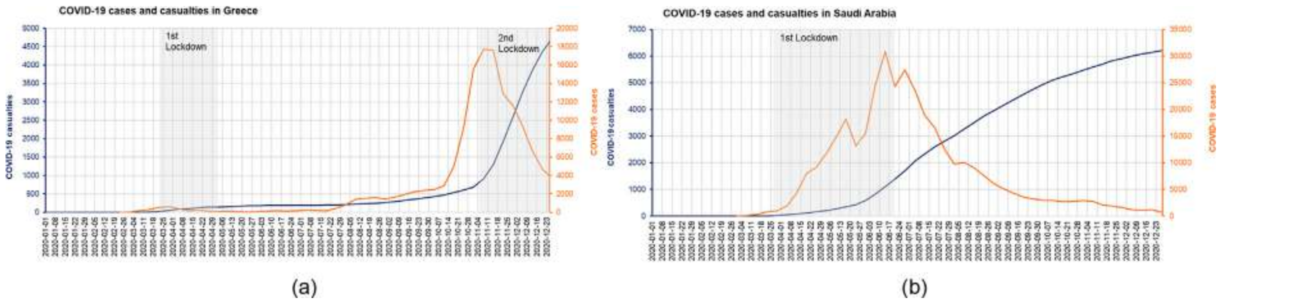


Fig. A1. The evolution of COVID-19 cases and casualties per week in (a) Greece (b) KSA (Source: Worldometer, Data Processing: NTUA).

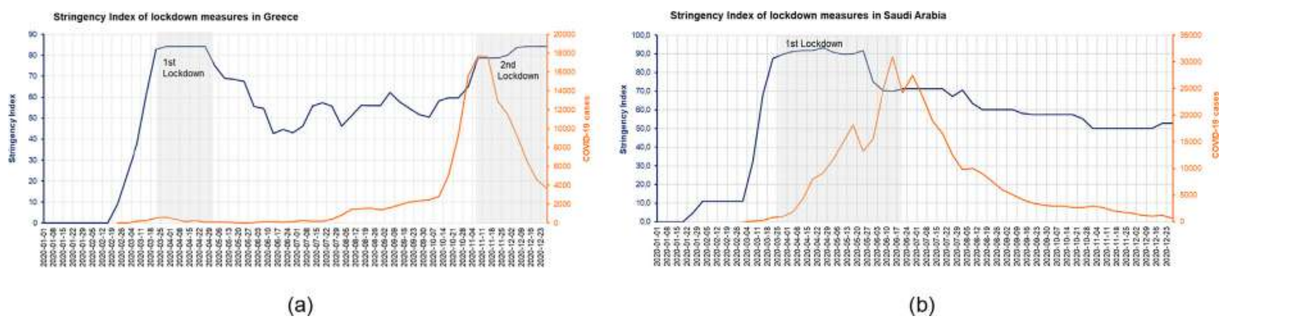


Fig. A2. Stringency index of lockdown measures per week along with the evolution of COVID-19 cases in (a) Greece (b) KSA (Source: Oxford, Data Processing: NTUA).

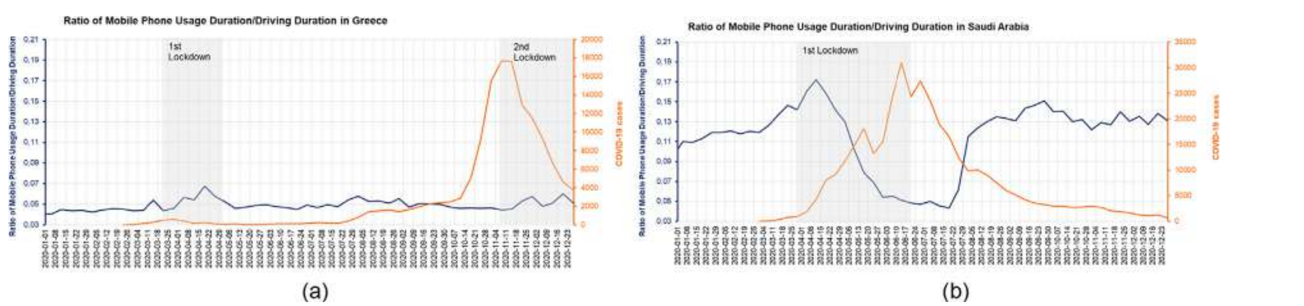


Fig. A3. Ratio of mobile phone usage duration/driving duration per week along with the evolution of COVID-19 cases in (a) Greece (b) KSA (Source: OSeven, Data Processing: NTUA).

References

Aloi, A., Alonso, B., Benavente, J., Cordera, R., Echániz, E., González, F., Ladisa, C., Lezama-Romanelli, R., López-Parra, Á., Mazzei, V., Perrucci, L., Prieto-Quintana, D., Rodríguez, A., & Sañudo, R. (2020). Effects of the COVID-19 Lockdown on Urban Mobility: Empirical Evidence from the City of Santander (Spain). *Sustainability, 12*(9), 3870. https://doi.org/10.3390/su12093870.

Barbieri, M. D., Lou, B., Passavanti, M., Hui, C., Antunes, D., Maharaj, B., Banerjee, A., Wang, F., Chang, K., Naik, B., Yu, L., Liu, Z., Sikka, G., Tucker, A., Foroutan, A., Naseri, S., Qiao, Y., Gupta, A., Abbas, M., . . . Adomako, S. (2020). A survey dataset to evaluate the changes in mobility and transportation due to COVID-19 travel restrictions in Australia, Brazil, China , Ghana , India , Iran , Italy , Norway , South Africa , United States. Data in Brief, 33, 106459. https://doi.org/10.1016/j.dib.2020.106459.

Apple, 2020. COVID 19 - Mobility Trends Reports - Apple [WWW Document]. Apple.

Bhaduri, E., Manoj, B. S., Wadud, Z., Goswami, A. K., & Choudhury, C. F. (2020). Modeling the effects of COVID-19 on travel mode choice behaviour in India. *Transportation Research Interdisciplinary Perspectives*, 100273. https://doi.org/10.1016/j.trip.2020.100273.

Calderon-Anyosa, R. J. C., & Kaufman, J. S. (2021). Impact of COVID-19 lockdown policy on homicide, suicide, and motor vehicle deaths in Peru. *Preventive Medicine, 143*(July 2020), 106331. https://doi.org/10.1016/j.ypmed.2020.106331.

Colonna, P., & Intini, P. (2020). Compensation effect between deaths from Covid-19 and crashes: The Italian case. *Transportation Research Interdisciplinary Perspectives, 6*, 100170. https://doi.org/10.1016/j.trip.2020.100170.

De Vos, J. (2020). The effect of COVID-19 and subsequent social distancing on travel behavior. *Transportation Research Interdisciplinary Perspectives, 5*, 100121. https://doi.org/10.1016/j.trip.2020.100121.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases, 20*(5), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1.

Formann, A. K. (1984). Die Latent-Class-Analyse: Einführung in Theorie und Anwendung. Weinheim [W. Germany], Beltz. http://catalog.hathitrust.org/api/volumes/oclc/16851238.html.

French Road Safety Observatory - ONISR (2021). Retrieved from https://www.itf-oecd.org/sites/default/files/france-road-safety.pdf.

Hale, T., Angrist, N., Cameron-Blake, E., Hallas, L., Kira, B., Majumdar, S., Petherick, A., Phillips, T., Tatlow, H., & Webster, S. (2020). Variation in government responses to COVID-19. BSG Working Paper Series. Blavatnik School of Government. University of Oxford, Version 8.0. www.bsg.ox.ac.uk/covidtracker%0Awww.bsg.ox.ac.uk/covidtracker%0A https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-covid-19.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics, 28*, 100–108. https://doi.org/10.2307/2346830.

Hong, J., & Mcarthur, D. (2020). Did Safe Cycling Infrastructure Still Matter During a COVID-19 Lockdown ?

Hotle, S., & Mumbower, S. (2021). The impact of COVID-19 on domestic U.S. air travel operations and commercial airport service. Transportation Research Interdisciplinary. *Perspectives, 9*(September 2020), 100277. https://doi.org/10.1016/j.trip.2020.100277.

Inada, H., Ashraf, L., & Campbell, S. (2020). COVID-19 lockdown and fatal motor vehicle collisions due to speed-related traffic violations in Japan: A time-series study. *Injury Prevention*, 98–100. https://doi.org/10.1136/injuryprev-2020-043947.

Jenelius, E., & Cebecauer, M. (2020). Impacts of COVID-19 on public transport ridership in Sweden : Analysis of ticket validations, sales and passenger counts. *Transportation Research Interdisciplinary Perspectives, 8*(June), 100242. https://doi.org/10.1016/j.trip.2020.100242.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k -means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 7*, 881–892.

Kathimerini.com. 2020. Last-minute shopping, city exodus causes traffic jams in Athens | Kathimerini. [online] Available at: <https://www.ekathimerini.com/258887/article/ekathimerini/news/last-minute-shopping-city-exodus-causes-traffic-jams-in-athens> [Accessed 11 February 2021].

Katrakazas, C., Michelaraki, E., Sekadakis, M., Ziakopoulos, A., Kontaxi, A., & Yannis, G. (2021). Identifying the impact of the COVID-19 pandemic on driving behavior using naturalistic driving data and time series forecasting. *Journal of Safety Research, 78*, 189–202.

Katrakazas, C., Michelaraki, E., Sekadakis, M., & Yannis, G. (2020). A descriptive analysis of the effect of the COVID-19 pandemic on driving behavior and road safety. *Transportation Research Interdisciplinary Perspectives, 7*, 100186. https://doi.org/10.1016/j.trip.2020.100186.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal, 1*(6), 90–95.

Kontaxi, A., Ziakopoulos, A., & Yannis, G. (2021). Investigation of the speeding behavior of motorcyclists through an innovative smartphone application. *Traffic Injury Prevention, 22*, 460–466.

Lee, M., & You, M. (2020). Psychological and Behavioral Responses in South Korea During the Early Stages of Coronavirus Disease. *International Journal of Environmental Research and Public Health, 17*(2977).

Lemke, M. K., Apostolopoulos, Y., & Sönmez, S. (2020). Syndemic frameworks to understand the effects of COVID-19 on commercial driver stress, health, and safety. *Journal of Transport and Health, 18*(May). https://doi.org/10.1016/j.jth.2020.100877.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition, 36*(2), 451–461.

Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2019). Identification of driving safety profiles from smartphone data using machine learning techniques. *Safety Science*.

Michelaraki, E., Sekadakis, M., Katrakazas, C., Ziakopoulos, A., & Yannis, G. (2021). A four-country comparative overview of the impact of COVID-19 on traffic safety behavior. In 10th International Congress on Transportation Research, Future Mobility and Resilient Transport: Transition to Innovation. ICTR 2021. September 1-3.

Muley, D., Shareef, M., Mohammad, A., & Kharbeche, M. (2021). Quantifying the impact of COVID – 19 preventive measures on traffic in the State of Qatar. *Transport Policy, 103*(February), 45–59.

Padmanabhan, V., Penmetsa, P., Li, X., Dhondia, F., Dhondia, S., & Parrish, A. (2021). COVID-19 effects on shared-biking in New York, Boston, and Chicago. Transportation Research Interdisciplinary. *Perspectives, 9*(November 2020), 100282. https://doi.org/10.1016/j.trip.2020.100282.

Papadimitriou, E., Argyropoulou, A., Tselentis, D. I., & Yannis, G. (2019). Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving. *Safety Science, 119*, 91–97.

Parady, G., Taniguchi, A., & Takami, K. (2020). Travel behavior changes during the COVID-19 pandemic in Japan: Analyzing the effects of risk perception and social in fl uence on going-out self-restriction. *Transportation Research Interdisciplinary Perspectives, 7*, 100181. https://doi.org/10.1016/j.trip.2020.100181.

Prasetijo, J., Zhang, G., Mohd, Z., Mohd, J., & Mahyeddin, E. (2021). Crash model based on integrated design consistency with low traffic volumes (due to health disaster (COVID - 19)/ movement control order). *Innovative Infrastructure Solutions, 1*–7. https://doi.org/10.1007/s41062-020-00388-7.

Qureshi, A. I., Huang, W., Khan, S., Lobanova, I., & Siddiq, F. (2020). Mandated societal lockdown and road traffic accidents. *Accident Analysis and Prevention, 146*(September), 105747. https://doi.org/10.1016/j.aap.2020.105747.

Roe, C. M., Rosnick, C. B., Colletta, A., & Babulal, G. M. (2021). Reaction to a pandemic: Social distancing and driving among older adults during COVID-19. *Journal of Applied Gerontology, 40*(3), 263–267.

Saladié, Ò., Bustamante, E., & Gutiérrez, A. (2020). COVID-19 lockdown and reduction of traffic accidents in Tarragona province. *Spain. Transportation Research Interdisciplinary Perspectives, 8*, 100218. https://doi.org/10.1016/j.trip.2020.100218.

Shamshiripour, A., Rahimi, E., Shabanpour, R., Mohammadian, A., & Kouros (2020). How is COVID-19 reshaping activity-travel behavior? Evidence from a comprehensive survey in Chicago. *Transportation Research Interdisciplinary Perspectives, 7*. https://doi.org/10.1016/j.trip.2020.100216.

Singer, M., Bulled, N., Ostrach, B., & Mendenhall, E. (2017). Syndemics and the biosocial conception of health. *The Lancet, 389*(10072), 941–950.

Singh, V., & Tech, M. (2020). Psychological Impacts on the Travel Behavior post-COVID-19. http://faculty.iitr.ac.in/~amitfce/publications.html.

Sekadakis, M., Katrakazas, C., Michelaraki, E., Kehagia, F., & Yannis, G. (2021). Analysis of the impact of COVID-19 on collisions, fatalities and injuries using time series forecasting: The case of Greece. *Accident Analysis & Prevention, 162* 106391.

Stavrakaki, A. M., Tselentis, D. I., Barmpounakis, E. N., Vlahogianni, E. I., & Yannis, G. (2019). How much driving data do we need to assess driver behavior. In *98th transportation research board annual meeting, Washington*.

Stavrinos, D., McManus, B., Mrug, S., He, H., Gresham, B., Albright, M. G., ... White, D. M. (2020). Adolescent driving behavior before and during restrictions related to COVID-19. *Accident Analysis and Prevention, 144*, 105686.

Swedish Transport Administration [STA] Annual Report (2021), STA, Borlänge.

Tselentis, D. I., Vlahogianni, E. I., & Yannis, G. (2019). Investigating the Temporal Evolution of Driving Safety Efficiency Using Data Collected from Smartphone Sensors (No. 19-03473).

Vingilis, E., Beirness, D., Boase, P., Byrne, P., Johnson, J., Jonah, B., Mann, R. E., Rapoport, M. J., Seeley, J., Wickens, C. M., & Wiesenthal, D. L. (2020). Coronavirus disease 2019: What could be the effects on Road safety? *Accident Analysis and Prevention, 144*(June). https://doi.org/10.1016/j.aap.2020.105687 105687.

Worldometer. (2020). Worldometer Coronavirus Update (Live). https://www.worldometers.info/coronavirus/.

Yannis, G., Tselentis, D. I., Vlahogianni, E. I., & Argyropoulou, A. (2017). Monitoring distraction through smartphone naturalistic driving experiment. In *6th International Naturalistic Driving Research Symposium, The Hague, Netherlands* (pp. 7–9).

Yannis, G., Papadimitriou, E., & Antoniou, C. (2007). Multilevel modeling for the regional effect of enforcement on road accidents. *Accident Analysis & Prevention, 39*(4), 818–825.

Ms. Eva Michelaraki is a Civil Engineer, PhD Candidate and Researcher at the Department of Transportation Planning and Engineering at the School of Civil Engineering of the National Technical University of Athens (NTUA). Since 2019, she holds a Civil Engineering Diploma from NTUA majoring in Transportation Engineering. Her PhD research is on "Road accident risk factors and big data". She has more than 2 years of experience in several aspects of traffic engineering and road safety. She has participated in 3 research projects and has published 14 scientific papers (5 in scientific journals and 9 in scientific conference proceedings after peer review).

Mr. Marios Sekadakis is a Civil – Transportation Engineer, PhD Candidate and Research Associate at the Department of Transportation Planning and Engineering at the School of Civil Engineering of the National Technical University of Athens (NTUA). In 2020, he graduated in Civil Engineering from the National Technical University of Athens (NTUA) majoring in Transportation Engineering. He has more than one year of experience in several aspects of traffic engineering and road safety.

Dr. Christos Katrakazas is a Research Associate at the Department of Transportation Planning and Engineering at the School of Civil Engineering of the National Technical University of Athens (NTUA). He holds a Civil Engineering Diploma from NTUA majoring in Transportation Engineering (2013) and a Ph.D. in safety of autonomous vehicles from Loughborough University (2017). After finishing his doctoral studies, he worked as a Research Associate at Loughborough University, and the Chair of Transportation Systems Engineering at the Technical University of Munich (TUM), Germany. He has more than 6 years of experience in several aspects of transport safety and engineering.

Dr. Apostolos Ziakopoulos, PhD, MSc DIC is a Research Associate at the Department of Transportation Planning and Engineering at the School of Civil Engineering of the National Technical University of Athens (NTUA). He holds a Civil Engineering Diploma from NTUA majoring in Transportation Engineering (2013). He also holds a Master of Science – DIC in Transport from Imperial College London and University College London (2014). His Ph.D. dissertation was in spatial analysis of road safety and traffic behavior from NTUA (2020). He has worked as a researcher for the University of Sheffield through the I.A.E.S.T.E. scheme. His main research interests involve road crash analyses, statistical and spatial modelling and traffic engineering. He has participated in 8 research and engineering projects and studies and has published 55 papers in scientific journals, textbooks and conferences.

Prof. George Yannis is a Professor in Traffic Safety and Management with particular focus on data management and analysis and currently Director of the Department of Transportation Planning and Engineering of the School of Civil Engineering at the NTUA. He has contributed extensively in more than 245 research and engineering projects and studies and in several scientific committees of the European Commission and other International Organizations (UN/ECE, OECD, WHO, World Bank, EIB, CEDR, ERF, UITP, ETSC, ECTRI, WCTR, TRB). He has published more than 560 scientific papers (169 in scientific journals) widely cited worldwide.

Journal of Safety Research 84 (2023) 404–410

Contents lists available at ScienceDirect

Journal of Safety Research

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

Predicting behavioral intentions for unsafe off-highway vehicle use

Brian J. Pugliese [1], Benjamin K. Barton, Gerardo Lopez

University of Idaho, United States

ARTICLE INFO

Article history:
Received 10 November 2021
Received in revised form 16 September 2022
Accepted 2 December 2022
Available online 12 December 2022

Keywords:
Theory of Planned Behavior
Off-highway vehicles
ATV
Safety
Behavioral control

ABSTRACT

Introduction: Hundreds of adults are killed or injured each year while operating off-highway vehicles. Four common risk-taking behaviors were identified on off-highway vehicles in the literature and examined intention to engage in such behaviors within the context of the Theory of Planned Behavior. *Method:* One hundred and sixty-one adults completed measures of experience on off-highway vehicles and injury exposure followed by a self-report created according to the predictive structure of the Theory of Planned Behavior. Behavioral intentions to engage in the four common injury risk behaviors on off-highway vehicles were predicted. *Results*: Similar to research on other risk behaviors, perceived behavioral control and attitudes emerged as consistently significant predictors. Subjective norms, the number of vehicles operated, and injury exposure showed varying relationships to the four injury risk behaviors. Results are discussed in the context of similar studies, intrapersonal predictors of injury risk behaviors, and with regard to implications for injury prevention efforts.

1. Introduction

Unintentional injury during off-highway vehicle (OHV) use is a significant transportation safety problem. In 2016, there were estimated 591 fatalities during OHV use and 101,200 injuries requiring treatment among users of all ages (U.S. Consumer Product Safety Commission, 2020). A total of 15,744 fatalities among users of all ages were reported from 1982 to 2018, or an average of 438 each year (U.S. Consumer Product Safety Commission, 2020). OHV-related injuries are especially a problem among younger users. For example, according to one source, an estimated 361,161 children under age 16 were treated for OHV-related injuries in 2013 (Shults, West, Rudd, & Helmkamp, 2013). The large annual number of injuries requiring medical attention highlights the need for understanding the etiology of risk-taking behaviors while using OHVs. The goal of this study was to identify common risk-taking behaviors on OHVs reported in the literature, then examine intention to engage in such behaviors within the context of the Theory of Planned Behavior.

1.1. Injury risk behaviors on off-highway vehicles

OHVs comprise various vehicles such as off-road motorbikes (i.e., dirt bikes), three-wheelers, four-wheelers, and utility vehicles. Many OHVs are powerful and capable of reaching speeds comparable to those of a regular automobile. Further still, off-highway vehicles generally do not afford protection during a crash. Injuries sustained while using an OHV can be significant, including fractures of lower limbs and intracranial damage (Helmkamp, Furbee, Coben, & Tadros, 2008). Despite the risk potential, several types of risk behaviors while operating OHVs remain common.

The literature on OHVs is primarily epidemiological and lacking in theoretically-driven research but has identified clear risk factors for severe injury. Evidence suggests three of the leading behaviors contributing to fatal crashes are lack of helmet use, having passengers on an OHV built for only one person to use, and riding while impaired by alcohol or other substances (Balthrop, Nyland, & Roberts, 2009). For example, one study examined 112 OHV crashes and found 85 % of operators who experienced a fatal injury were not wearing a helmet, and 50 % were intoxicated (Hall, Bixler, Helmkamp, Kraner, & Kaplan, 2009). A fourth common risk behavior associated with injury while operating OHVs is riding on roads intended for regular vehicle use (Williams, Oesch, McCartt, Teoh, & Sims, 2014). Research suggests operating an OHV on any road, paved or unpaved, increases injury risk significantly beyond that associated with off-road riding (Denning & Jennissen, 2016). OHV crashes on roads may involve a second vehicle, significantly

[1] Communication can be directed to Brian Pugliese, The Virginia Tech Transportation Institute, Virginia Tech, 3500 Transportation Research Plaza, Blacksburg, Virginia, 24061.

E-mail address: bpugliese@vtti.vt.edu (B.J. Pugliese)

https://doi.org/10.1016/j.jsr.2022.12.003
0022-4375/

increasing the severity of trauma (Denning, Jennissen, Harland, Ellis, & Buresh, 2013).

Riding on roads often occurs in conjunction with other risk factors such as carrying passengers and intoxication (Denning, Harland, Ellis, & Jennissen, 2013). For example, one study found 92 % of OHV riders carried passengers, 81 % operated their OHV on public roads, 64 % rode without a helmet, but 60 % reported engaging in all three risk behaviors at the same time (Jennissen et al., 2014). Risk behaviors while operating OHVs are therefore likely to co-occur.

Despite evidence pointing to a common set of risk factors, little is known about why OHV users engage in such risk behaviors. The lack of knowledge of the reasons behind recurring risk behaviors among OHV users can be addressed by examining the underpinning of such behaviors. One useful theoretical structure for examining health and injury risk behaviors is the Theory of Planned Behavior.

1.2. Theory of planned behavior

The Theory of Planned Behavior examines the intention to engage in behaviors by examining subjective norms, attitudes, and perceived behavioral control and has been used extensively in health research (Sleet, Diekman, Ikeda, & Carlson Gielen, 2010). Eating habits, exercise, lifestyle, sexual health, drug and alcohol use, and hygiene all have been studied using the Theory of Planned Behavior as a framework (Albarracin, Johnson, Fishbein, & Muellerleile, 2001; Armitage, Armitage, Conner, Loach, & Willetts, 1999; Blue, 1995; Conn, Tripp-Reimer, & Maas, 2003; Courneya & McAuley, 1995; Downs & Hausenblas, 2005; Fila & Smith, 2006; French et al., 2005; Godin & Kok, 1996; Guo et al., 2007; Norman & Conner, 2006; O'Boyle, Henly, & Larson, 2001; Rhodes & Courneya, 2003; Schifter & Ajzen, 1985; Sheeran & Taylor, 1999). One meta-analytic review showed at least 185 studies have employed the theory (Armitage & Conner, 2001), and many more have applied the Theory of Planned Behavior in the 17 years since the last meta-analytic investigation.

In practice, the Theory of Planned Behavior employs self-report questionnaire measures of attitudes, subjective norms, and perceived behavioral control as predictors of intentions to engage in health risk behaviors (Ajzen, 1985, 1991). A review of studies suggested attitudes and subjective norms accounted for approximately 27 % of the variance in behavior and 39 % in intention, and perceived behavioral control explained a significant additional amount of variance (Armitage & Conner, 2001). Behavioral intention, taken to be the antecedent of actual risk behaviors, makes the theory valuable for understanding psychological processes leading to increased risk regardless of the health topic in question.

The Theory of Planned Behavior has been applied repeatedly in studies of injury etiology (e.g., Trifiletti, Gielen, Sleet, & Hopkins, 2005). Topics such as bike safety, seatbelt usage, and other driving violations have been examined (Conner et al., 2007; Forward, 2009; Lajunen & Rasanen, 2004; Parker, Manstead, Stradling, Reason, & Baxter, 1992; Quine, Rutter, & Arnold, 1998, 2001; Şimşekoğlu & Lajunen, 2008; Warner & Åberg, 2006). A more recent study examined pedestrians' intentions to cross streets under conditions of distraction (Barton, Kologi, & Siron, 2016).

The pattern of predictive utility of the three components in the Theory of Planned Behavior appears to vary across studies. Previous inquiries suggest perceived behavioral control is the strongest predictor of behavioral intentions (Armitage & Conner, 2001), and this was found in at least three studies of risk behaviors (Johnson & Hall, 2005; Evans & Norman, 1998; Zhou, Horrey, & Yu, 2009). However, several other studies found all three components predicted behavioral intentions (Diaz, 2002; Malekpour, Moeini, Tapak, Sadeghi-Bazargani, & Rezapur-Shahkolai, 2021; Zhou & Horrey, 2010). Still other research on injury behavior found only attitude (Holland & Hill, 2007) or both attitude and perceived behavioral control predicted behavioral intention (Barton et al., 2016). One explanation for the variability of the component's predictability may be the differences in behaviors being examined.

The application of the Theory of Planned Behavior suggests great utility for examining health and safety topics. Research from the past two decades also suggests the components in the theory may perform differently across behavior types. Although the usefulness of the theory has been demonstrated in studies of other topics, the theory has never been applied to risk behaviors on OHVs. To our knowledge, the present study is the first to use *any* theoretical framework in research addressing risk behaviors during OHV use.

1.3. Aim and hypotheses

The aim of this study was to examine intentions to engage in risky behaviors on OHVs guided by the Theory of Planned Behavior framework. The expected results aligned both with the structure of the theory and with applications of the theory in other studies. Attitudes, subjective social norms, and perceived behavioral control were expected to significantly predict intentions to engage in four common types of risky behavior identified by the literature: riding without a helmet, riding with passengers, riding on paved roads, and riding while intoxicated. Consistent with the literature, perceived behavioral control was expected to emerge as a significant and positive predictor of intention to engage in risk behaviors on OHVs.

Research using the Theory of Planned Behavior to examine other kinds of injury risk behaviors has found an inconsistent prediction pattern among the theory's components. In addition, the theory has never been used to help explain risk behaviors on OHVs. Therefore, it was expected that the pattern of the predictive utility of the three components may differ from other research or may change across the four risk behaviors.

2. Method

2.1. Sample

One hundred and sixty-one adults ages 18 to 48 (M = 19.70; SD = 3.50; 35 % male) were recruited from the undergraduate population at a university in the Pacific Northwest. Corresponding to the demographic characteristics of the local population, the sample was primarily Caucasian (81 %) but also included participants identifying as Hispanic (10 %), Asian (5 %), African American (2 %), and Native American (2 %). All participants had experience operating some form of OHV. The study was approved by the university's Institutional Review Board.

2.2. Measures and Procedure

Demographic information and OHV experience. Participants first reported age, sex, and ethnicity. Next, participants reported their frequency of use of OHVs, types of vehicles operated, and the first age at which they operated an OHV. Third, participants reported injury experience related to operating OHVs: whether they had been injured or had a "close call" (nearly being injured) while operating an OHV; whether they knew anyone who had been injured while operating an OHV requiring a stay in a hospital; whether they knew anyone who had been injured while operating an OHV who was treated and released; or whether they knew anyone who had been injured while operating an OHV and received only untreated scrapes and bruises. The number of "yes" answers to the five questions was tallied to create an injury experience score.

Theory of Planned Behavior Questionnaire. Consistent with the literature, a self-report instrument was used to measure risky behaviors guided by the structure of the Theory of Planned Behavior. The questionnaire was organized following published guidelines (Ajzen, 2013). Preparation of a questionnaire typically involves creating hypothetical scenarios and new response items (e.g., Barton et al., 2016; Evans & Norman, 1998, Holland & Hill, 2007; Zhou & Horrey, 2010). In this study, participants read and responded to scenarios depicting risk behaviors identified in the literature: riding without a helmet (A), riding with unsecured passengers (B), riding on paved roads (C), and riding while intoxicated (D). All scenarios are presented in Appendix A. Twelve questions, based on the three predictor variables (attitudes, subjective norms, and perceived behavioral control) followed the presentation of each scenario to the participant. Responses to all questions were scored on a Likert scale ranging from one to seven. The descriptive statistics for the Theory of Planned Behavior components, including behavioral intention, can be found in Table 1.

The first four questions measured *attitudes* toward the behavior presented in the scenario. Questions measuring attitudes were constructed in two pairs (i.e., 1 & 2, 3 & 4), with the first question in each pair measuring strength of response and the second question measuring evaluation of the behavior. Questions 5–8 measured *subjective norms* (i.e., the participant's indication of what they thought other relevant individuals would want them to do). As with attitudes, questions in this section were organized in pairs, with the first question in each pair measuring the strength of response and the second question measuring motivation to comply. Reference groups chosen for subjective norms were friends and other OHV users. Questions 9 and 10 measured *perceived behavioral control*. Finally, responses to the last two questions indicated *behavioral intention*. Similar to previous versions of TPB measures (e.g., Ajzen, 2013), a larger number of items were employed for the constructs of attitude and subjective norms.

Scores were calculated generally by multiplying items, with some items being reverse scored prior to multiplication. Calculation for attitudes and subjective norms proceeded in two steps. First scores from questions concerning the strength of response and evaluation of the behavior were multiplied, yielding two sub-scores for attitudes and two for subjective norms. For attitudes, question 1 multiplied by question 2 reverse-scored; 3 multiplied by 4 reverse-scored. For subjective norms, question 5 multiplied by question 6; 7 multiplied by 8. Second, the two sub-scores were averaged within each measure to create an attitude and a subjective norm score. The perceived behavioral control score was calculated by reverse-scoring item 9, then multiplying item 9 by item 10. Items 11 and 12 were multiplied to calculate a score for behavioral intention.

Calculations were performed separately for the four scenarios. Aggregate variables were then created by averaging across variables generated in the four scenarios. For example, attitude scores across the four scenarios were averaged. Final aggregates represented belief-based measures of attitudes and subjective norms in which higher scores indicated greater strength. Higher scores for perceived behavioral control indicated greater perception of control if the behavior were to be attempted. Higher scores for the aggregate behavioral intention represented greater intention to engage in the type of behaviors outlined in the four scenarios.

Procedure. Participation proceeded in two steps. First, the informed consent process began. Second, participants completed the Theory of Planned Behavior questionnaire.

Analyses. Analyses comprised four steps. First, sex differences were examined across all Theory of Planned Behavior components in a series of ANOVA to determine if sex would be included in subsequent analyses. Second, participant OHV experience and injury exposure were examined. Third, a set of correlation matrices were performed to determine what variables should be retained as predictors for the final regression. Finally, a regression analysis examined significantly correlated variables as predictors of behavioral intention.

3. Results

3.1. Sex differences

Sex differences were examined across all Theory of Planned Behavior components from each OHV scenario in a series of ANOVAs. No statistically significant differences between males and females were found. Sex was excluded from further analyses.

3.2. Self-Reported OHV experience and exposure to injury

All participants reported having used OHVs. Among the participants, 118 reported riding once or twice per year, 29 reported riding once per month, 9 reported riding once per week, and 5 participants reported riding daily. Participants reported using a variety of OHVs: four-wheelers (131), dirt bikes (68), snow mobiles (35), three-wheelers (14), rhinos (15), snow cats (6), dune ATV's (20), rally cars (6), and side-by-sides (21). Participants reported being an average age of 10.78 years ($SD = 4.11$) the first time they operated an OHV, ranging from 3 years of age to 21.

Some participants reported first-hand experience with injury on an OHV. Forty participants reported having been injured, and 90 reported having had a "close call." One hundred and twenty-three reported knowing someone who had been hospitalized following an injury on an OHV, 113 reported knowing someone who was treated and released, and 154 reported knowing someone who experienced minor injuries not requiring treatment.

3.3. Theory of Planned Behavior prediction of behavioral intention

Examination of predictors of intentions to engage in risky behaviors on OHVs proceeded in several stages. First, correlations between potential predictors and behavioral intentions were examined within each scenario to identify predictors for regression analyses. Variables were retained for examination as predictors if significantly correlated to behavioral intentions within each scenario. Correlations between predictors and behavioral intentions by scenario are reported in Table 2. Second, selected predictors were regressed on behavioral intentions for each risk behavior scenario using data from the entire sample.

Behavioral intentions were regressed on relevant predictors identified in Pearson correlations. Regression results are shown in Table 3. Predictors in each model accounted for 37 % of the variance in behavioral intention concerning helmet use, 40 % in riding with passengers, 41 % in riding on paved roads, and 15 % in riding while intoxicated.

Perceived behavioral control and attitudes again emerged as statistically significant positive predictors of behavioral intentions in all scenarios except for helmet use, in which the predictions were negative. Subjective norms presented as a significant positive predictor of behavioral intentions in all the scenarios except for helmet use, in which the predictive nature was non-significant. Finally, the number of vehicles operated was found to be a significant positive predictor of behavioral intention.

Table 1
Descriptive statistics for Theory of Planned Behavior components for each scenario.

	A	B	C	D	**Total**
Component	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Attitudes	20.84 (11.97)	23.51 (11.31)	28.97 (12.57)	10.20 (8.54)	20.88 (7.70)
Subjective Norms	19.19 (10.89)	19.74 (11.22)	17.65 (9.77)	6.00 (5.51)	15.64 (6.98)
Perceived Behavioral Control	11.42 (12.32)	17.08 (14.92)	15.48 (13.92)	6.39 (7.83)	12.59 (8.97)
Behavioral Intention	35.16 (16.26)	21.40 (17.79)	21.70 (17.72)	3.63 (7.60)	20.47 (8.14)

Table 2
Pearson Correlations between potential predictor variables and components of Theory of Planned Behavior.

Variables	2	3	4	5	6	7	8
Scenario				A B C D	A B C D	A B C D	A B C D
1. Age	0.02	0.06	-0.17*	0.21†/-0.08/-0.01/-0.06	0.21†/0.19*/0.08/-0.04	0.26†/0.10/0.04/-0.08	-0.14/0.01/0.05/-0.06
2. Number of vehicles operated		0.34†	-0.18*	-0.04/-0.11/0.04/-0.04	-0.03/0.11/0.13/0.20†	-0.21†/-0.14/-0.13/0.16*	0.11/0.11/0.30†/0.04
3. Injury exposure			-0.04	-0.04/0.02/0.01/-0.07	-0.07/0.13/0.11/0.05	0.11/0.11/0.19*/-0.07	-0.01/0.15/0.28†/-0.08
5. Attitude					0.11/0.07/0.15/0.13	0.38†/0.31†/0.25†/0.08	-0.49†/0.42†/0.43†/0.20†
6. Subjective norms						0.12/0.17*/0.13/0.17*	-0.19*/0.27†/0.35†/0.22†
7. Perceived control							-0.51†/0.54†/0.35†/0.30†
8. Behavioral intentions							

Note. $N = 161$. * $p < .05$. † $p < .01$. A = helmet use, B = paved roads, C = passengers, D = intoxication.

Table 3
Linear regression predicting behavioral intentions by scenario.

Predictors	*B*	*SE*	β	R^2
Scenario A: Helmet use				
Attitudes	-0.46	0.09	-0.34**	
Subjective norms	-0.17	0.10	-0.11	
Perceived control	-0.48	0.09	-0.36**	0.37**
Scenario B: Riding with passengers				
Attitudes	0.43	0.10	0.27**	
Subjective norms	0.28	0.10	0.18**	
Perceived control	0.52	0.08	0.43**	0.40**
Scenario C: Riding on paved roads				
Number of vehicles operated	3.26	0.88	0.25**	
Injury Exposure	1.44	0.89	0.11	
Attitudes	0.46	0.09	0.33**	
Subjective norms	0.40	0.12	0.22**	
Perceived control	0.32	0.09	0.25**	0.41**
Scenario D: Riding while intoxicated				
Attitudes	0.14	0.07	0.16*	
Subjective norms	0.21	0.10	0.16*	
Perceived control	0.25	0.07	0.26**	0.15**

Note. $N = 161$. * $p < .05$. ** $p < .01$.

4. Discussion

The framework provided by the Theory of Planned Behavior was used to examine intentions to engage in four common risk behaviors on OHVs: riding without a helmet, riding with passengers, riding on paved roads, and riding while intoxicated. Consistent with the literature, a questionnaire was constructed and used to assess behavioral intention. It was expected that all components of the Theory of Planned Behavior would be useful in predicting intentions to engage in risky behavior but perceived behavioral control would emerge as a strong and consistent predictor. It was also expected that experience operating OHVs and exposure to OHV-related injury might be predictive of behavioral intention.

Components of the Theory of Planned Behavior proved useful for predicting intentions to engage in the risk behaviors that was identified in the literature. However, as suggested by other studies, the usefulness of each component varied by risk behavior. Perceived behavioral control emerged as a consistently significant predictor. Attitudes also significantly predicted behavioral intentions across all of the common risk behaviors. Subjective norms were also significantly predictive of behavioral intention across all scenarios, except helmet use. Injury exposure was not significantly predictive of behavioral intention in any scenario. Finally, OHV experience was significantly predictive of behavioral intention while riding on paved roads and riding with passengers.

4.1. Perceived Behavioral Control

Perceived behavioral control significantly predicted behavioral intentions across scenarios. The results are similar to other studies that have used the Theory of Planned Behavior in the examination of pedestrian behavior (Evans & Norman, 1998, 2003; Holland & Hill, 2007; Zhou et al., 2009, 2010; Xu, Li, & Zhang, 2013). The strong and consistent relation of perceived behavioral control to behavioral intentions in these data may be due to the nature of the types of OHV use described in the scenarios in this study. All four scenarios described common behaviors on OHVs. Previous researchers (Barton et al., 2016; Evans & Norman, 1998) have posited perceived behavioral control is a significant predictor of intention to engage in mundane types of behavior, partly because such behaviors are perceived to be easier to execute and thus any associated risk is minimal. Alternatively, perceived behavioral control could tap into a sense of efficacy concerning the risk behavior (Barton, Davis, & Pugliese, 2021).

The results of this study concerning perceived behavioral control are consistent with other research concerning risky behavior while operating vehicles. For example, other researchers have found perceived behavioral control to be predictive of speeding (Conner et al., 2007; Parker et al., 1992; Warner & Åberg, 2006), intent to commit violations such as intoxicated driving, close following, and risky passing (Parker et al., 1992). Collectively, our results and evidence in the literature highlight the importance of perceived behavioral control as a predictor of unsafe operation of vehicles and for transportation safety.

Perceived behavioral control significantly and negatively predicted behavioral intention in the helmet use scenario. Participants who responded that they would have no problem riding without a helmet and that wearing a helmet was largely up to them tended to rate their intention to wear a helmet in our scenario as lower. Two previous studies that used the Theory of Planned Behavior found perceived behavioral control to be positively predictive of wearing a helmet while bicycling (Lajunen & Räsänen, 2004; O'Callaghan & Nausbaum, 2006). At first glance, our result seems unusual. However, previous work concerned bicycling among teenagers, and no evidence exists to which we might directly compare our results concerning OHV use. A greater sense of control or agency may very well be predictive of not using a helmet for this form of transportation. Regardless, our result suggests perception of control is indeed a relevant factor for helmet use among OHV operators.

4.2. Attitudes

Attitudes were significantly predictive of behavioral intention across all scenarios, negatively predictive of riding without wearing a helmet and positively predictive of behaviors in all other scenarios. We offer a few interpretations of this result. OHV operators may view some use of roads as necessary depending on how vast an area of the terrain is being crossed. Roads also afford opportunities to ride faster, constituting an important environmental context for injury risk behavior of this type (Barton, Davis, & Pugilese 2021). Perhaps operators viewed riding on paved roads as an opportunity to have fun by riding faster than they can on rough terrain. In fact, other research found the attitude component of the Theory of Planned Behavior was a significant predictor of speeding (Conner et al., 2007). More broadly speaking, use of OHVs may simply be often used in a social context supporting riding on paved roads, with passengers, or while intoxicated. Future work could investigate the social context surrounding OHV use for evidence of such social support.

4.3. Subjective Norms

Subjective norms was a significant positive predictor of behavioral intention for all examined risk behaviors, except helmet use. Similar to the predictive pattern of attitudes, subjective norms showed a different pattern only in relation to riding without wearing a helmet. We posit that perhaps the broader social context is at play, as suggested above for the pattern of results for attitudes. The four risk behaviors in our scenarios were chosen because they are statistically common. Research suggested that, while participants understand certain behaviors are unsafe, the behaviors are not seen as deviating from moral or ethical standards (Barton et al., 2016; Evans & Norman, 2003). OHV operators may know the behaviors are unsafe, but the behaviors are common and seen as excusable, thus reducing the importance of subjective norms for predicting intention to engage in risk behaviors. Perhaps riding without wearing a helmet is the exception to the influence of the social context surrounding OHV use (i.e., engaging in the other three behaviors is acceptable to some degree, but riding without a helmet is not).

4.4. Injury Exposure and OHV Experience

Injury exposure was not significantly related to any of the examined risk behaviors. As with some of our findings, no context exists in which to evaluate injury exposure on OHVs in relation to behavioral intention (i.e., in the context of the Theory of Planned Behavior). However, we may offer at least one explanation. Perhaps users who have been injured or have seen friends injured associated the other three risk behaviors more strongly with injury likelihood. A roadway may appear a less likely context for potential injury. At least in our scenario, the terrain would be assumed to be level or smooth, no passengers are present, and one is not intoxicated. The user would simply be traveling between points and need only be wary of other vehicles.

OHV experience was significantly predictive for riding on paved roads among only those with OHV experience. One reason experience may not have been a useful predictor of intended risk behaviors in our data is the nature of our measure. Our measure of experience was blunt; merely the self-reported number of vehicles operated. A more comprehensive measure of OHV experience might include not only the number of vehicles operated but also how much training a person has had (if any), the types of terrain on which the person usually rides, and the purpose behind OHV use (e.g., farm, recreational riding, hunting). Another more comprehensive measure of experience might also include some report of successes versus failures concerning the risk behavior (Barton, Davis, & Pugliese, 2021). Alternatively, another explanation might be the social context surrounding OHV use overrides the effect of previous experience (including injury) on avoidance of certain risk behaviors.

4.5. Limitations

Several limitations are worth mentioning. First, a questionnaire was used which, although common in the literature, is not predictive of actual behavior on OHVs. Extending intention to behavior is conjecture. The frequency of risk behaviors on OHVs and actual risk of injury among participants could not be determined. However, the intention was to apply the Theory of Planned Behavior to common risk behaviors among OHV users, not to predict risk behaviors or measure increments in actual risk.

Second, the sample did not comprise demographic variations in the national population or include participants under the age of 18. The usefulness of the components in the Theory of Planned Behavior as predictors of risk taking among OHV operators may vary somewhat in a nationally representative sample. Evidence also may vary between samples drawn from various rural areas or between samples of operators who use OHVs for specific purposes (e.g., farm or law enforcement use). Operators under age 18 are at significant risk for injury on OHVs. More research is needed to demonstrate any predictive utility of perception of behavioral control (or other predictors) among juvenile operators.

Third, future work might investigate how much perceived risk or danger participants associate with the tasks in the questionnaire. Previous work (Barton et al., 2016) examined the rather mundane task of street crossing, and yet the Theory of Planned Behavior predicted intention to engage in risky behavior. Perception of associated injury risk might be much higher for other types of risk behaviors and could moderate predictive utility of the Theory of Planned Behavior.

Finally, scenarios in this study did not include co-occurring risk behaviors. The four risk behaviors predominant in the literature occur together in real settings (Jennissen et al., 2014). Our assessment focused on each risk behavior separately, but future efforts should consider the additive or interactive implications of co-occurring risk behaviors on OHVs.

4.6. Implications

Theoretical guidance is missing from the study of risk behaviors on OHVs. These results offer a glimpse of the usefulness afforded by the guidance of theoretical structures applied to OHV safety. To our knowledge, no other studies have yet applied any theoretical framework to risk taking behaviors on OHVs. Many avenues of inquiry into OHV safety remain to be explored, and implications for injury prevention wait to be developed. The primary suggestion for using knowledge gained from the application of the Theory of Planned Behavior concerns the role of perceived behavioral control.

Researchers might consider the greater implications of perception of behavioral control as a predictor of risk taking on OHVs for injury prevention efforts and creation of policy. Efforts to reduce risk taking behaviors may take many forms. One possible method is to target the perception of control over injury risk through efforts such as public service announcements about common risk behaviors during peak months of OHV usage. Citizen contacts by law enforcement (e.g., traffic stops and regular patrol activities) are another method. The *Alive at 25* program already uses traffic stops as one way to advertise the availability of safety courses. Similarly, points of contact between law enforcement and citizens offer a way to convey safety messages about OHV use. For example, fish and game officers often encounter people using OHVs for hunting and recreation. Finally, yet another suggestion is packaging OHV safety training as part of courses citizens are generally required to take, such as hunter safety. Through these, and potentially other methods, stakeholders in OHV safety may more effectively raise awareness among the public about OHV risk behaviors, their outcomes, and simple ways to prevent serious injury.

Conflict of Interest

The authors have had no conflicts of interest concerning the preparation of this manuscript. This research did not receive any external funding.

Appendix A

A. You and several friends are going out to ride trails in the woods for the afternoon. Some people wear helmets when they ride, some don't. You have only a moment to decide whether to wear your helmet, and you choose to leave it at home.
B. You're visiting family who live outside of town. You and several family members, including two children under 10, want to spend the afternoon riding the skid roads in the woods near the family home. There are only two ATV's and five people who want to ride. Although the seat on your ATV is made for only one person, you decide to let one or two of the children ride with you.
C. You and your friend live several miles out of town. Your homes are about a mile apart. This afternoon, you are about to go to your friend's house to hang out. The road to your friend's house is paved and you see traffic on the road every day. Even though the road is used by motorists, you decide to ride your ATV and take the road to your friend's house.
D. You and some friends are camping. Every-one was up late last night drinking and some are still intoxicated. Now it's 8:00am and you're thinking about going into town for breakfast. A couple of people brought ATV's along. You can still feel the effects of drinking last night, but you can walk and speak with no problem. You decide to ride into town.

References

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 11–39). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211.

Ajzen, I. (2013). Theory of Planned Behaviour Questionnaire. Measurement Instrument Database for the Social Science. Retrieved from www.midss.ie.

Albarracin, D., Johnson, B. T., Fishbein, M., & Muellerleile, P. A. (2001). Theories of reasoned action and planned behavior as models of condom use: A meta-analysis. *Psychological Bulletin, 127*(1), 142.

Armitage, C. J., Armitage, C. J., Conner, M., Loach, J., & Willetts, D. (1999). Different perceptions of control: Applying an extended theory of planned behavior to legal and illegal drug use. *Basic and Applied Social Psychology, 21*(4), 301–316.

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British journal of social psychology, 40*(4), 471–499.

Balthrop, P. M., Nyland, J., & Roberts, C. S. (2009). Risk factors and musculoskeletal injuries associated with all-terrain vehicle accidents. *The Journal of emergency medicine, 36*(2), 121–131.

Barton, B. K., Kologi, S. M., & Siron, A. (2016). Distracted pedestrians in crosswalks: An application of the theory of planned behavior. *Transportation research part F: Traffic psychology and behaviour, 37*, 129–137.

Barton, B. K., Davis, S. J., & Pugliese, B. J. (2021). A Risk Appraisal Framework for Injury Etiology. Health Promotion Practice, 15248399211018167.

Blue, C. L. (1995). The predictive capacity of the theory of reasoned action and the theory of planned behavior in exercise research: An integrated literature review. *Research in Nursing & Health, 18*(2), 105–121.

Conn, V. S., Tripp-Reimer, T., & Maas, M. L. (2003). Older women and exercise: Theory of planned behavior beliefs. *Public Health Nursing, 20*(2), 153–163.

Conner, M., Lawton, R., Parker, D., Chorlton, K., Manstead, A. S., & Stradling, S. (2007). Application of the theory of planned behaviour to the prediction of objectively assessed breaking of posted speed limits. *British Journal of Psychology, 98*(3), 429–453.

Courneya, K. S., & McAuley, E. (1995). Cognitive mediators of the social influence-exercise adherence relationship: A test of the theory of planned behavior. *Journal of Behavioral Medicine, 18*(5), 499–515.

Denning, G. M., & Jennissen, C. A. (2016). All-terrain vehicle fatalities on paved roads, unpaved roads, and off-road: Evidence for informed roadway safety warnings and legislation. *Traffic injury prevention, 17*(4), 406–412.

Denning, G. M., Harland, K. K., Ellis, D. G., & Jennissen, C. A. (2013). More fatal all-terrain vehicle crashes occur on the roadway than off: Increased risk-taking characterises roadway fatalities. *Injury prevention, 19*(4), 250–256.

Denning, G., Jennissen, C., Harland, K., Ellis, D., & Buresh, C. (2013). All-terrain vehicles (ATVs) on the road: A serious traffic safety and public health concern. *Traffic injury prevention, 14*(1), 78–85.

Downs, D. S., & Hausenblas, H. A. (2005). Elicitation studies and the theory of planned behavior: A systematic review of exercise beliefs. *Psychology of Sport and Exercise, 6*(1), 1–31.

Fila, S. A., & Smith, C. (2006). Applying the theory of planned behavior to healthy eating behaviors in urban Native American youth. *International Journal of Behavioral Nutrition and Physical Activity, 3*(1), 11.

Forward, S. E. (2009). The theory of planned behaviour: The role of descriptive norms and past behaviour in the prediction of drivers' intentions to violate. *Transportation Research Part F: Traffic Psychology and Behaviour, 12*(3), 198–207.

French, D. P., Sutton, S., Hennings, S. J., Mitchell, J., Wareham, N. J., Griffin, S., ... Kinmonth, A. L. (2005). The Importance of Affective Beliefs and Attitudes in the Theory of Planned Behavior: Predicting Intention to Increase Physical Activity1. *Journal of Applied Social Psychology, 35*(9), 1824–1848.

Godin, G., & Kok, G. (1996). The theory of planned behavior: A review of its applications to health-related behaviors. *American Journal of Health Promotion, 11*(2), 87–98.

Guo, Q., Johnson, C. A., Unger, J. B., Lee, L., Xie, B., Chou, C. P., ... Pentz, M. (2007). Utility of the theory of reasoned action and theory of planned behavior for predicting Chinese adolescent smoking. *Addictive Behaviors, 32*(5), 1066–1081.
Hall, A. J., Bixler, D., Helmkamp, J. C., Kraner, J. C., & Kaplan, J. A. (2009). Fatal all-terrain vehicle crashes: Injury types and alcohol use. *American journal of preventive medicine, 36*(4), 311–316.
Helmkamp, J. C., Furbee, P. M., Coben, J. H., & Tadros, A. (2008). All-terrain vehicle–related hospitalizations in the United States, 2000–2004. *American journal of preventive medicine, 34*(1), 39–45.
Jennissen, C. A., Harland, K. K., Wetjen, K., Peck, J., Hoogerwerf, P., & Denning, G. M. (2014). A school-based study of adolescent all-terrain vehicle exposure, safety behaviors, and crash experience. *The Annals of Family Medicine, 12*(4), 310–316.
Johnson, S. E., & Hall, A. (2005). The prediction of safe lifting behavior: An application of the theory of planned behavior. *Journal of Safety Research, 36*(1), 63–73. https://doi.org/10.1016/J.JSR.2004.12.004.
Lajunen, T., & Räsänen, M. (2004). Can social psychological models be used to promote bicycle helmet use among teenagers? A comparison of the Health Belief Model, Theory of Planned Behavior and the Locus of Control. *Journal of Safety Research, 35*(1), 115–123.
Malekpour, F., Moeini, B., Tapak, L., Sadeghi-Bazargani, H., & Rezapur-Shahkolai, F. (2021). Prediction of Seat Belt Use Behavior among Adolescents Based on the Theory of Planned Behavior. *Journal of Research in Health Sciences, 21*(4), e00536 https://doi.org/10.34172/JRHS.2021.71.
Norman, P., & Conner, M. (2006). The theory of planned behaviour and binge drinking: Assessing the moderating role of past behaviour within the theory of planned behaviour. *British Journal of Health Psychology, 11*(1), 55–70.
O'Boyle, C. A., Henly, S. J., & Larson, E. (2001). Understanding adherence to hand hygiene recommendations: The theory of planned behavior. *American Journal of Infection Control, 29*(6), 352–360.
O'Callaghan, F. V., & Nausbaum, S. (2006). Predicting bicycle helmet wearing intentions and behavior among adolescents. *Journal of Safety Research, 37*(5), 425–431.
Parker, D., Manstead, A. S., Stradling, S. G., Reason, J. T., & Baxter, J. S. (1992). Intention to commit driving violations: An application of the theory of planned behavior. *Journal of Applied Psychology, 77*(1), 94.
Quine, L., Rutter, D. R., & Arnold, L. (1998). Predicting and understanding safety helmet use among schoolboy cyclists: A comparison of the theory of planned behaviour and the health belief model. *Psychology and Health, 13*(2), 251–269.
Quine, L., Rutter, D. R., & Arnold, L. (2001). Persuading school-age cyclists to use safety helmets: Effectiveness of an intervention based on the Theory of Planned Behaviour. *British Journal of Health Psychology, 6*(4), 327–345.
Rhodes, R. E., & Courneya, K. S. (2003). Investigating multiple components of attitude, subjective norm, and perceived control: An examination of the theory of planned behaviour in the exercise domain. *British Journal of Social Psychology, 42*(1), 129–146.
Schifter, D. E., & Ajzen, I. (1985). Intention, perceived control, and weight loss: An application of the theory of planned behavior. *Journal of personality and Social Psychology, 49*(3), 843.
Sheeran, P., & Taylor, S. (1999). Predicting Intentions to Use Condoms: A Meta-Analysis and Comparison of the Theories of Reasoned Action and Planned Behavior1. *Journal of Applied Social Psychology, 29*(8), 1624–1675.
Shults, R. A., West, B. A., Rudd, R. A., & Helmkamp, J. C. (2013). All-terrain vehicle–related nonfatal injuries among young riders in the United States, 2001–2010. *Pediatrics, 132*(2), 282–289.
Şimşekoğlu, Ö., & Lajunen, T. (2008). Social psychology of seat belt use: A comparison of theory of planned behavior and health belief model. *Transportation Research Part F: Traffic Psychology and Behaviour, 11*(3), 181–191.
Sleet, D. A., Diekman, S., Ikeda, R., & Carlson Gielen, A. (2010). Preventing Unintentional Injury: A Review of Behavior Change Theories for Primary Care. *American Journal of Lifestyle Medicine, 4*(1), 25–31. https://doi.org/10.1177/1559827609349573.
Trifiletti, L. B., Gielen, A. C., Sleet, D. A., & Hopkins, K. (2005). Behavioral and social sciences theories and models: Are they used in unintentional injury prevention research? *Health Education Research, 20*(3), 298–307.
U.S. Consumer Product Safety Commission. (2020). 2018 Annual Report of ATV-Related Deaths and Injuries. Washington, DC: U.S. Government Printing Office.
Warner, H. W., & Åberg, L. (2006). Drivers' decision to speed: A study inspired by the theory of planned behavior. *Transportation Research Part F: Traffic Psychology and Behaviour, 9*(6), 427–433.
Williams, A. F., Oesch, S. L., McCartt, A. T., Teoh, E. R., & Sims, L. B. (2014). On-road all-terrain vehicle (ATV) fatalities in the United States. *Journal of safety research, 50*, 117–123.
Xu, Y., Li, Y., & Zhang, F. (2013). Pedestrians' intention to jaywalk: Automatic or planned? A study based on a dual-process model in China. *Accident Analysis & Prevention, 50*(1), 811–819.
Zhou, R., & Horrey, W. J. (2010). Predicting adolescent pedestrians' behavioral intentions to follow the masses in risky crossing situations. *Transportation research part F: traffic psychology and behaviour, 13*(3), 153–163.

Brian Pugliese is currently a research associate at the Virginia Tech Transportation Institute. Brian received his Ph.D. at the University of Idaho in experimental psychology with an emphasis on human factors. His areas of expertise are human error modeling, accident analysis, and unintentional injury risk behaviors.

Benjamin K. Barton is Professor and Chair of the Department of Psychology and Communication at the University of Idaho. He directs the Palouse Injury Research Laboratory. Dr. Barton's areas of expertise are injury etiology and prevention, human factors, transportation safety, and modeling the processes preceding risk behaviors.

Gerardo Lopez received is M.S. in human factors at the University of Idaho. Gerardo was a member of the Palouse Injury Research Lab and currently works as a Human Factors Engineer at Polycom.