Accident Analysis and Prevention 150 (2021) 105920

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Influence of built environment on the severity of vehicle crashes caused by distracted driving: A multi-state comparison

Youngbin Lym, Zhenhua Chen *

City and Regional Planning, The Ohio State University, United States

ARTICLE INFO

Keywords:
Distracted driving
Crash severity safety
Built environment
Multiple states comparison

ABSTRACT

With recent increased attention to the consequences of distracted driving (DD), this research provides a comprehensive investigation of the role of the built environment on the severity of vehicle crashes caused by DD. Utilizing crash data collected from fifteen states in the United States for the period 2013–2017, the association between distracted driving crash severity and various built environment indicators was examined by the generalized ordered logit regression model. The results show that at a lower severity level, DD related crashes were found to be less severe at roundabouts or in urban areas, whereas the probability of injuries rather than property damage only (PDO) increases if an accident involves speeding or when occurring at an intersection or a curved road. Comparatively, at a higher severity level, the odds of severe (or fatal) injury involvement compared to minor injuries and PDO was found to be higher in a work-zone, a curved roadway, or when excessive speed was involved. Conversely, roundabouts and urban areas affected negatively in severe DD crash, which is consistent with the lower-level case. The study also reveals a state-specific variability of the influence of the built environment on the severity of DD related crashes. These findings provide a comprehensive understanding of the severity of DD related crashes for transportation safety planners or policymakers to develop customized policy recommendations, such as designing policies or roadway safety treatments, to curb the negative consequences of distracted driving.

1. Introduction

Distracted driving (DD) has existed since the invention of the cars, along with a driver's inattention (Regan et al., 2011). In conjunction with technological advancements, distracted driving has worsened in recent years due to the wide adoption of multiple in-vehicle electronic devices, such as cellphones, navigation support systems, easily-accessible internet devices, and onboard entertainment systems (Overton et al., 2015; Knapper et al., 2015; Oviedo-Trespalacios et al., 2016; Lipovac et al., 2017), as well as the conventional sources of distracted driving (e.g. eating, drinking, and/or applying makeup) (Yannis et al., 2010; Lansdown, 2012; Lansdown and Stephens, 2013; Papantoniou et al., 2017). Subsequently, distracted driving and drivers' inattention have become major contributing factors in vehicle crashes (Klauer et al., 2006; Olson et al., 2009; Wilson and Stimpson, 2010; Regan et al., 2011) and they are more likely to result in severe crash outcomes (Chen and Lym, 2020); Fatmi and Habib, 2019).

Recently, the danger of distracted driving has gained significant attention among government organizations. The increased awareness of the severity of vehicle crashes associated with distracted driving allowed the Ohio Department of Transportation (ODOT) and the Ohio Department of Public Safety (ODPS) to assemble a Distracted Driving Task Force with more than 30 stakeholders, including insurance companies, private firms, transportation planners, law enforcement agencies, and policymakers (Ohio Department of Transportation, 2019). In addition, the severe consequences of distracted driving have brought nationwide attention to the issue. According to the National Highway Traffic Safety Administration (NHTSA) Traffic Safety Facts (National Highway Traffic Safety Administration, 2019), about 9 % of fatal crashes (2935 crashes out of 34,247 accidents) that occurred in 2017 were reported as distraction-affected (D-A) crashes, with 401 cases (14 % of D-A crashes) caused by cellphone use, resulting in 3166 fatalities (9 % of total fatalities in 2017). Although the total number of fatalities by distracted driving was lower in 2017 than in previous years, the relative proportion of D-A crashes (9 %) remained the same, suggesting that the situation did not improve (National Highway Traffic Safety Administration, 2017, 2018, 2019). The report also reveals that specific age cohorts (i.e., 15–19 and 20–29) are more prone to be engaged in DD fatal crashes, who

* Corresponding author at: City and Regional Planning, The Ohio State University, 275 West Woodruff Avenue, Columbus, OH 43210-1138, United States.
E-mail address: chen.7172@osu.edu (Z. Chen).

https://doi.org/10.1016/j.aap.2020.105920
Received 10 July 2020; Received in revised form 18 November 2020; Accepted 19 November 2020
Available online 11 December 2020

together constituted 15 % of distracted drivers in 2017 (National Highway Traffic Safety Administration, 2019). Therefore, it can be inferred that if an accident involves cellphone use by younger age cohorts, the crash severity is more likely to be severe or fatal rather than resulting in property damage only (PDO) or minor injuries.

Although a significant number of studies have investigated the risk and behavioral response of DD (Brace et al., 2007; Collet et al., 2010; Overton et al., 2015; Oviedo-Trespalacios et al., 2016; Lipovac et al., 2017; Papantoniou et al., 2017), a comprehensive understanding of the linkages between the built environment and DD related crashes has yet to be fully explored. For instance, a previous study by Kidd et al. (2016) revealed that the types of DD behavior vary with changes in the roadway environment and conditions. Fitch et al. (2013) further indicated that while the road traffic environment plays a major role in mobile phone-use related DD, its effects on the severity of crash outcomes have not been properly investigated. This is because earlier works examined the influence of the roadway environments on driving performance under various simulated settings (Rakauskas et al., 2004; Birrell and Young, 2011; Kircher and Ahlstrom, 2012; Knapper et al., 2015; Kountouriotis and Merat, 2016; Oviedo-Trespalacios et al., 2017). Meanwhile, (Chen and Lym, 2020) utilized an actual vehicle crash dataset from 2013 to 2017 in the state of Ohio to i investigate the effects of the built environment on both the frequency and severity of DD-related vehicle crashes. Their study confirmed that the frequency and severity of DD crashes do vary among various built envionments. In a similar vein, Fatmi and Habib (2019) studied the effects of land use and the built environment on injury severity caused by DD in Nova Scotia, Canada. With a few exceptions of recent works, there is still a very limited number of studies that contribute to a thorough understanding of the impact of different built environments on the severity of vehicle accidents caused by DD through accident data (i.e., based on a post-accident perspective). Furthermore, while previous studies have investigated the issues with a focus on a specific state or province, it remains unclear to what extent the severity of DD-related crashes tend to vary among different states, in other words, to what extent the results are generalizable.

Hence, the objective of this research is to provide a more comprehensive assessment, by broadening the scope of the study area to fifteen states in the U.S., to investigate the relationship between built environment and the severity of DD related vehicle crashes. In particular, we developed a micro-level DD crash dataset, including 1.74 million vehicle accidents across 15 states in the U.S. for the period 2013–2017. The built environment can be defined as the physical environment that is constructed by human activity. In general, the built environment consists of the following dimensions: density and intensity, land use mix, street connectivity and scale, aesthetic qualities and regional structure (Handy et al., 2002). Our study focuses on the evaluation of the built environment features, including road configuration, intersection, roundabout, and functional classification of the roads. Specifically, the following three research questions are addressed: How does the severity of DD-related crashes vary in various built environments and roadway conditions? In addition, to what extent the influence of the built environment on the severity of crashes caused by distracted driving vary among different states. If the state-specific characteristics exist, what implications can we draw from the outcomes to improve transportation safety?

Through an empirical assessment based on actual crash datasets collected from 15 states in the U.S., this research helps to improve the understanding of the relationship between the built environment characteristics and the severity of DD-related crashes. More importantly, the study provides implications for public safety and transportation infrastructure planning agencies to identify effective strategies to improve roadway safety, by mitigating the negative consequences (i.e. injuries or fatality) of the built environment on crashes associated with DD.

The rest of this paper is organized as follows. Section 2 reviews the relevant literature followed by data description and methodology in Sections 3 and 4, respectively. Section 5 presents the outcomes of an empirical assessment and discusses the findings in detail, while Section 6 concludes with a summary and future study suggestions.

2. Literature review

2.1. General description of distracted driving

A significant number of DD studies have concentrated on the investigation of the relationship between driving performance and distraction factors, behavioral responses, and crashes. Examples include Oviedo--Trespalacios et al. (2016), who provide a systemic review of the impacts of mobile phone distraction on driving performance based on a wide variety of studies of DD. They proposed a novel human-machine system framework to cope with the complexity as well as to consider various factors (e.g. internal and external interactions related to the mobile phone distracted driving). They found that the systemic approach helps explain the association between drivers' behavioral response (adaptation) and in-vehicle tasks and their interaction to distraction. In a similar vein, several studies (Hancox et al., 2013; Christoph et al., 2019; Oviedo-Trespalacios et al., 2019) investigated the association between driving contexts (e.g. phone function (texting, calling, taking a selfie), roadway driving complexity (driving highways, rural areas, intersection, pedestrians, and passengers) and drivers' behavioral response of mobile phone engagement. When the driving demands are low (drivers are perceived to be less risky), they are more likely to use the mobile phone, whereas an increase of driving complexity prohibits the engagement of mobile phone related tasks, revealing various levels of drivers' self-regulating behavior. In addition, Oviedo-Trespalacios et al. (2020) found that as driving complexity increases (e.g. demanding secondary tasks while behind the wheel or driving sophisticated road traffic environments), drivers are more likely to initiate risk-compensating behaviors to reduce the crash risks. They also found that drivers' decisions to engage in multiple types of risk-compensating behavior are correlated, showing the influence of the unobserved heterogeneity.

Papantoniou et al. (2017) offer a comprehensive review of critical driving-performance parameters for distracted driving based on driving simulators. They argue that due to the multi-dimensional nature of distracted driving, no single driving performance measure can capture all the influences of distraction. One should note that most of the studies related to distracted driving focus on the behavioral responses of drivers in terms of teen behavior (Gershon et al., 2017, 2019), age and execution (Pope et al., 2017), behaviors among young adults (Braitman and Braitman, 2017), driving performance as a response to receiving information via cellphone (Horrey et al., 2017), and the impact of in-vehicle devices on experienced drivers (Knapper et al., 2015). However, there is still a lack of understanding of the relationship between the built environment and crash consequences, such as frequency or severity of DD using actual crash datasets (i.e. from the post-accident perspective).

2.2. General framework for crash severity analysis

Within the crash assessment context, crash severity has been one of the key research areas of transportation safety, along with crash frequency. Several recent studies have provided comprehensive reviews and methodological frameworks for analyzing crash severity (Savolainen et al., 2011; Yasmin and Eluru, 2013; Mannering and Bhat, 2014;

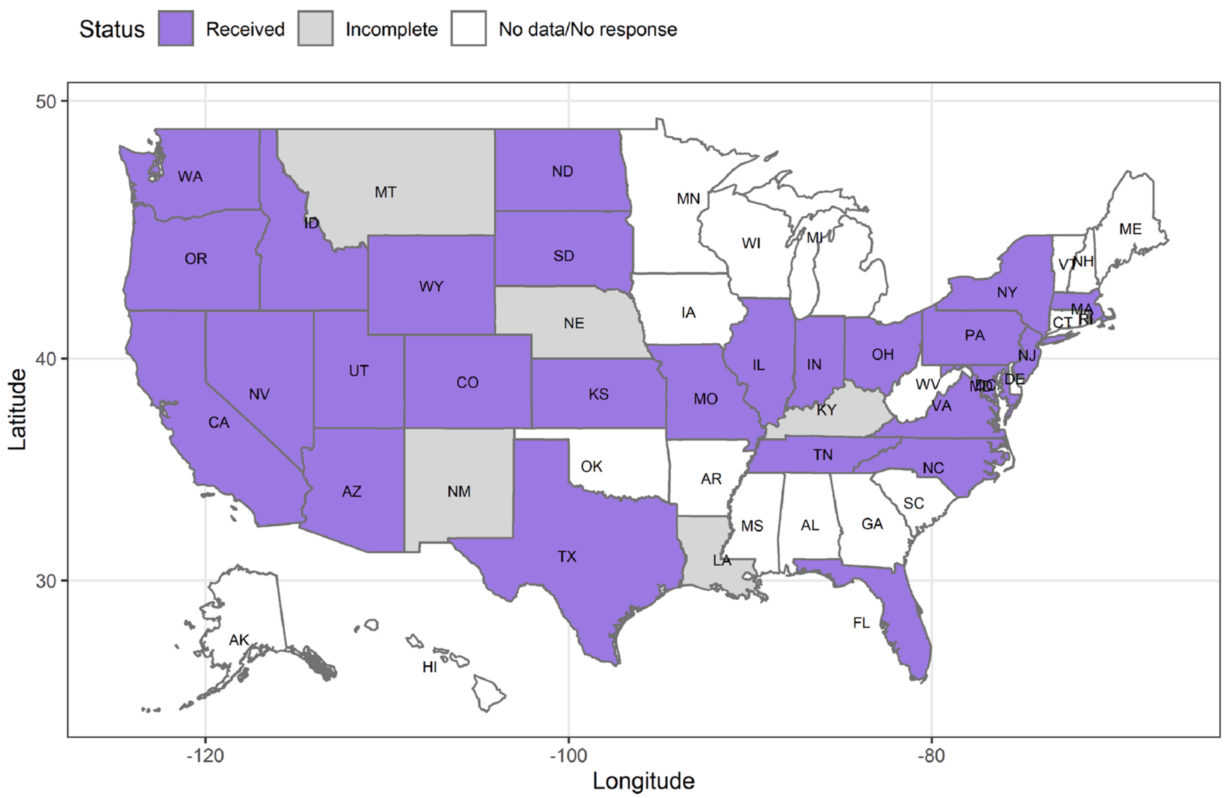

Fig. 1. Availability of DD crash data.
Note: Hawaii (HI) and Alaska (AK) were shifted and not to scale.
Source: Data collected by The Risk Institute, Fisher College of Business, The Ohio State University (As of April. 2020).

Table 1
Identified variables for the severity analysis.

Variable	Description	Variable Coding
DD	Distraction induced vehicle crashes	Yes = 1, No = 0
Severity	Severity levels including PDO, Minor Injury, and Severe Injury	PDO = 1, Injury = 2, Severe Injury = 3
No. Lanes	Number of lanes at the roadway segment	Number of Lanes
Work Zone	Crash location within a work zone	Work Zone = 1, Else = 0
School Zone	Crash location within a school zone	School Zone = 1, Else = 0
Light Condition	The type/level of a light condition at the time of the crash	Daylight = 1, Dawn = 2, Dark = 3
Road Contour	The characteristics of the road such as straight, curved, hill, and grade	Curved = 1, Else = 0
Weather Condition	Weather condition when the crash occurred	Clear = 1, Cloudy (Including Fog, Smog) = 2, Rain = 3, Snow = 4
Divided	Whether the road segment is divided or not	Divided = 1, Else = 0
Roadway System	Hierarchical grouping of roads, streets, and highways based on the type of service	Local (Including collectors) = 1, Minor Arterial = 2, Principal Arterial = 3
Intersection	Crash location involving intersection	Intersection = 1, Else = 0
Roundabout	Crash location at roundabouts	Roundabout = 1, Else = 0
Speeding	Whether the cause of crash is associated with an exceeded speed limit[1]	Speeding = 1, Else = 0
Urban Area	Crash location within urban areas	Urban = 1, Else = 0
Year	Year of crash occurrence	2013–2017
Month	Month to further check for seasonal variability of crashes	
Day of Week	Day of week	
Time of Day	Time of crash within a day	
Location	Spatial coordinate of crash location including latitude and longitude	

Source: Authors' summary.
We were not able to obtain the actual speed limit data of each crash, given that we deal with multi-million crash records across fifteen states in the U.S. and not every state provides this information.

Mannering et al., 2016; Behnood and Mannering, 2017; Xin et al., 2017). Various modeling techniques have been employed to evaluate the influences of contributing factors on the severity of a wide range of traffic accidents, including those involving pedestrians (Abdul Aziz et al., 2013; Mohamed et al., 2013; Yasmin et al., 2014; Xin et al., 2017; Uddin and Ahmed, 2018; Ding et al., 2018), bicycles (Zahabi et al., 2011; Kaplan et al., 2014; Chen and Shen, 2016, 2019) and motor vehicles (Dabbour et al., 2017, 2019; (Chen and Lym, 2020); Fatmi and Habib, 2019; Wang and Kim, 2019; Tang et al., 2019; Zeng et al., 2019; Lym and Chen, 2020). Crash severity, the dependent variable in severity modeling frameworks, is typically characterized by ordered categories of outcomes: fatality, incapacitating injury (severe injury), non-incapacitating injury (evident injury), possible injury, and property damage only (PDO) (Savolainen et al., 2011; Xin et al., 2017; (Chen and Lym, 2020)). Some models consider a binary response outcome (e.g. injury vs. non-injury or fatal vs. non-fatal crashes) while other models account for either the ordinal nature (e.g. ordered probability models such as ordered logit, probit, and their random parameter extensions) or nominal structure (e.g. multinomial logit, nested logit, and mixed logit models) (Savolainen et al., 2011).

In terms of contributing factors (independent variables) on pedestrian crash severity, Xin et al. (2017) summarized five categories of variables adopted from previous studies. They classified the influential factors as 1) neighborhood characteristics and the built environment; 2) driver characteristics and driving behaviors; 3) pedestrian features and behavior; 4) vehicle types and traffic characteristics; and 5) roadway characteristics and environmental circumstances. Meanwhile, Xin et al. (2017) specifically focused on neighborhood features and the built environment to investigate the effects on pedestrian injury severity by means of a random parameter generalized ordered probability model. As discussed in Mannering et al. (2016), the advantage of the random parameters approach is that it relaxes the basic assumptions of standard probability models, allowing unobserved heterogeneity into the modeling framework (discussion on the limitation of the standard models can be found elsewhere such as Washington et al., 2011; Eluru and Yasmin, 2015). The adoption of a random parameter ordered model is also found in Dabbour et al. (2017), which investigated significant factors affecting the severity of injuries in vehicle-train collisions and in Dabbour et al. (2019), which explored the temporal stability of these factors.

Chen and Shen (2016(Chen and Shen, 2016); 2019(Chen and Shen, 2019)) utilized the influence of the built environment to analyze bicycle injury severity. They consider five sets of contributing factors, including crash characteristics, socio-demographics, traffic controls, roadway network circumstances, and land use to examine their impacts on the severity of bicycle crashes. Using the generalized ordered logit (GOL) modeling framework, their studies address the ordered nature of crash severity and successfully capture the effects of various roadway environments with respect to changes in the thresholds. As a further extension of the GOL method, Zeng et al. (2019) employed a Bayesian framework to account for the spatial influence of crash severity by means of a conditional autoregressive prior specification.

In addition, unlike the mainly regression-based approaches discussed above, recent advances in methodologies provide different insights into the influence of contributing factors on the crash severity research domain. For instance, the adoption of machine learning techniques, such as tree-based algorithms (Iranitalab and Khattak, 2017; Ding et al., 2018; Wang and Kim, 2019; Tang et al., 2019), support vector machine (Iranitalab and Khattak, 2017; Tang et al., 2019), and clustering methods (Mohamed et al., 2013; Iranitalab and Khattak, 2017), which relax the linearity assumption between the covariates and the response, addresses non-linear effects of independent variables so as to concentrate on the predictive performance and factor identification (Ding et al., 2018).

2.3. Gaps in knowledge: limited focus on the role of the built environment within distracted driving

Previous studies have investigated the effects of the built and/or road environment and drivers' behavioral responses under distraction in a simulated setting. For instance, Kountouriotis and Merat (2016) evaluated driving performance between non-distracted and distracted drivers in different road environments (straight and curved sections). They showed that road environment does affect driving behaviors as driving difficulty increases: in those circumstances, drivers might be forced to restrict their attention to secondary tasks in order to focus on the primary task, driving. Similarly, Oviedo-Trespalacios et al. (2017) investigated the speed adaptation behaviors of distracted drivers under varying road infrastructure and traffic complexity conditions. Their study revealed that distracted drivers behave differently from

Table 2
The comparison of variable availability across multiple states.

State ID[1)]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Variable	CA[5)]	FL	ID	IN	MA[3)]	NC	ND[4)]	NV	NY	OH	OR	PA	TX	VA	WA
DD	Y[2)]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Severity[6)]	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Lanes	N[2)]	Y	Y	N	Y	Y	N	N	N	Y	Y	Y	Y	N	N
WorkZone	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	N
SchoolZone	N	N	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
Light	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
Contour	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
Weather	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Divided	N	Y	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	N
Roadway	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y
Intersection	Y	Y	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y
Roundabout	N	Y	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y	N	Y
Speeding	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y	Y	N
Urban	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N
Day	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Location	Y	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	Y	N	N

Notes: 1) Initially, we have identified seventeen states and then two states (NJ and SD) were excluded at the statistical analysis process.
2)"Y" represents the variable availability, whereas "N" refers to the variable "not available".
3) Data for MA and ND are collected between 2013 and 2016, while TX provides datasets for the period 2013-2018.
4) Roadway system in ND identified using Ohio Codebook manual (i.e. based on the functional classification).
5) In the construction of the work zone variable in CA, we used "D – Construction or Repair Zone" class of Roadcondition1 & 2 columns under the *collision* file.
6) Severity has five different classes having an ordered nature. It consists of Property Damage Only (PDO), Possible Injury (PI), Evident Injury (EI), Severe Injury (SI), and Fatality (FA).

non-distracted drivers, particularly when driving conditions are complicated (e.g. a curved road or car-following situations). Haque et al. (2016) examined the decisions and actions of distracted drivers at the onset of yellow lights using data collected from the National Advanced Driving Simulator located at the University of Iowa, in which they found that the propensity of yellow light running is significantly higher among older (aged 50–60) distracted drivers. While these pioneering studies provide a solid understanding of the driving behavior in various road environment based on simulation analysis, there is, however, limited understanding of to what extent the crash consequence as a result of DD varies among different built environment on the actual crash outcomes.

Meanwhile, (Chen and Lym, 2020) employed a post-accident perspective utilizing actual crash data in a single state, Ohio, to analyze the role of the built environment on the frequency and severity of DD-induced crashes. Their study suggests that roundabouts are an effective measure in reducing the severity of crashes associated with DD, while other road environments tend to have heterogeneous effects on severity among different types of crashes. To complement their findings from a different perspective, Lym and Chen (2020) examined the influence of space on the severity and frequency of DD crashes simultaneously utilizing aggregated crash data in central Ohio. They verified the existence of unobserved heterogeneity stemming from spatial correlation as well as within-severity level correlation in DD crashes, even after accounting for the effects of a few built environment characteristics. In addition, Fatmi and Habib (2019) assess the effects of land use and the built environment on injury severity involving DD in Nova Scotia, Canada. They used police-reported collision data and attempted to account for the latent heterogeneity by means of the latent segmentation-based ordered logit model. Their findings suggest that rain, curved roads, freeways, and mid-block collisions can all aggravate vehicle injury severity, while higher land use mix and population density reduce the severity level. These studies suggest that the built environment may have a heterogeneous effect on crash outcomes due to the variations of driving behaviors and location-specific attributes in different localities.

As revealed by Oviedo-Trespalacios et al. (2020), it is worth noting that some built environments may increase the driving complexity in a certain built environment. As a result, drivers may not engage in DD as a form of risk-compensating behavior. These variations of behavior responses further suggest the need for a comprehensive evaluation of the relationship between the built environment and crash consequences caused by DD. Our objective is to examine this issue through a multi-state comparison to understand to what extent the relationship varies among different states. One should note that such a multi-state comparison is particularly important in countries, such as the U.S. where driving behavior could be heterogeneous due to the influences of different natural and built environments, and socioeconomic factors. To the best of our knowledge, no attempt has been made to compare the role of the built environment on the crash severity by DD across multiple states in the U.S. at the micro-level. We believe that by comparing the results through a consistent modeling framework and data structure, our assessment helps to understand the inherent heterogeneity of DD related crash outcomes and the built environment in different states.

3. Data

3.1. Data generation

The multi-state crash data was developed systematically through the following procedures. In the first step, key variables, such as DD-related vehicle crash severity, built environment, roadway design, land use, and environmental conditions at crash locations were selected for the assessment. These variables were selected following similar works (e.g. (Chen and Lym, 2020); Fatmi and Habib, 2019). In addition, we also adjusted the list based on the feedback from the field experts of the Ohio Department of Transportation and academic scholars. In the second step, we contacted the transportation agencies (whose vehicle accidents are recorded by law enforcement officers) of the 50 states and the District of Columbia in the U.S. to obtain the crash datasets for the period 2013–2017. Fig. 1 presents the outcome of the outreach efforts. In total, 1.74 millions of DD related crashes were collected. Among all the agencies, 26 states responded to our requests and shared the completed dataset. Five states (colored in grey) responded with incomplete datasets. The remaining states neither responded to our request nor indicated

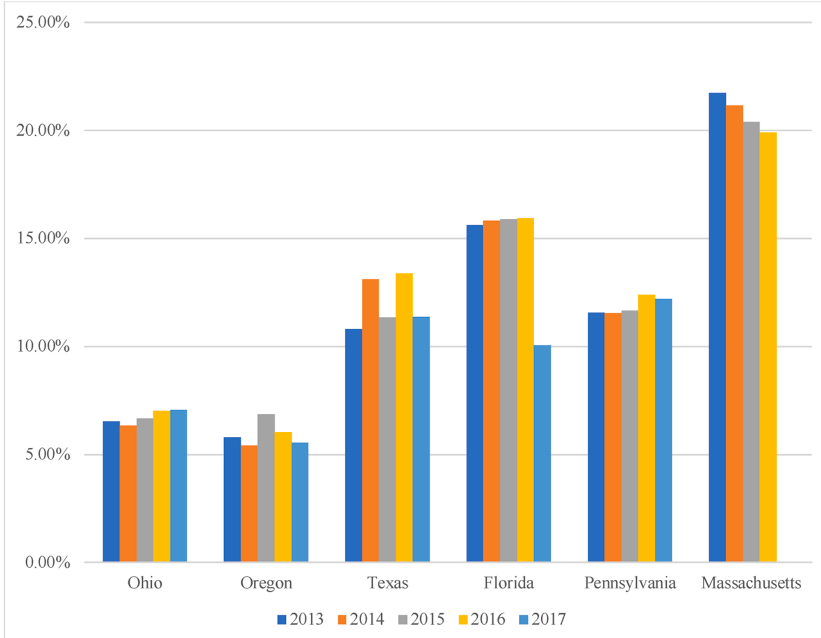

Fig. 2. The annual relative share of DD crashes in response to total crashes (2013–2017).
Note: Data in Massachusetts (MA) are only available during 2013–2016.

data availability.

In step three, a comprehensive data validation was conducted. One of the most challenging tasks was the identification of DD related crashes as the definition of DD crashes in the crash report system varies across different states. We assume that when the cause of the crash or contributing factors (or other associated factors depending on the crash manuals in each state) include cellphone, texting, distraction, or inattention, we regard it as a DD crash, separating from other types of crashes. (For details, see Appendix Table A1). Next, the key candidate variables were selected using Ohio as a benchmark (as shown in (Chen and Lym, 2020). Table 1 summarizes the key variables of interest identified from the crash datasets (although some of them are missing in several states, as shown in Table 2). In the end, only the ones with higher relevance to the research theme were selected. Specifically, variables selected to measure the roadway environment include: number of lanes, road contour, divided or not, roadway system, intersection, and roundabout); and variables represent land-use characteristics include: work zone, school zone, and urban area.

To minimize variability of the explanatory variables and also to achieve a consistent comparison across the states (as shown Appendix Table A2), all the variables were reclassified following a consistent guideline. The detailed procedure for the variable construction (e.g. reclassification of several independent variables, how to identify DD crashes, how to treat redundancies) is explained in Appendix Tables A1–A3. In the end, Table 2 summarizes the availability of variables identified across the selected states.

The states for the comparison were selected given the consideration of data availability and quality. In addition, given that our previous study has confirmed that both roundabouts and work zones have a significant effect on the severity of DD-related crashes in Ohio ((Chen and Lym, 2020)), priority in the selection of states was given to the ones with both variables available (e.g. Florida, Pennsylvania, Massachusetts, North Dakota, Texas, Ohio, and Oregon). As a result, fifteen states were selected to provide a comprehensive understanding of the association between crash severity by DD and the built environment features.

While our dataset provides a comprehensive reflection of crash outcomes of DD in various states, one should note that there are two limitations of the dataset. First, since all the information on DD crashes was originally collected by police officers, the issue of underreporting is inevitable. Hence, one should note that the research findings were developed only based upon the sample datasets, which may not fully reflect the real-world scenario. Second, given that different states have different legislation requirements and reporting systems, even though we have tried our best to capture the heterogeneous effects of the results through a consistent procedure of variable creation, it is worth noting that the data still has a limitation in terms of providing generalizable implications to the regions and states beyond the analysis.

3.2. Descriptive analysis

Fig. 2 presents the variations in vehicle crashes caused by DD of six selected sample states for the period 2013–2017 where both roundabout and work zone indicators are available in their datasets. In general, the five-year average share of DD crashes is the highest in Massachusetts, where it exceeded 20 % of the total crashes. Florida ranks the second with a share of approximately 15 %. Conversely, the share of DD crashes is only about 7 % in Ohio and Oregon. One should also note that the relative proportions of DD crash in Florida and Massachusetts decreased from 2013 to 2017, whereas it has increased slightly in Ohio and roughly stabilized in Pennsylvania.

In Table 3, the frequency of DD crashes is also compared in terms of a ratio attributed to different baselines, such as the number of total

Table 3
Five-year average DD crashes with respect to total crashes and residents.

	OH	OR	TX	FL[1)]	PA	MA[2)]
Five-year average						
DD crashes (A)	19,707	1241	69,990	53,431	14,989	27,921
Total crashes (B)	292,599	20,942	586,171	366,725	126,173	134,511
Licensed drivers (C)	7,971,285	2,826,741	15,991,020	14,316,604	8,943,374	4,903,124
Driving age (16+) (D)	9,293,367	3,268,097	20,442,981	16,635,696	10,422,150	5,541,718
Total residents (E)	11,610,275	4,033,104	27,408,291	20,262,854	12,790,655	6,761,108
Proportion (%)						
DD/Crash (A/B)	6.74 %	5.92 %	11.94 %	14.57 %	11.88 %	20.76 %
DD/Resident (A/E)	0.17 %	0.03 %	0.26 %	0.26 %	0.12 %	0.41 %
Crash/Resident(B/E)	2.52 %	0.52 %	2.14 %	1.81 %	0.99 %	1.99 %
Five-year total						
DD	98,536	6,204	349,952	267,156	74,944	111,683
Total crashes	1,462,994	104,710	2,930,854	1,833,625	630,865	538,042

Source: 1) For DD and Total Crashes: Authors' Database collected from each state agency.
2) Licensed Drivers and Others: Highway Statistics 2017 (Federal Highway Administration and U.S. Department of Transportation, 2019a).
Note: 1) Total Crashes in Florida are extracted from Crash Facts Annual Report for the period 2013–2017 (Florida Highway Safety and Motor Vehicles (FHSMV), 2019).
2) DD Crash data are only available during the periods 2013–2016 in MA.

Table 4
Injury vs. PDO.

	CA	FL	ID	IN	MA	NC	ND	NV	NY	OH	OR	PA	TX	VA	WA
No.Lanes		0.06	-			-				-					
WorkZone	-	0.01		-	-0.27		-	0.1	-0.15	-	-	0.18	-0.01	-	
SchoolZone				-	0.65	-	-		-	-0.29		0.5	-	-	
Light_Dawn	-	0.07	-	-	-	-	-	0.18	0.19	-0.11	-0.36	-	0.07		-
Light_Dark	-0.1	-0.01	-	0.04	0.15	0.18	-	0.08	0.1	-0.08	-0.41	-0.25	-0.05		-
Road_Curved		0.18	-	-		0.49	-		-0.29	0.05	-	-	0.2	0.05	-
Weather_Cloudy	0.07	0.09	-	-	0		-	0.03	-0.28	-	0.26	-	0.07	-	-0.1
Weather_Rain	-	-0.1	-	-	-0.37		-	-	-	-0.08	-	-0.08	-0.11	-0.12	-0.1
Weather_Snow	-	0.56	-	-	-		-	-0.26	-0.22	-0.27	-	-	-0.25	-0.31	-0.4
Divided		0.18	0.34	0.31	0.07		-			-0.05	-	-0.24	0.45	-	
Minor_Road	-0.23	-0.38	0.45	0.59	0.26	0.18	-			0.17	-	-	-	0.22	-
Principal_Road	-0.36	-0.39	0.75	0.47	0.23	0.21	0.78			0.17	-0.25	0.24	-0.11	0.18	-0.2
Intersection	0.46	0.41		0.13	0.25	0.06	0.21			0.07	-	0.26	0.22	0.43	-
Roundabout		-0.25		-	-0.17	-0.45	-			-0.46	-	-	-0.29		-0.6
Speeding	0.12	1.34	2.77		1.62			0.34	0.51	0.76	3.02	0.69	0.62	0.26	
Urban_Area	0.23	-0.35	-0.24	-0.1	-0.3	-0.12	-0.46		-0.44	-0.29	0.72	0.05	0.06	-0.15	-0.1

<< -0.75	-0.5 ~ 0.75	-0.25 ~ 0.5	>> - 0.25	0	<< 0.25	0.25 ~ 0.5	0.5 ~ 0.75	>> +0.75	-

Notes: 1) Each value represents differences from baseline 1 in the odds ratio. For example, the Speeding variable in CA (Appendix Table A5) has an estimated odds ratio 1.12 and is statistically significant. As compared to baseline 1, the difference is 0.12 (1.12 – 1 = 0.12) and this is interpreted as speeding in CA is associated with 0.12 (12 %) increased relative risk (Injury rather than PDO).
2) "- "denotes Non-Significant and Blank refers to variables Not-Available.
3) Reference categories: WorkZone (No), SchoolZone (No), Light (Daylight), Road_Curved (No), Weather (Clear), Divided (No), Minor_Road (Local_Road), Principal_Road (Local_Road), Intersection (No), Roundabout (No), Speeding (No).

crashes, licensed drivers, and the total population. The rate of total crashes with reference to the total population is highest (2.52 %) in Ohio, followed by Texas (2.14 %) and Florida (1.81 %). For DD-related crashes per total residents, statistics show that the five-year annual average ratio of Ohio is 0.17 %, which is higher than that of Oregon and Pennsylvania, but lower than the rate in Florida, Texas, and Massachusetts. The state with the highest rate of DD crashes relative to total crashes, as well as total residents, is Massachusetts. Moreover, although Ohio and Pennsylvania have similar population sizes, the relative proportion of DD to total crashes in Pennsylvania is twice that of Ohio. These findings clearly reveal that states differ in the relative risk of DD crashes, and suggest that Massachusetts and Pennsylvania especially need to take action to reduce the frequency of DD crashes. To validate these findings, statistics were also calculated based on the most recent data in 2017 (Appendix Table A4), which confirms that the findings are consistent.

Table 5
Severe Injury vs. Injury & PDO.

	CA	FL	ID	IN	MA	NC	ND	NV	NY	OH	OR	PA	TX	VA	WA
No.Lanes		0.02	-			0.1				-0.04					
WorkZone	-	0.18		-	-		1.33	0.16	-	-	-	0.93	-0.1	-	
SchoolZone				-0.3	-	-	-		-	-		-	-	-	
Light_Dawn	-	0.2	-	-	-	-	-	0.21	0.38	-	-0.53	0.76	0.58		-
Light_Dark	0.76	0.35	-	0.11	0.41	0.51	-	0.34	0.46	0.09	-0.54	0.17	0.4		0.43
Road_Curved		0.66	-	-		3.58	-		0.32	0.14	-	0.54	0.67	0.53	0.37
Weather_Cloudy	-	0.15	-	-	-		-	0.11	-0.15	-	-	0.49	-0.02	-	-
Weather_Rain	-	-0.16	-	-0.2	-0.52		-	-0.41	-0.1	-0.24	-	-	-0.31	-0.16	-0.3
Weather_Snow	-	13.09	-	-	-0.21		-	-	-0.24	-0.48	-	-	-0.47	-	-
Divided		0.08	0.48	0.22	-		-			0.16	-	0.65	2.06	-0.06	
Minor_Road	-	-0.97	-	0.9	-	0.23	-			-	-	0.69	0.3	0.26	-
Principal_Road	-	-0.97	0.57	0.66	-	-0.34	-			-	-0.6	0.23	-0.17	0.39	-0.3
Intersection	-	0.26		-0.1	0.19	-0.22	0.6			-0.11	-	-0.12	-0.17	0.54	-0.3
Roundabout		-0.39		-	-0.5	-	-			-	-	-	-		-
Speeding	-0.62	9.83	1.41		2.23			211.3	0.86	1.19	2.12	13.93	2.58	1	
Urban_Area	-0.42	-0.76	-0.63	-0.5	-0.41	-0.77	-0.94		-0.29	-0.54	-	-0.46	-0.48	-0.49	-0.5

<< -0.75	-0.5 ~ 0.75	-0.25 ~ 0.5	>> - 0.25	0	<< 0.25	0.25 ~ 0.5	0.5 ~ 0.75	>> +0.75	-

Notes: 1) Each value represents differences from the baseline 1 in the odds ratio. Thus, the negative value of Roundabout in FL is given by 0.61–1 = –0.39, indicating a reduction of the relative risks (i.e. Injury & PDO rather than Severe Injury involvement).
2) "- "denotes Non-Significant and Blank refers to variables Not-Available.
3) Reference categories: WorkZone (No), SchoolZone (No), Light (Daylight), Road_Curved (No), Weather (Clear), Divided (No), Minor_Road (Local_Road), Principal_Road (Local_Road), Intersection (No), Roundabout (No), Speeding (No).

4. Methodology

This study employs the Generalized Ordered Logit (GOL) regression model to examine the association between DD crash severity and the built environment, following Chen and Shen (2016); (2019) and Williams (2016). The GOL model was selected due to the following considerations. Given that our objective is to examine the relationship between the built environment and DD related crash severity among 15 states, we adopt the GOL model as the standard model to account for the variable specific influence on each severity level within states, while also providing consistent outcomes for the comparison.

Our model is able to address the heterogeneity in the following aspects. First, the state-specific heterogeneity will be understood through the comparison of the estimates based on the samples of fifteen states. Second, the heterogeneous effects of independent variables (built environment) on different crash severities will be revealed by comparing the results of the GOL model.[1]

To address the key research questions, we hypothesized that each explanatory variable would act differently with respect to multiple severity levels. As Williams (2016) noted, when the assumptions of the conventional ordered logit model are met, all of the corresponding coefficients (except the intercepts as each provides differences among ordinal outcomes) should be the same across different categories of the response variable. This leads to the parallel lines that each class can be only distinguished by the intercepts (i.e. the odds ratio will stay the same (equal logit slope) so that the model is also called the proportional odds model). Conversely, the GOL model estimates ordered response outcomes based on the non-parallel or partial-parallel proportional odds assumptions that differentiate the effects of independent variables with respect to each class in the dependent variable (Williams, 2016). Such flexibility allows us to obtain different sets of estimated coefficients corresponding to the ordered response classes, leading to unveiling heterogeneous influences of each explanatory variable in response to the different severity outcomes.

The dependent (response) variable is a categorical variable with an ordered nature, which we reclassify the severity of vehicle crashes caused by DD into three categories: Property Damage Only (PDO), Minor Injury (Injury), and Severe Injury (Severe Injury)[2] . The GOL model assumes that there is a continuum of potential severity levels, but what we observe is data that is censored into these particular severity levels (i.e. an ordered ranking has a cumulative distribution) (Agresti, 2010; Williams, 2016). The model was applied to estimate the coefficients separately for each state, which provides outcomes for a multi-state comparison (i.e. allowing us to understand the state-specific heterogeneous influence of the various built environments on DD crash severity outcomes through the estimated coefficients). For simplicity, the implicit state identifier is omitted here. If there are K categories of crash severity in a random variable Y, for the i-th observation, Y_i, the GOL model can be written as:

$$\log\left(\frac{\Pr(Y_i > j)}{1-\Pr(Y_i > j)}\right) = \log\left(\frac{g(\boldsymbol{\beta}_j \boldsymbol{X}_i)}{1-\ g(\boldsymbol{\beta}_j \boldsymbol{X}_i)}\right) = \alpha_j + \boldsymbol{\beta}_j \boldsymbol{X}_i \tag{1}$$

$$Pr(Y_i > j) = g(\boldsymbol{\beta}_j \boldsymbol{X}_i) = \frac{\exp(\alpha_j + \boldsymbol{\beta}_j \boldsymbol{X}_i)}{1+\exp(\alpha_j + \boldsymbol{\beta}_j \boldsymbol{X}_i)} \tag{2}$$

where $j = 1, 2, 3, 4, \ldots, K$-1, α_j denotes the cutoff points with respect to the label j and $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \cdots, \beta_{jp})^T$ represents a set of vectors of

[1] We acknowledge that one limitation of this analysis is that our model does not address the unobserved heterogeneity within each state, which might be caused by the random effects. This issue definitely deserves a further exploration in future research.

[2] As of Table 2, there are five classes (i.e. Property Damage Only (PDO), Possible Injury (PI), Evident Injury (EI), Severe Injury (SI), and Fatality (FA)) in DD crash severity outcomes. We re-categorize them as PDO, Injury (PI + EI), and Severe Injury (SI + FA) to overcome too few observations in the FA severity category so as to improve reliability of estimation.

corresponding coefficients (p-dimension) to be estimated for each j. $\boldsymbol{X}_i = (x_{i1}, x_{i2}, \cdots, x_{ip})^T$ is a vector of p independent variables, which represent various built and natural environmental characteristics of i-th observation (assumed to be observed without errors) and $g(\cdot)$ refers to a logit link function. Eq. (2) allows us to find that when $\boldsymbol{\beta}_j$ does not vary with respect to each severity level j (i.e. $\boldsymbol{\beta}_j = \boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$), then the GOL model is equivalent to the ordered logit model (i.e. equal slope or parallel line assumption holds).

The probability functions of each Y_i being in category j derived from Eq. (2) can be denoted as the following (Agresti, 2010):

$$\Pr(Y_i = 1) = 1 - g(\boldsymbol{\beta}_1 \boldsymbol{X}_i) \tag{3}$$

$$\Pr(Y_i = j) = g(\boldsymbol{\beta}_{j-1} \boldsymbol{X}_i) - g(\boldsymbol{\beta}_j \boldsymbol{X}_i),\ j = 2,\ 3,\ \cdots,\ K - 1 \tag{4}$$

$$\Pr(Y_i = K) = g(\boldsymbol{\beta}_{K-1} \boldsymbol{X}_i) \tag{5}$$

where j is the label of a crash severity level in the dependent variable, Y_i and K is the highest ordered response, which in our case equals 3. Therefore, two thresholds are derived, which include the cutoff points (α_1 and α_2) between $j = 1$ vs. $j = 2$ and $j = 1$ & 2 vs. $j = 3$ (showing a cumulative structure), corresponding to PDO vs. Injury and PDO & Injury vs. Severe Injury, respectively. The probabilities of the first category (PDO) and the last category (Severe Injury) are represented by Eqs. (3) and (5), respectively.

5. Results and discussions

Initially, we conducted the GOL regression for seventeen states using the data based on Tables 1 and 2 as well as Appendix Tables A1–A3. The ordinal response (dependent variable) has five levels of severity in sixteen states: Property Damage Only (PDO), Possible Injury (PI), Evident Injury (EI), Severe Injury (SI), and Fatality (FA), while in New Jersey the crash severity was coded: PDO, Injury, and Fatality. With five classes, some states did not produce the estimated coefficients for corresponding severity levels due to very few observations in the Fatality category, resulting in unstable as well as unreliable estimation.

As an alternative to the initial five levels, we aggregated them into three: PDO, Injury (PI + EI), and Severe Injury (SI + FA) in order to circumvent the limitation of small samples at the highest severity rank (FA). Subsequently, we performed the GOL regression analysis for seventeen states with the revised datasets and the results became more stable except those in South Dakota. The GOL model for South Dakota neither converged nor provided estimated coefficients, ultimately leading to its exclusion in the final model. Meanwhile, we also identified that a set of estimated parameters in New Jersey could not be directly comparable to that of other fifteen states because of different severity classifications (i.e. in New Jersey, the reported crash severity categories are PDO, Injury, and Fatality, whereas in other states, the severity levels are PDO, Injury, and Severe Injury). Therefore, we exploit the estimated results of the fifteen sampled states to verify disparities of the influence of the built environment on the severity of DD-related crashes via a multi-state comparison approach.[3]

To achieve an informative interpretation, key results are presented in Tables 4 and 5 in lieu of Appendix Table A5 (which shows detailed outcomes by the estimated odds ratio from the GOL regression analysis). To appreciate the results in following Tables 4 and 5, it is useful to revisit that a coefficient β_p associated with the single independent variable x_p in the linear predictor (the right-hand side of Eq. (1)) can be interpreted as the difference in the logit (log-odds) when x_p is increased by one unit from x_{p0} to $x_{p0} + 1$. That is, if we are considering the odds O_{jk} of outcome j over outcome k in the response variable, then we can easily show that $\beta_p = log(O_{jk}x_{p0} + 1) - log(O_{jk}x_{p0})$ and taking the antilog (i.e. exponential) on both sides gives the proportional change in the odds in response to a unit change of x_p (as presented in Appendix Table A5). With all that in mind, here we discuss the results in terms of the proportional changes in odds, so that when $Exp(\beta_p) > 1$, the marginal influence of x_p on the response (dependent) variable is positive (i.e. the likelihood of the occurrence of severe outcomes increases in reference to the baseline category or is more likely to have elevated risks).

Table 4 delineates the results for the lower levels of severity threshold, i.e. Injury vs. PDO, with respect to independent variables across fifteen states. The red color indicates that there is a positive statistical association between the corresponding factor and the severity of crashes, whereas the green color reflects a negative statistical association. In the lower severity level, when a DD crash is involved in speeding (i.e. whether the DD crashes is related to exceeding the posted speed limit), it is prone to be Injury rather than PDO across those sampled states, which confirms that speeding is a key factor that influences the level of crash severity.

The work zone variable is found to be an influential factor in Florida, Pennsylvania, Nevada, Massachusetts, New York, and Texas in a different way. The results indicate that a DD crash is more likely to result in Injury rather than PDO, suggesting an elevated relative risk at work zone area in FL, NV, and PA by 1 %, 10 % and 18 %, respectively. Conversely, the result shows that the probability of DD related crashes involving Injury as compared to PDO is likely to be lower by 27 %, 15 %, and 1 % a work zone in Massachusetts, New York, and Texas, respectively. Similarly, odds of an accident that results in injuries in a school zone is likely to be 29 % lower in Ohio, whereas a relative risk of injury involvement by a DD crash appears to be higher in Pennsylvania and Massachusetts. These results confirm that the severity of DD crashes does vary substantially among different states. Such heterogeneous effects could either be due to the different driving behaviors among different states. It may also be due to the different safety regulations at the work zone and school zone among different states.

In terms of the influence of the roadway system[4], our analysis also reveals a large variation exists among states. For instance, the probability of involving injuries as opposed to PDO is likely to be higher in Idaho, Indiana, Massachusetts, North Carolina, North Dakota, Ohio, Pennsylvania and Virginia, if a crash occurs on a principal arterial (Principal_Road). Conversely, opposite effects were found in states such as California, Florida, Oregon, Texas, and Washington, suggesting that the odds ratio of a DD related crash involving injuries instead of PDO is likely to be relatively lower on principal arterials as compared with local roads. Again, the results reveal that the severity of DD related crashes tend to vary substantially among states, even in a similar road environment condition. Such a difference may reflect the outcomes of complex heterogeneous effects, such as driving behaviors, built environment conditions and enforcement efforts, among various states.

The estimated odds ratio associated with an intersection is greater than 1 for most of the states. Thus, the probability of a DD vehicle crash at an intersection shifting from PDO to Injury increases by between 7 % and 46 % in our sampled states. Conversely, the influence of a roundabout is found to have an opposite association with the DD crash severity. Specifically, the probability of an injury of DD crash at a roundabout was found to be lower by 17 %–60 % in Florida, Massachusetts, North Carolina, Ohio, Texas, and Washington, which confirms

[3] One should note that our assessment did not consider the drivers' exposure to the completing activity under the various driving conditions (e.g. built environments, weather conditions) due to data limitation. However, as Christoph et al. (2019) and Oviedo-Trespalacios et al. (2020) suggested, drivers may not engage in DD when they perceive too complex or demanding environment. Hence, such risk compensating or self-regulating behaviors, in fact, could have potential impacts on the regression model outcomes.

[4] Here, we set "local roads" as a reference category for Minor_Road and Principal_Road variables.

its positive influence on risk reduction. Such a result suggests that roundabouts could be an effective countermeasure to mitigate the severity of DD crashes, given that the introduction of curvature at intersections may force drivers to slow down, which hence, reduces the risk of crash severity. The finding is consistent with (Chen and Lym, 2020). In terms of the urban versus rural environment, the estimated odds ratio reveals that in states, such as California, Oregon, Pennsylvania, and Texas, DD crash is more likely to result in an injury instead of PDO in an urban area, suggesting that the risk of DD tends to be relatively higher in an urbanized area in these states.

In addition, Table 5 provides a summary of the impacts of the variables on higher levels of crash severity. Severe Injury vs. Injury & PDO is calculated in a cumulative manner. Similar to the lower case of severity, the speeding variable was also found to be an important factor that affects the severity of crashes. In general, speeding is associated with an increase in the odds of severe crash outcomes. In particular, we observed an extreme high odds ratio in the case in Nevada, which is likely due to a very few observations in this category in this state (i.e. Severe Injury and Speeding; for details, see Appendix Table A6). Conversely, speeding was found to be associated with lower severity risks in the state of California. Such a result could be due to the relatively small samples being captured in the dataset.[5]

In terms of the work zone variable, while the odds ratio of involving a more severe injury is much higher in several states, such as Florida, North Dakota, Nevada and Pennsylvania, the odds of a DD accident involving a severe injury was found to be around 10 % lower in Texas. Such an outcome could be due to the successful implementation of the Move Over/Slow Down law[6] in Texas. The strategy was developed dedicatedly to improve the safety of work zones (Texas Department of Transportation, 2019).

The assessment also reveals that when a crash occurs along a curved roadway segment, the probability of severe injuries rather than PDO or injuries increases. This is in line with our intuition that as driving difficulty increases, so does the relative risk of crash severity. When it comes to roadway systems, overall trends are similar to what we observed at the lower severity level threshold in Table 4, depicting state-specific variability that produces mixed outcomes.

In addition, the impact of intersections on higher levels of severity varies by state. Unlike the lower severity levels, where intersections are consistently associated with an increase in the relative risk of injuries, we see that a DD crash occurrence at an intersection is likely to have a 10–22 % lower severe injury risk (i.e. PDO or Injury over Severe Injury) in Indiana, North Carolina, Ohio, Pennsylvania, Texas and Washington. Conversely, states such as Florida, Massachusetts, North Dakota and Virginia, appear to have higher relative risks around an intersection. The findings in Ohio and Pennsylvania[7] could be partially explained by the fact that both states have successfully implemented Intersection Safety Implementation Plans (ISIP) in 2010 and 2014, respectively. For instance, since the implementation of ISIP, Ohio achieved a 23 % reduction in fatalities and a 14 % reduction in severe injuries at all intersections during 2003–2013 (Federal Highway Administration and U. S. Department of Transportation, 2019b). The ISIP seems to be an effective countermeasure to all severe vehicle crashes, including ones associated with distracted driving.

Meanwhile, if a DD crash at a roundabout in the states of Florida and Massachusetts, the probability of the occurrence of severe injuries is expected to decrease by 39 % and 50 %, respectively. Again, roundabouts appear to be an effective measure to mitigate the risk of severe outcomes of DD-related crashes. Lastly, all states showed a significant reduction in the likelihood of severe injuries of DD crashes ranging from 24 to 94 % in an urban environment. This is consistent with the Traffic Safety Facts (National Highway Traffic Safety Administration, 2017, 2018), in which the fatality rate of motor vehicle-related crash was found to be 2.5 times higher in rural areas than that in urban areas (1.96 and 0.79, respectively).

6. Conclusion

With the increase in distracted driving (DD) due to the adoption of numerous in-vehicle equipment and devices behind the wheel, DD has gained significant attention among many stakeholders. However, an empirical investigation examining the association between the contributing factors of DD-related crashes and the severity of the outcome has not yet been fully achieved. This study evaluates the role of the built environment, such as roadway infrastructure, zoning specification, and the urban/rural distinction, on the severity of vehicle crashes caused by distracted driving. Our study is one of the first attempts to compare these influences across different states in the U.S., so as to understand the heterogeneous effects of the outcomes and provide sound policy implications to improve roadway safety.

The study shows that at the lower severity level (Injury vs. PDO), the probability of injury is likely to be lower when a DD crash occurs at a roundabout or in an urban area. However, the probability of injury is likely to be higher when the crashes involve speeding and when it occurs in the environments, such as intersections and curved roads. In addition, mixed outcomes were found in work zones, school zones, and various types of road among various states, indicating state-specific heterogeneity. Similarly, at a higher crash severity level (Severe Injury vs. Injury & PDO), if a DD crash involves a work zone, curved roadway, or exceeding the speed limit, then it is more likely to result in severe injuries rather than minor injuries or PDO.

Overall, the research findings provide the following policy implications for improving traffic safety. First, the research outcome provides evidence to help transportation planners and policymakers better understand the impact of the built environment on the severity of crashes caused by distracted driving. Furthermore, the comparison of the results among different states helps to identify the best practice of infrastructure planning and policies for distracted driving countermeasures. For instance, given that the empirical findings show that the severity of vehicle crashes caused by distracted driving tends to be relatively higher in work zones than in other road environments, we suggest that strong work zone safety regulations are needed to reduce severe injuries or fatalities among these areas. In the meantime, more efforts, such as providing better signage and adjusting vehicle speed at the work zone, are needed to enhance public awareness of work zone safety.

Secondly, given that roundabouts were found to be associated with a lower probability of crash severity of distracted driving-related crashes, more attention are needed to further assess the effectiveness of roundabout for urban roadway safety.[8]

[5] For instance, as shown in Appendix A6, the relative proportion of severe DD crashes caused by speeding only accounts for 1.22% of the total number of DD crashes in CA, which is much lower than other states. However, the case of speeding accounts for 45% of the total DD crash case, which is, again, significantly higher than other states. These disparities suggest that the relative small proportion of DD crashes that involve speeding is much lower than the other states. Such an outcome could reflect state-specific characteristics regarding DD, such as an effective roadway management and law enforcement. It may also be due to the different crash report systems, which should be considered as a caveat of the data source and needs to be interpreted carefully.

[6] Move Over/Slow Down law: Effective September 1, 2013, drivers must move over or slow down when approaching work zones.

[7] The Intersection Safety Implementation Plan (ISIP) in PA was developed and released in 2010. But the plan was not implemented until 2014.

[8] While our assessment reveals that roundabout has a positive role in reducing the severity of DD-related crashes, one should note that roundabouts are also found to be unsafe for pedestrians and cyclists (Harkey and Carter, 2006; Cumming, 2011). Hence, we suggests that more research efforts are needed to understand the benefits and costs of roundabout design for roadway safety.

Thirdly, given that both the severity of vehicle crashes caused by distracted driving were found to be different in the urban/non-urban road environment, more attention should be paid to both a more targeted driver's education and stronger enforcement of distracted driving regulation in suburban and rural areas.

One should note that this study has several limitations that should be further improved in future research. Different from Oviedo-Trespalacios et al. (2020), our assessment was based on the GOL model, which only allowed us to capture the heterogeneity among various states and the heterogeneous influences of built environments on the severity of DD crashes. Hence, the results should be read with a caution, given that the study only presents the influence of the fixed effects (mean structure), whereas we did not consider the unobserved heterogeneity and/or random fluctuation from missing covariates. Therefore, future research could be expanded by adopting more advanced modeling frameworks, such as the generalized additive model and generalized linear mixed model to address non-linear relationships by smoothing functions and random effects such as unobserved heterogeneity (Faraway, 2016). In addition, a latent segmentation-based ordered logit (LSOL) model may also be adopted to capture heterogeneity or variation in crash severity levels (Fatmi and Habib, 2019).

Last but not least, while the analysis was based on standardized data of various states for comparison, it is worth noting that distracted driving behavior tends to be underreported. Therefore, the pattern revealed in this study may not fully capture the actual DD related crashes. Future research should also focus on the utilization of different data sources, such as the telematics data with integration of crash data in order to provide a more comprehensive understanding of the DD driving behavior.

Funding

The authors would like to thank to Michelle May, the Program Manager of the Highway Safety Program of ODOT, who provided funding support for this study. The research team also benefited from the interactive discussions with the Distracted Driving Leadership Team, including Philip Renaud II, Dr. Brittany Shoots-Reinhard, Sharon Montgomery, and Jim Noble. We also thank various state DOTs for kindly providing us access to their state crash datasets, and the research supports from Kartik Mukundan, Fahd Jehangir, Andy Barley, and Yuxuan (Norman) Wang, for collecting and processing the data. The authors would also like to express their gratitude to Dr. Philip Viton and the two anonymous reviewers for their valuable comments to improve the quality of this paper. Views presented in this study do not represent any organizations, and authors are responsible for any errors or mistakes.

CRediT authorship contribution statement

Youngbin Lym: Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Zhenhua Chen:** Conceptualization, Resources, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

All authors declare that there is no conflict of interest.

Appendix A

Generating consistent variables is the most time-consuming and

Table A1
Variable identification for distracted driving.

State	Variable	File	Define
CA	Other Associated Factor 1	Party	A - Violation E - Vision Obscurements F - Inattention (beginning 1/1/01, inattention not stated) G - Stop and Go Traffic H - Entering/Leaving Ramp ~ M - Other N - None Apparent O - Runaway Vehicle P - Inattention, Cell Phone (1/1/01) Q - Inattention, Electronic Equip. (1/1/01) R - Inattention, Radio/CD (1/1/01) S - Inattention, Smoking (1/1/01) T - Inattention, Eating (1/1/01) U - Inattention, Children (1/1/01) V - Inattention, Animal (1/1/01) W - Inattention, Personal Hygiene (1/1/01) X - Inattention, Reading (1/1/01) Y - Inattention, Other (1/1/01) - - Not Stated
NY	Cft_cde1 ~ 2	newyorkveh	Cft_cde1 CFT_DESCR 2Alcohol Involvement 3Backing Unsafely 4Driver Inattention/Distraction* 5Driver Inexperience* 6Drugs (Illegal) 7Failure to Yield Right-of-Way 8Fell Asleep 21Fatigued/Drowsy 9Following Too Closely 10Illness 11Lost Consciousness 12Passenger Distraction 13Passing or Lane Usage Improper 14Pedestrian/Bicyclist Error/Confusion 15Physical Disability ~ 69View Obstructed/Limited 80Other* −1Unknown −2Not Applicable −3Not Entered 22Cell Phone (hand held) 23Cell Phone (hands-free) 24Other Electronic Device* 25Outside Car Distraction* 26Reaction to Other Uninvolved Vehicle 27Failure to Keep Right 28Aggressive Driving/Road Rage 50Driverless/Runaway Vehicle 31Texting 29Passing Too Closely 30Vehicle Vandalism 32Using On Board Navigation Device 33Eating or Drinking 34Listening/Using Headphones 51Tinted Windows
OH	Distracted_driver_ind		Y - yes N - no
PA	Distracted	Flag	0.no 1. Yes

Table A2
Database creation form (using Pennsylvania as an example).

Variable	Name	File	Definition	Note
DD	DISTRACTED	pennsylvaniaflag	At Least One Driver Action Indicating a Distraction 0 = No, 1 = Yes	Key variable CRN 1 -> 1 Else -> 0
Severity	MAX_SEVERITY_LEVEL	~crash	INJ_SEVERITY 0 - Not injured 1 – Fatal Injury 2 – Suspected Serious Injury 3 – Suspected Minor Injury 4 - Possible Injury 8 - Injury/ Unknown Severity 9 - Unknown if Injured	Code 0 -> 1 Else -> 2 1, 2 -> 3 Key variable CRN
Number of Lanes	LANE_COUNT	~roadway		Use a key variable CRN
Workzone	Work_zone_ind	~crash	Did the crash occur in a work zone (Y/N)	Y -> 1 N -> 0
School zone	SCH_ZONE_IND	~crash	Did the crash occur in a School Zone? (Y/N)	Y -> 1 N -> 0
Light Condition	ILLUMINATION	~crash	ILLUMINATION 1 – Daylight 2 – Dark – no street lights 3 – Dark – street lights 4 – Dusk 5 – Dawn 6 – Dark – unknown roadway lighting 8 – Other 9 – Unknown (expired)	Re-Encoding 1 -> 1 5 -> 2 4 -> 3 2,3,6,8 -> 4 9 -> 9 Priority 4 > 3 > 2 > 1 > 9
Road Contour	RDWY_ALIGNMENT;	~vehicle	RDWY_ALIGNMENT 1 - Straight 2 - Curved 9 - Unknown GRADE 1 – Level Roadway 2 - Uphill 3 - Downhill 4 - Sag or bottom of hill 5 - Crest or top of hill 9 - Unknown	Re-Encode 1 -> 1 2 -> 2 Blank, 9 -> 9 Priority 2 > 1 > 9
Weather Condition	WEATHER	~crash	WEATHER 1 – No adverse conditions 2 – Rain 3 – Sleet (hail) 4 – Snow 5 – Fog 6 – Rain and fog 7 – Sleet and fog 8 – Other 9 – Unknown	Re-Encode 1 -> 1 5,6,7 -> 2 8 -> 3 2 -> 4 3,4 -> 5 9 -> 9 Priority 5 > 4 > 3 > 2 > 1 > 9
Undivided	Relation_To_Road	~crash	RELATION_TO_ROAD 1 – On roadway 2 – Shoulder 3 – Median 4 – Roadside (off trafficway; on vehicle area) 5 – Outside trafficway (in area not meant for vehicles) 6 – In parking lane 7 – Gore (intersection of ramp and highway) 9 – Unknown	Recoding If 3 -> 1 Else 0
Roadway System	Road_Owner	~roadway	ROAD_OWNER 1 - Interstate - non turnpike 2 - State highway 3 - County road 4 - Local road or street 5 - East-West portion of turnpike 6 - Turnpike spur (extension) 7 - Private Road 9 - Other or Unknown	Re-encoding 4,7 -> 1 3 -> 2 5,6 -> 3 1,2 -> 4 Blank,9 -> 9
Intersection	Intersection	~flag	0 = No, 1 = Yes	Keep it
Roundabout	Intersect_Type	~crash	INTERSECT_TYPE 00 – Mid-block 01 – Four way intersection 02 – "T" intersection 03 – "Y" intersection 04 – Traffic circle or Round About 05 – Multi-leg intersection 06 – On ramp 07 – Off ramp 08 – Crossover 09 – Railroad crossing 10 – Other 99 – Unknown (expired)	04 -> 1 Else -> 0
Speeding	Speeding	~flag	0 = No, 1 = Yes	Keep this
Urban Area	Urban_rural	~crash	1= Rural, 2=Urbanized, 3=Urban	Re-encoding Else -> 0 2,3 -> 1
Time of the day	Time_Of_Day	~crash		Keep this
Day of the week	Day_Of_Week	~crash		Keep this
Year	Crash_Year;	~crash		Keep this
Month	Crash_Month			Keep this
Location	Dec_Lat; Dec_Long	~crash	Latitude Longitude	Keep both columns
Crash_ID	CRN	~crash	Identifier	

Table A3
Variable coding suggestion and coding rules.

	Sample Coding (Ohio as an example)		
Variable	Original	Suggestion	Rules for Selection (Redundancy)
DD	Distracted Driving	Only Primary Cause as DD Considered	
		1 PDO or NONE	
Severity	1/2/3/4/5	2 Injury A/B/C	Fatality comes first 3 > 2 > 1
		3 Fatality / Kill	
Number of Lanes			Maximum
Workzone		1 Yes 0 No	1 > 0
Schoolzone		1 Yes 0 No	1 > 0
	1 Daylight	1 Daylight (1 and 7)	Darkest
	2 Dawn	2 Dawn	4 > 3 > 2 > 1 > 9
	3 Dusk	3 Dusk	
	4 Dark – Lighted Roadway	4 Dark (including 4,5,6,7,8 from	1 -> 1
Light Condition	5 Dark – Roadway Not Lighted	Left Column)	2, 3 -> 2
	6 Dark – Unknown Roadway Lighting	9 Else (9)	4 -> 3
	7 Glare		
	8 Other		
	9 Unknown		
	1 'Straight-Level'	1 Straight (1, 2)	Curve comes first
	2 'Straight-Grade'	2 Curve (3, 4)	
Road Contour	3 'Curve-Level'		2 > 1 > 9
	4 'Curve-Grade'	9 Else	
	9 'Unknown'		
	1 Clear	1 Clear	Snow given priority
	2 Cloudy	2 Cloudy (2, 3)	5 > 4 > 3 > 2 > 1 > 9
	3 Fog, Smog, Smoke	3 Wind (7, 8)	
	4 Rain	4 Rain	
Weather Condition	5 Sleet, Hail	5 Snow (5, 6)	1 -> 1
	6 Snow	9 Else (9)	2 -> 2
	7 Severe Crosswinds		4 -> 3
	8 Blowing Sand, Soil, Dirt, Snow		3, 5 -> 4
	9 Other/Unknown		
Undivided	Divided	1 Divided	Divided given priority
	Undivided	0 Undivided	1 > 0
	'01' 'Principal Arterial (Rural Interstate)'	1 Local (09, 19)	Principal Arterial given top priority
	'02' 'Principal Arterial (Rural Others)'	2 Collector (07, 08, 17)	1,2 -> Local
	'06' 'Minor Arterial (Rural)'	3 MArteral (06, 16)	3 -> Minor
	'07' 'Major Collector (Rural)'	4 PArterial (01, 02, 11, 12, 14)	4 -> Principal
	'08' 'Minor Collector (Rural)'	9 Else	
	'09' 'Local (Rural)'		
Roadway System	'11' 'Principal Arterial (Urban Interstate)'		
	'12' 'Principal Arterial (Urban-Freeway & Expressway'		
	'14' 'Principal Arterial (Urban-Other)'		4 > 3 > 2 > 1 > 9
	'16' 'Minor Arterial (Urban)'		
	'17' 'Collector (Urban)'		
	'19' 'Local (Urban)'		
	' ' 'Not Coded'		
Intersection		1 Yes 0 No	1 > 0
Roundabout		1 Yes 0 No	1 > 0
Speeding		1 Yes 0 No	1 > 0
Urban Area		1 Yes 0 No	1 > 0
Location	Latitude; Longitude		Take Average

Table A4
Comparison of most recent DD crashes related to other standards.

	OH	OR	TX	FL[3)]	PA	MA[4)]	ND[5)]
Counts (2017)							
DD Crash (A)	21,415	1152	70,584	40,485	15,630	28,562	798
Total Crash (B)	303,284	20,714	620,787	402,385	128,177	143,474	15,296
Licensed Drivers [2)] (C)	8,011,705	2,910,592	17,099,340	15,076,358	8,964,855	5,040,662	545,027
Driving Age (16+) (D)	9,363,729	3,369,446	21,761,872	17,276,704	10,460,707	5,602,296	600,642
Total Resident (E)	11,658,609	4,142,776	28,304,596	20,984,400	12,805,537	6,811,779	756,927
Proportion (%)							
DD/Crash (A/B)	7.06 %	5.56 %	11.37 %	10.06 %	12.19 %	19.91 %	5.22 %
DD/License (A/C)	0.27 %	0.04 %	0.41 %	0.27 %	0.17 %	0.57 %	0.15 %
Crash/License (B/C)	3.79 %	0.71 %	3.63 %	2.67 %	1.43 %	2.85 %	2.81 %
Licensed/Resident (C/E)	68.72 %	70.26 %	60.41 %	71.85 %	70.01 %	74.00 %	72.01 %
DD/Resident (A/E)	0.18 %	0.03 %	0.25 %	0.19 %	0.12 %	0.42 %	0.11 %
Crash/Resident(B/E)	2.60 %	0.50 %	2.19 %	1.92 %	1.00 %	2.11 %	2.02 %

Source: 1) For DD and Total Crashes: Authors' Database collected from each state agency.
2) Licensed Drivers and Others: Highway Statistics 2017 (Federal Highway Administration and U.S. Department of Transportation, 2019a).
Notes: 1) Crash Data are based on the 2017 crash report of each state.
2) The number of Licensed Drivers, Driving Age (16+), and Total Resident is based on the 2017 record.
3) Total Crashes in Florida are extracted from Crash Facts Annual Report for the period 2013–2017 (FHSMV, 2019).
4) MA: based on the 2016 crash dataset.
5) ND: based on the 2015 crash dataset (complete).

demanding procedure, in which we need to identify and to build a linkage between our selected variables and the actual crash database by means of an appropriate variable coding scheme. As emphasized, each state has a different crash database management system with its own variable definition as well as variable coding structure. For example, unlike the case of Ohio or Pennsylvania, where distracted driving is already coded as an indicator variable Yes or No, most states do not have this context in their crash database tables. The information on crashes caused by distracted driving is often hidden in contributing factors (New York), other associated factors (California) under multiple files. The upper panel of Appendix Table A1 presents the differences across the states in their distracted driving classification. To be more specific, under the Party file in California crash database, there are variables named "Other Associated Factor 1∼" and they are recorded by definition in HSIS Guidebook for California State Data Files (2014) which we had to further examine the attributes one by one. Similarly, in New York, crash contribution factors are located under the newyorkveh file with the variable name as "Cft_cdel1∼", also requesting further subjective judgment on the cause of distracted driving. Conversely, Ohio has a single crash database file that includes a distracted driving indicator, while the state of Pennsylvania provides an indicator for DD crashes under Flag file with "Distracted" variable.

Regarding our purpose of this research, the foremost step is proper identification of vehicle crashes caused by distracted driving that clearly draws the line for the further variable selection. Of several states that we have examined so far, we have identified that the state of Maryland does show a mismatch between their crash database and a variable codebook that we are forced to exclude as one of our candidate states.

The following step consists of an array of variable collection procedures and the aggregation of multiple crash database sources via the key variable usage. Unlike Ohio where each row represents unique crash cases with corresponding variables such as built environment, roadway infrastructure feature, the severity level of vehicle crashes, and environmental circumstances, those in other states are distributed in multiple file sources (e.g. crash/collision/accident, vehicle, and person/individual) that require linking. Regarding each variable described in Table 2 for the selected sample states, we have developed an initial database creation form, exemplified in Appendix Table A2. As observed in the third column of Table A2, there are several sources of crash database in Pennsylvania where we have to come up with a proper aggregation scheme in order to ensure consistency across states.

The final step prior to the modeling-based analysis requires an appropriate variable coding scheme for our variables of interest. For example, recording weather condition and light condition that varies over different states need a unified coding strategy to achieve consistency. States such as MA, NV, and NJ have three levels of classification for crash severity, whereas other states have five categories. To overcome these challenges, we developed variable coding suggestions and coding rules based on a comparison of three sample states (OH, OR, and TX) for our initial analysis, presented in Appendix Table A3. One thing worth noting is the last column of the variable selection rule. As mentioned, some states have multiple crash database files that include redundancies in their observations, which create unexpected problems with the merging and aggregation process. We had identified this issue when we attempted to test OR and TX for our initial case study. Processes up to Appendix Table A2, including relevant steps, seemed to allow us to generate a complete and consistent dataset for each state. When there exist redundancies in the key variable such as CrashID or an identifier while merging or aggregating, however, algorithms arbitrarily choose attributes without any consideration, clearly leading to errors in statistical analysis.

Table A5
Multi-state comparison of DD crash severity (by Odds ratio).

	CA	FL	ID	IN	MA	NC	ND	NV	NY	OH	OR	PA	TX	VA	WA	NJ[4)]
Injury vs. PDO																
Intercept	0.818***	1.219***	0.599***	0.206***	0.360***	0.412***	0.412***	0.996	1.266***	0.596***	0.494***	0.846***	0.339***	0.471***	0.657***	0.281***
No.Lanes		1.059***	1.042			0.985				1.005						
WorkZone	1.076	1.010***		1.01	0.732**		0.809	1.099***	0.849**	0.994	1.747	1.175**	0.988***	0.959		
SchoolZone				0.936	1.646**	1.609	0.231		0.895	0.710**		1.503***	0.875	1.013		
Light_Dawn	0.962	1.068***	0.944	0.921	1.054	1.041	0.863	1.177***	1.189***	0.895***	0.637***	0.926	1.074***		0.953	1.089
Light_Dark	0.896***	0.995***	1.047	1.043***	1.147***	1.178***	0.952	1.079***	1.096***	0.921***	0.594***	0.755***	0.947***		0.978	1.075
Road_Curved		1.177***	1.034	0.982		1.488***	0.902		0.706***	1.053*	1.032	1.03	1.196***	1.046**	1.015	1.156*
Weather_Cloudy	1.069**	1.081***	0.993	1.05	1		0.891	1.025***	0.720***	0.973	1.256**	0.903	1.067***	1.056	0.916***	0.975
Weather_Rain	0.972	0.900***	0.856	0.993	0.632***		0.692	0.774	0.975	0.919**	1.168	0.919**	0.889***	0.882***	0.871***	1.089
Weather_Snow	0.867	1.556***	0.923	0.97	0.97		1.247	0.739*	0.578***	0.730***	0.515	0.936	0.748***	0.687***	0.640***	0.696*
Divided		1.179***	1.342***	1.310***	1.074***		1.117			0.953*	1.044	0.761***	1.445***	1.014		0.888**
Minor_Road	0.774***	0.624***	1.448***	1.586***	1.259***	1.175***	1.344			1.169***	0.872	1.012	0.995	1.218***	0.956	1.653***
Principal_Road	0.636***	0.611***	1.750***	1.474***	1.232***	1.206**	1.776***			1.165***	0.746*	1.242***	0.889***	1.176***	0.826***	1.765***
Intersection	1.455***	1.406***		1.125***	1.246***	1.057***	1.205*			1.072***	1.239	1.255***	1.215***	1.434***	1.023	
Roundabout		0.752***		1.288	0.829***	0.554***	0.251			0.536**	1.391	0.572	0.709**		0.380***	
Speeding	1.123***	2.335***	3.767***		2.615***			1.338***	1.509***	1.759***	4.023***	1.694***	1.624***	1.260***		
Urbarea	1.225***	0.654***	0.757***	0.869***	0.699***	0.877***	0.543***		0.560***	0.710***	1.721***	1.049**	1.062***	0.848***	0.944**	
Year	1.006	0.995***	1.002	0.972***	0.996	0.968***	1.069	1.120***	0.999	0.988**	1.039	0.962***	0.973***	0.975***	0.968***	1.044**
Severe Injury vs. Injury & PDO																
Intercept	0.036***	5.305***	0.063***	0.028***	0.029***	0.007***	0.089***	0.028***	0.041***	0.072***	0.043***	0.015***	0.023***	0.051***	0.025***	0.001***
No.Lanes		1.017***	0.996			1.096***				0.959*						
WorkZone	1.321	1.183***		1.148	0.404		2.325*	1.159***	0.779	1.105	1.959	1.928***	0.897***	0.901		
SchoolZone				0.746*	2.14	0.394	0.069		1.589	0.479		0.935	1.426	0.915		
Light_Dawn	1.232	1.199***	1.159	0.99	1.012	1.138	0.895	1.214***	1.382***	1.038	0.474*	1.758***	1.580***		1.181	8.196***
Light_Dark	1.759***	1.353***	1.258	1.114***	1.405***	1.506***	0.748	1.339***	1.456***	1.089*	0.458**	1.171***	1.395***		1.429***	1.7
Road_Curved		1.655***	0.97	0.91		4.582***	1.012		1.316***	1.139*	1.129	1.539***	1.670***	1.531***	1.373***	8.092***
Weather_Cloudy	0.924	1.150***	0.935	0.936	0.981		0.842	1.107***	0.850***	0.929	1.389	1.487**	0.977***	1.092	0.858	2.281*
Weather_Rain	0.979	0.837***	0.471	0.809*	0.480***		0.378	0.594*	0.896**	0.760***	0.632	0.954	0.695***	0.844***	0.678***	1.182
Weather_Snow	1.46	14.090***	0.779	0.793	0.788**		0.901	1.658	0.759***	0.517***	0.92	1.208	0.529***	0.832	0.547	0.369
Divided		1.084***	1.484*	1.216***	1.105		0.936			1.162*	1.103	1.652***	3.064***	0.939*		0.556*
Minor_Road	1	0.026***	0.994	1.899***	1.13	1.227***	1.26			1.087	0.706	1.691***	1.301***	1.258***	1.05	1.846
Principal_Road	0.958	0.027***	1.568**	1.659***	1.101	0.662***	1.365			1.094	0.403**	1.231***	0.835***	1.390***	0.663*	4.647***
Intersection	1.035	1.262***		0.886**	1.188***	0.782***	1.597*			0.892**	0.351	0.884*	0.826***	1.538***	0.742***	
Roundabout		0.608***		0.701	0.499*	0.557	0.056			0.205	0	0.331	0.51		0	
Speeding	0.385***	10.828***	2.409*		3.229***			212.3***	1.858***	2.193***	3.118***	14.932***	3.576***	1.998***		
Urbarea	0.585***	0.240***	0.369***	0.549***	0.591***	0.227***	0.059***		0.711***	0.458***	0.762	0.537***	0.518***	0.506***	0.466***	
Year	1.047**	1.037***	1.039	1.309***	0.890***	1.176***	1.313**	1.032***	0.979**	0.943***	1.034	1.130***	0.947***	0.983**	1.029	0.976
Log Likelihood	−81,782		−4,160.9		−67,420				−233,666	−72,627	−4,183.4			−130,071	−47,240	
Num. obs.	211,786	534,312	9,860	70,802	218,940	183,600	5,494	24,038	574,724	196,840	12,408	149,888	835,446	319,354	141,850	26,608

Note: 1) Statistical significance level: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.
2) Reference categories: WorkZone (No), SchoolZone (No), Light (Daylight), Road_Curved (No), Weather (Clear), Divided (No), Minor_Road (Local_Road), Principal_Road (Local_Road), Intersection (No), Roundabout (No), Speeding (No).
3) Blank cells: data not available.

Table A6
Descriptive of speeding and severe DD crashes.

States	CA	FL	ID	MA	ND	NV	NY	OH	OR	PA	TX	VA
Counts												
Severe Speeding (A)	588	275	6	20	6	21	746	740	18	130	472	2502
Speeding (B)	48059	1552	41	495	88	126	10188	10620	327	1799	8018	24108
Severe (C)	2416	13556	219	1727	62	299	11101	3296	120	1470	8467	7990
DD Crashes (D)	105893	267156	4159	109470	2747	10784	287362	98536	6204	74944	349952	132155
Proportions (%)												
SevSpeed/Speeding (A/B)	1.22 %	17.72 %	14.63 %	4.04 %	6.82 %	16.67 %	7.32 %	6.97 %	5.50 %	7.23 %	5.89 %	10.38 %
SevSpeed/Severe (A/C)	24.34 %	2.03 %	2.74 %	1.16 %	9.68 %	7.02 %	6.72 %	22.45 %	15.00 %	8.84 %	5.57 %	31.31 %
Speeding/DD (B/D)	45.38 %	0.58 %	0.99 %	0.45 %	3.20 %	1.17 %	3.55 %	10.78 %	5.27 %	2.40 %	2.29 %	18.24 %
Severe/DD (C/D)	2.28 %	5.07 %	5.27 %	1.58 %	2.26 %	2.77 %	3.86 %	3.34 %	1.93 %	1.96 %	2.42 %	6.05 %
SevSpeed/DD (A/D)	0.56 %	0.10 %	0.14 %	0.02 %	0.22 %	0.19 %	0.26 %	0.75 %	0.29 %	0.17 %	0.13 %	1.89 %

References

Agresti, A., 2010. Analysis of ordinal categorical data. Hoboken, NJ. Wiley.

Aziz, H.A., Ukkusuri, S.V., Hasan, S., 2013. Exploring the determinants of pedestrian–vehicle crash severity in New York City. Accid. Anal. Prev. 50, 1298–1309. https://doi.org/10.1016/j.aap.2012.09.034.

Behnood, A., Mannering, F., 2017. Determinants of bicyclist injury severities in bicycle vehicle crashes: a random parameters approach with heterogeneity in means and variances. Anal. Methods Accid. Res. 16, 35–47.

Birrell, S.A., Young, M.S., 2011. The impact of smart driving aids on driving performance and driver distraction. Transp. Res. Part F Traffic Psychol. Behav. 14 (6), 484–493. https://doi.org/10.1016/j.trf.2011.08.004.

Brace, Charlotte L., Young, Kristie L., Regan, Michael A., 2007. Analysis of the literature: the use of mobile phones while driving. Analysis 37.3, 261–265.

Braitman, K.A., Braitman, A.L., 2017. Patterns of distracted driving behaviors among young adult drivers: exploring relationships with personality variables. Transp. Res. Part F Traffic Psychol. Behav. 46, 169–176. https://doi.org/10.1016/j.trf.2017.01.015.

Chen, Z., Lym, Y., 2020. The influence of built environment on distracted driving related crashes in Ohio. Transport Policy. https://doi.org/10.1016/j.tranpol.2020.11.011. In press.

Chen, P., Shen, Q., 2016. Road environment effects on cyclist injury severity in automobile-involved bicycle crashes. Accid. Anal. Prev. 86, 239–246. https://doi.org/10.1016/j.aap.2015.11.002.

Chen, P., Shen, Q., 2019. Identifying high-risk built environments for severe bicycling injuries. J. Safety Res. 68, 1–7. https://doi.org/10.1016/j.jsr.2018.11.002.

Christoph, M., Wesseling, S., van Nes, N., 2019. Self-regulation of drivers' mobile phone use: the influence of driving context. Transp. Res. Part F Traffic Psychol. Behav. 66, 262–272. https://doi.org/10.1016/j.trf.2019.09.012.

Collet, C., Guillot, A., Petit, C., 2010. Phoning while driving I: a review of epidemiological, psychological, behavioural and physiological studies. Ergonomics 53.5, 589–601.

Cumming, B., 2011. Roundabouts: why they are dangerous for cyclists and what can Be done about it. Trans. Eng. Aust. Vol. 13 (No. 1), 27–40, 2011.

Dabbour, E., Easa, S., Haider, M., 2017. Using fixed-parameter and random-parameter ordered regression models to identify significant factors that affect the severity of drivers' injuries in vehicle-train collisions. Accid. Anal. Prev. 107, 20–30. https://doi.org/10.1016/j.aap.2017.07.017.

Dabbour, E., Haider, M., Diaa, E., 2019. Using random-parameter and fixed-parameter ordered models to explore temporal stability in factors affecting drivers injury severity in single-vehicle collisions. J. Traffic Transp. Eng. 6 (2), 132–146. https://doi.org/10.1016/j.jtte.2018.04.002.

Ding, C., Chen, P., Jiao, J., 2018. Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: a machine learning approach. Accid. Anal. Prev. 112, 116–126. https://doi.org/10.1016/j.aap.2017.12.026.

Eluru, N., Yasmin, S., 2015. A note on generalized ordered outcome models. Anal. Methods Accid. Res. 8, 1–6. https://doi.org/10.1016/j.amar.2015.04.002.

Faraway, J.J., 2016. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. CRC Press, Boca Raton.

Fatmi, M.R., Habib, M.A., 2019. Modeling vehicle collision injury severity involving distracted driving: assessing the effects of land use and built environment. Trans. Res. Rec., 036119811984906 https://doi.org/10.1177/0361198119849060.

Federal Highway Administration, U.S. Department of Transportation, 2019a. Highway Statistics Series. Accessed Oct 1, 2019. https://www.fhwa.dot.gov/policyinformation/statistics.cfm.

Federal Highway Administration, U.S. Department of Transportation, 2019b. Intersection Safety. Accessed Dec 21, 2019. https://safety.fhwa.dot.gov/intersection/other_topics/fhwasa18047/.

Fitch, G.A., Soccolich, S.A., Guo, F., McClafferty, J., Fang, Y., Olson, R.L., Perez, M.A., Hanowski, R.J., Hankey, J.M., Dingus, T.A., 2013. The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk. (Report No. DOT HS 811 757). Washington, DC. National Highway Traffic Safety Administration.

Florida Highway Safety and Motor Vehicles (FHSMV), 2019. Traffic Crash Facts Annual Report 2013-2017. Accessed Oct 10, 2019. https://www.flhsmv.gov/resources/crash-citation-reports/.

Gershon, P., Zhu, C., Klauer, S.G., Dingus, T., Simons-Morton, B., 2017. Teens distracted driving behavior: prevalence and predictors. J. Safety Res. 63, 157–161. https://doi.org/10.1016/j.jsr.2017.10.002.

Gershon, P., Sita, K.R., Zhu, C., Ehsani, J.P., Klauer, S.G., Dingus, T.A., Simons-Morton, B.G., 2019. Distracted driving, visual inattention, and crash risk among teenage drivers. Am. J. Prev. Med. 56 (4), 494–500. https://doi.org/10.1016/j.amepre.2018.11.024.

Hancox, G., Morris, A., Richardson, J., 2013. Drivers' willingness to engage with their mobile phone: the influence of phone function and road demand. Iet Intell. Transp. Syst. 7 (2), 215–222. https://doi.org/10.1049/iet-its.2012.0133.

Handy, S.L., Boarnet, M.G., Ewing, R., Killingsworth, R.E., 2002. How the built environment affects physical activity: views from urban planning. Am. J. Prev. Med. 23 (2), 64–73.

Haque, M.M., Oviedo-Trespalacios, O., Debnath, A.K., Washington, S., 2016. Gap acceptance behavior of mobile phone–Distracted drivers at roundabouts. Trans. Res. Rec. 2602, 43–51. https://doi.org/10.3141/2602-06.

Harkey, D.L., Carter, D.L., 2006. Observational analysis of pedestrian, bicyclist, and motorist behaviors at roundabouts in the United States. Trans. Res. Rec. 1982 (1), 155–165. https://doi.org/10.1177/0361198106198200120.

Horrey, W.J., Lesch, M.F., Garabet, A., Simmons, L., Maikala, R., 2017. Distraction and task engagement: how interesting and boring information impact driving performance and subjective and physiological responses. Appl. Ergon. 58, 342–348. https://doi.org/10.1016/j.apergo.2016.07.011.

Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 108, 27–36. https://doi.org/10.1016/j.aap.2017.08.008.

Kaplan, S., Vavatsoulas, K., Prato, C.G., 2014. Aggravating and mitigating factors associated with cyclist injury severity in Denmark. J. Safety Res. 50, 75–82.

Kidd, D.G., Tison, J., Chaudhary, N.K., McCartt, A.T., Casanova-Powell, T.D., 2016. The influence of roadway situation, other contextual factors, and driver characteristics on the prevalence of driver secondary behaviors. Transportation Research Part F. Traffic Psychology and Behaviour 41, 1–9. https://doi.org/10.1016/j.trf.2016.06.004.

Kircher, K., Ahlstrom, C., 2012. The impact of tunnel design and lighting on the performance of attentive and visually distracted drivers. Accident Analysis & Prevention 47, 153–161. https://doi.org/10.1016/j.aap.2012.01.019.

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., Ramsey, D.J., 2006. The Impact of Driver Inattention on near-crash/crash Risk: an Analysis using the 100-car Naturalistic Driving Study Data. Report No. DOT HS 810 594. National Highway Traffic Safety Administration, Washington, D.C.

Knapper, A.S., Hagenzieker, M.P., Brookhuis, K.A., 2015. Do in-car devices affect experienced users driving performance? IATSS Res. 39 (1), 72–78. https://doi.org/10.1016/j.iatssr.2014.10.002.

Kountouriotis, G., Merat, N., 2016. Leading to distraction: driver distraction, lead car, and road environment. Accid. Anal. Prev. 89, 22–30. https://doi.org/10.1016/j.aap.2015.12.027.

Lansdown, T.C., 2012. Individual differences and propensity to engage with in-vehicle distractions—a self-report survey. Transp. Res. Part F Traffic Psychol. Behav. Vol. 15 https://doi.org/10.1016/j.trf.2011.09.001.

Lansdown, T.C., Stephens, A.N., 2013. Couples, contentious conversations, mobile telephone use and driving. Accid. Anal. Prev. 50, 416–422.

Lipovac, K., Đerić, M., Tešić, M., Andrić, Z., Marić, B., 2017. Mobile phone use while driving-literary review. Transp. Res. Part F Traffic Psychol. Behav. 47, 132–142.

Lym, Y., Chen, Z., 2020. Does space influence on the frequency and severity of the distraction-affected vehicle crashes? An empirical evidence from the Central Ohio. Accid. Anal. Prev. 144, 105606 https://doi.org/10.1016/j.aap.2020.105606.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. Anal. Methods Accid. Res. 1, 1–22. https://doi.org/10.1016/j.amar.2013.09.001.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16.

Mohamed, M.G., Saunier, N., Miranda-Moreno, L.F., Ukkusuri, S.V., 2013. A clustering regression approach: a comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada. Saf. Sci. 54, 27–37. https://doi.org/10.1016/j.ssci.2012.11.001.

National Highway Traffic Safety Administration, 2017. Traffic Safety Facts Research Note. National Highway Traffic Safety Administration (NHTSA), Washington, DC.

Report No. DOT HS 812 456 Accessed June 20, 2019. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812456.

National Highway Traffic Safety Administration, 2018. Traffic Safety Facts Research Note. National Highway Traffic Safety Administration (NHTSA), Washington, DC. Report No. DOT HS 812 517 Accessed June 20, 2019. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812517.

National Highway Traffic Safety Administration, 2019. Traffic Safety Facts Research Note. National Highway Traffic Safety Administration (NHTSA), Washington, DC. Report No. DOT HS 812 700 Accessed June 20, 2019. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812700.

Ohio Department of Transportation. (2019). http://www.dot.state.oh.us/Divisions/Planning/ProgramManagement/HighwaySafety/SHSP/DD_TaskForce/Forms/AllItems.aspx, Accessed June 21, 2019.

Olson, R.L., Hanowski, R.J., Hickman, J.S., Bocanegra, J., 2009. Driver Distraction in Commercial Vehicle Operations. U.S. Department of Transportation, Washington, D.C, 2009 Report No. FMCSA-RRR-09-042.

Overton, T.L., Rives, T.E., Hecht, C., Shafi, S., Gandhi, R.R., 2015. Distracted driving: prevalence, problems, and prevention. Int. J. Inj. Contr. Saf. Promot. 22 (3), 187–192.

Oviedo-Trespalacios, O., Haque, M.M., King, M., Washington, S., 2016. Understanding the impacts of mobile phone distraction on driving performance: a systematic review. Transp. Res. Part C Emerg. Technol. 72, 360–380.

Oviedo-Trespalacios, O., Haque, M.M., King, M., Washington, S., 2017. Effects of road infrastructure and traffic complexity in speed adaptation behaviour of distracted drivers. Accid. Anal. Prev. 101, 67–77. https://doi.org/10.1016/j.aap.2017.01.018.

Oviedo-Trespalacios, O., Haque, M.M., King, M., Washington, S., 2019. "Mate! I'm running 10 min late": an investigation into the self-regulation of mobile phone tasks while driving. Accid. Anal. Prev. 122, 134–142. https://doi.org/10.1016/j.aap.2018.09.020.

Oviedo-Trespalacios, O., Afghari, A.P., Haque, M.M., 2020. A hierarchical Bayesian multivariate ordered model of distracted drivers' decision to initiate risk-compensating behaviour. Anal. Methods Accid. Res. 26, 100121 https://doi.org/10.1016/j.amar.2020.100121.

Papantoniou, P., Papadimitriou, E., Yannis, G., 2017. Review of driving performance parameters critical for distracted driving research. Transp. Res. Procedia 25, 1796–1805. https://doi.org/10.1016/j.trpro.2017.05.148.

Pope, C.N., Bell, T.R., Stavrinos, D., 2017. Mechanisms behind distracted driving behavior: the role of age and executive function in the engagement of distracted driving. Accid. Anal. Prev. 98, 123–129. https://doi.org/10.1016/j.aap.2016.09.030.

Rakauskas, M.E., Gugerty, L.J., Ward, N.J., 2004. Effects of naturalistic cell phone conversations on driving performance. J. Safety Res. 35 (4), 453–464. https://doi.org/10.1016/j.jsr.2004.06.003.

Regan, M.A., Hallett, C., Gordon, C.P., 2011. Driver distraction and driver inattention: definition, relationship and taxonomy. Accid. Anal. Prev. 43 (5), 1771–1781. https://doi.org/10.1016/j.aap.2011.04.008.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accid. Anal. Prev. 43 (5), 1666–1676. https://doi.org/10.1016/j.aap.2011.03.025.

Tang, J., Liang, J., Han, C., Li, Z., Huang, H., 2019. Crash injury severity analysis using a two-layer Stacking framework. Accid. Anal. Prev. 122, 226–238. https://doi.org/10.1016/j.aap.2018.10.016.

Texas Department of Transportation. (2019). https://www.txdot.gov/driver/share-road/move-over-slow-down.html, Accessed December 21, 2019.

Uddin, M., Ahmed, F., 2018. Pedestrian injury severity analysis in motor vehicle crashes in Ohio. Safety. https://doi.org/10.20944/preprints201804.0035.v1.

Wang, X., Kim, S.H., 2019. Prediction and factor identification for crash severity: comparison of discrete choice and tree-based models. Trans. Res. Rec., 036119811984445 https://doi.org/10.1177/0361198119844456.

Washington, S., Karlaftis, M., Mannering, F., 2011. Statistical and Econometric Methods for Transportation Data Analysis, second edition. Chapman and Hall/CRC, BocaRaton, FL.

Williams, R., 2016. Understanding and interpreting generalized ordered logit models. The Journal of Mathematical Sociology 40 (1), 7–20. https://doi.org/10.1080/0022250x.2015.1112384.

Wilson, F.A., Stimpson, J.P., 2010. Trends in fatalities from distracted driving in the United States, 1999 to 2008. Am. J. Public Health 100 (11), 2213–2219. https://doi.org/10.2105/ajph.2009.187179.

Xin, C., Guo, R., Wang, Z., Lu, Q., Lin, P., 2017. The effects of neighborhood characteristics and the built environment on pedestrian injury severity: a random parameters generalized ordered probability model with heterogeneity in means and variances. Anal. Methods Accid. Res. 16, 117–132. https://doi.org/10.1016/j.amar.2017.10.001.

Yannis, G., Papadimitriou, E., Karekla, X., Kontodima, F., 2010. Mobile phone use by young drivers: effects on traffic speed and headways. Transp. Plan. Technol. 4 (33), 385–394.

Yasmin, S., Eluru, N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accid. Anal. Prev. 59, 506–521.

Yasmin, S., Eluru, N., Ukkusuri, S.V., 2014. Alternative ordered response frameworks for examining pedestrian injury severity in New York City. J. Transp. Saf. Secur. 6 (4), 275–300.

Zahabi, S., Strauss, J., Manaugh, K., Miranda-Moreno, L., 2011. Estimating potential effect of speed limits, built environment, and other factors on severity of pedestrian and cyclist injuries in crashes. Trans. Res. Rec. 2247, 81–90.

Zeng, Q., Gu, W., Zhang, X., Wen, H., Lee, J., Hao, W., 2019. Analyzing freeway crash severity using a Bayesian spatial generalized ordered logit model with conditional autoregressive priors. Accid. Anal. Prev. 127, 87–95. https://doi.org/10.1016/j.aap.2019.02.029.