

# Predicting Pedestrian Crashes in Texas' Intersections and Midblock Segments

**Natalia Zuniga-Garcia, Ph.D.**

Researcher Fellow,  
Department of Civil, Architectural and Environmental Engineering  
The University of Texas at Austin,  
[nzuniga@utexas.edu](mailto:nzuniga@utexas.edu)

**Kenneth A. Perrine**

Research Associate, Center for Transportation Research,  
The University of Texas at Austin  
[kperrine@utexas.edu](mailto:kperrine@utexas.edu)

**Kara M. Kockelman, Ph.D., P.E.**

(Corresponding Author)  
Professor and Dewitt Greer Centennial Professor of Transportation Engineering  
Department of Civil, Architectural and Environmental Engineering  
The University of Texas at Austin,  
[kkockelm@mail.utexas.edu](mailto:kkockelm@mail.utexas.edu)

Under review for publication in *Sustainability*,  
Special Issue "Vulnerable Road Users in Safe System Approach"

## ABSTRACT

This study analyzes pedestrian crash counts at more than one million intersections and midblock segments using Texas police reports over ten years. Key contributions include methods for obtaining so many points and related variables across a vast network, while controlling for traffic control variables, highway design details, traffic attributes, and land use information across multiple sources. A negative binomial model for crash counts across the State of Texas and within the City of Austin suggests that signalized intersections, arterial roads, more lanes, narrower or non-existent medians, and wider lanes coincide with higher crash rates, per vehicle-mile traveled (VMT) and per walk-mile traveled. The analysis suggests that daily VMT increases the likelihood of pedestrian crashes, and midblock segments are more vulnerable than intersections, where one standard deviation increase of VMT causes an increase in crashes at intersections and midblock sections of 52% and 187%, respectively. Also, the number of intersection crashes in Austin is higher than in the rest of Texas, but the number of midblock crashes is lower. The analysis of the Austin area suggests that the central business district location is critical, with midblock crashes being more sensitive (240%) in this area than intersection (78%) crashes. Also, a significant inequity was found in the area: an increase of \$41,000 in average household income leads to a reduction of 32% (intersections) and 39% (midblock) in pedestrian crash rates.

**Keywords:** Pedestrian crashes, intersection safety, midblock crashes, roadway inventory, negative binomial crash count models.

## INTRODUCTION

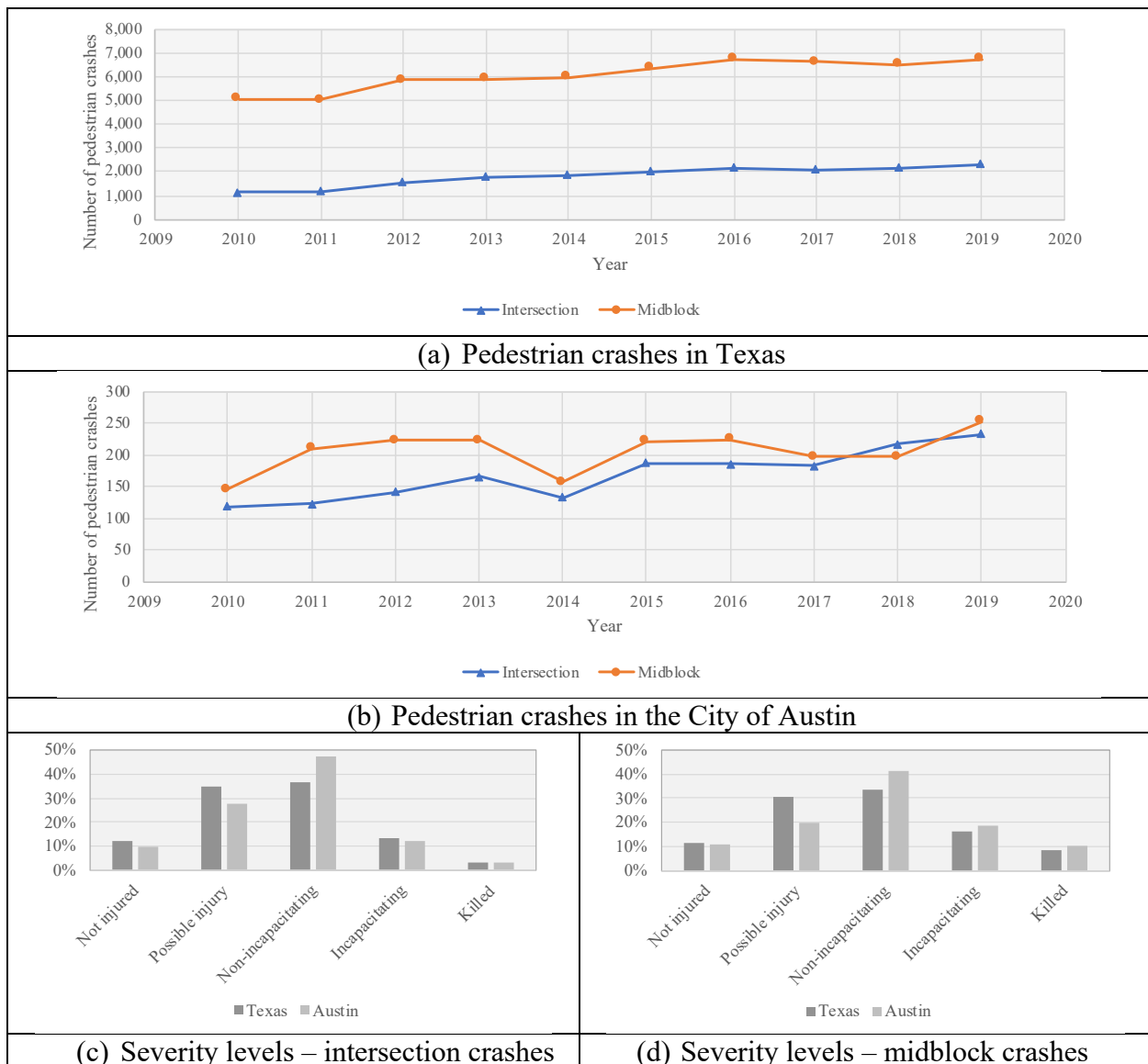
Despite being the oldest and most environmentally friendly form of transportation, walking has become increasingly risky in the U.S. While the total walk-miles traveled (WMT) is estimated to have risen 16% (BTS, 2019) between 2009 and 2017, the number of (reported) pedestrian deaths rose 46% (GHSA, 2020). Texas averaged 1.14 pedestrian deaths per 100,000 residents in 2019 (GHSA, 2020), which is 26% higher than the U.S. average of 0.900. Transportation planners and policymakers can reduce crash risks by implementing countermeasures based on benefit-cost analyses (BCA). However, such analyses do best with site-specific evaluations that are difficult to do at scale, due to a lack of detailed road feature variables. Between 2010 and 2020, the number of intersection crashes doubled in both Texas and the City of Austin, while midblock segment crashes rose 30% and 75%, respectively (Figure 1). Figure 1 shows how midblock pedestrian fatalities are more prevalent than those at intersections (where vehicle speeds are often lower, and pedestrians are more expected), at a rate of more than 3 to 1.

Most past pedestrian safety studies use macro-level information, with data aggregated at traffic analysis zones (TAZs) (Siddiqui et al., 2012), census tracts (Wier et al., 2009), census blocks groups (Noland et al., 2013), and zip codes (Ukkusuri et al., 2012) while studies at point level are more limited. In the past, a key limitation of intersection-safety studies is the sample size due to a lack of information across wide networks. Xie et al. (2018) used a Bayesian measurement error model with 262 signalized intersections in Hong Kong and found that the number of crossing pedestrians and passing vehicles, the presence of curb parking, and the presence of shops were associated with higher pedestrian crash counts, while the presence of playgrounds came with lower counts. Pulugurtha and Sambhara (2011) used 176 randomly selected signalized intersections in Charlotte, North Carolina, to understand the factors affecting pedestrian crash counts using a negative binomial (NB) count model using different buffers widths (0.25-mile, 0.5-mile, and 1-mile) to extract data at intersection levels. They found that using a 0.5-mile buffer width to extract demographic, socio-economic, and land use characteristics would yield better estimates for low pedestrian activity signalized intersections, 1-mile buffer width would yield better estimates in case of high pedestrian activity signalized intersections. Unfortunately, estimation of exposure variables, like site-level/highly local WMT, is challenging since pedestrian volumes are rarely available (unlike annualized vehicle counts). Studies have relied on surrogate measures, like the presence of schools and businesses, car ownership, pavement condition, sidewalk width, bus ridership, intersection control type, and presence of sidewalk barriers, to develop analyses, as seen in Lee et al. (2019).

Midblock level analyses are more common in past research since midblock crashes tend to be more severe and more common, and most transportation departments maintain corridor design variables. Kwayu et al. (2019) analyzed two years of crash information from Michigan and found that the average mid-block pedestrian crash took place while pedestrians were crossing 130 feet from the nearest intersection/crosswalk. Key predictors of more pedestrian deaths in these settings were a lack of lighting during a nighttime crash, crashes involving an older pedestrian, and crashes along corridors that carry higher traffic volumes. Diogenes and Lindau (2010) developed a Poisson regression using 21 midblock crosswalks and found that pedestrian crash counts rise with the presence of busways and bus stops, road widths, more traffic lanes, and higher volumes of pedestrians and vehicles. Analyses at both intersections and midblock segments for the same locations are limited and tend to focus on small areas (Lightstone et al., 2001) or are based on

comparisons of crash characteristics (Sandt and Zegeer, 2006). To the authors' knowledge, no published work has yet developed crash count models for both intersections and midblock segments across thousands of locations.

The main goal of this research is to develop a micro-level analysis of pedestrian crashes at intersections and midblock segments using historical pedestrian crash information from police reports in Texas. The contributions of this work include (1) a method to estimate intersection and midblock segment geometries with roadway characteristics that can be used to better understand crashes at different locations. (2) Using the developed geometries, NB pedestrian count models are estimated using historical pedestrian crash information from police reports in Texas to understand the factors associated at intersection and midblock levels. (3) A specific case study of the City of Austin is also estimated to understand local affecting factors.



**FIGURE 1 Description of pedestrian crashes at intersections and midblock (Texas Department of Transportation, 2020).**

## **DATA DESCRIPTION**

### **Crash Count Data**

The Texas Crash Records Inventory System is called CRIS (Texas Department of Transportation, 2020), and contains records from police crash reports across Texas' 254 counties and 268,597 square miles. Crash variables include time and location, persons and vehicles involved, injury severities, and road conditions. Many crashes are never reported to police or are not flagged for CRIS inclusion. These are typically proper-damage-only or no-injury crashes, but drivers and pedestrians will leave the scene for other reasons as well. And police who deem a reported crash to be worth less than the \$1,000 minimum crash cost threshold (for recording purposes) often do not record the crash formally.

From January of 2010 through December of 2019, 5.63 million crash incidents were recorded in the TxDOT CRIS database. A total of 78,498 crashes involving a pedestrian (including those caused or contributed to by a pedestrian, such as a driver swerving into a tree to avoid a pedestrian) are analyzed here. These are 1.40% of all reported and then recorded crashes.

### **Road Inventory Data**

The TxDOT Roadway Inventory database was used to obtain road-specific attributes (Texas Department of Transportation, 2018). The database is available in GIS shapefile and tabular format. Both on-system (under the jurisdiction of TxDOT) and off-system roads (not under TxDOT jurisdiction) are included in the database. The centerline miles show the mileage of a segment, regardless of the number of lanes, while the lane miles include the mileage of all lanes. Important road attributes include highway design and traffic characteristics such as daily vehicle miles traveled (DVMT), annual average daily traffic (AADT), percentage of truck, shoulder and median types and width, number of lanes, and speed limit.

### **Other Data Sources**

Other data sources include school, hospital, and transit stops, such as school locations from the Texas Education Agency, hospital locations from the Homeland Infrastructure Foundation, and transit stop locations from Capital Metro Transit Agency. The Capital Area Metropolitan Planning Organization (CAMPO) TAZ data were used for population and employment density for household income for the City of Austin. The WMT data are obtained at the individual respondent level, via the 2017 National Household Travel Survey (NHTS) and modeled as a function of respondent-level demographics and local land-use variables (such as population and job density of the respondent's home census tract), and then scaled up to Public Use Microdata Area demographics, based on methods found in (Rahman et al., 2021).

## **GEOMETRY ESTIMATION**

### **Purpose of Geometry Estimation**

To facilitate the analysis of pedestrian-related crashes, it is necessary to spatially model crash locations with respect to known roadways, intersections, and traffic signals. This allows analyzing crash-prone "hotspots" tied to intersections and segments. This research focuses upon crashes that take place at mid-block locations or places that are not associated with an intersection between public streets. Two criteria are considered in determining whether a crash is "midblock." First, the original police report contains a field that indicates whether the officer categorizes the crash as happening at an intersection. Unfortunately, it is sometimes unclear as to whether that intersection

is the center “box” area where public streets typically cross, a right turn yield link, or a roundabout. Second, the geographic crash location analysis against a map may show that the crash occurs far enough away from an intersection that it may be considered a midblock crash.

#### Uniform Segments

To support research analysis efforts, an underlying representation of roadway segments is prepared, and pedestrian-related crashes nearest to these segments are then associated with these segments. A geographic database of road segments with “multi-line string” geometry found in the TxDOT Roadway Inventory serves as the starting representation for all roadways in Texas. Each of these come labeled with street name, physical roadway characteristics such as functional class and lane count, estimated daily traffic volume, and maintenance information. In that inventory, each roadway consists of one or more segments. These segments are generally bidirectional except for those that represent one-way streets and divided highways.

A challenge in using the TxDOT Roadway Inventory is that individual roadway segments may be extremely short—a minimum of 5 feet—to represent a high rate of changes in inventory values along a roadway, or extremely long—up to 44 miles—for roadways that see few changes. Project activities benefit from the use of fairly consistently spaced segments, necessitating a remapping effort to create derived datasets of mostly uniform segments.

Two derived sets are created for this research: 1-mile-long target segment length and 0.1-mile-long target segment length. In each, key criteria of the underlying segments that overlap the most from the TxDOT Roadway Inventory are mapped to these new segments. To create these, an algorithm divides up roadways to a new set of segments targeted at length  $\ell$  using these rules:

- If the original roadway is less than  $1.25 \cdot \ell$  miles, then the derived roadway is represented with one segment of the same length.
- If the original roadway is less than  $2(1.25 \cdot \ell)$  miles, then the derived roadway is represented with two segments of equal length.
- Otherwise, the derived roadway is represented as starting and ending with segments no less than  $0.75 \cdot \ell$  miles on either end, with  $\ell$ -mile segments in-between.

#### Intersections and Signals

Mapping locations of crashes to nearby intersections is also important for this research to distinguish from midblock crashes. Unfortunately, the TxDOT Roadway Inventory does not offer explicit intersection point geometry. Although one can analyze the Roadway Inventory to evaluate where roadway segments intersect, it is impossible to understand where bridges exist (especially near expressway interchanges), leading to numerous false-positive intersections, especially around expressways.

OpenStreetMap (OpenStreetMap contributors, 2021a) was leveraged to positively map intersection and signal locations and apply them to appropriate locations in the TxDOT Roadway Inventory geometry. To do this, queries to a local instance of the Overpass API (OpenStreetMap contributors, 2021b) were created among approx. 30x30-mile tiles to return all candidate intersection seed points throughout Texas. This initial set came with caveats that many OpenStreetMap intersections on divided roads of all types exist for each direction. In contrast, the

intersections needed for the TxDOT Roadway Inventory would be needed per roadway. Other criteria were sought in positively identifying this first set of candidate intersections:

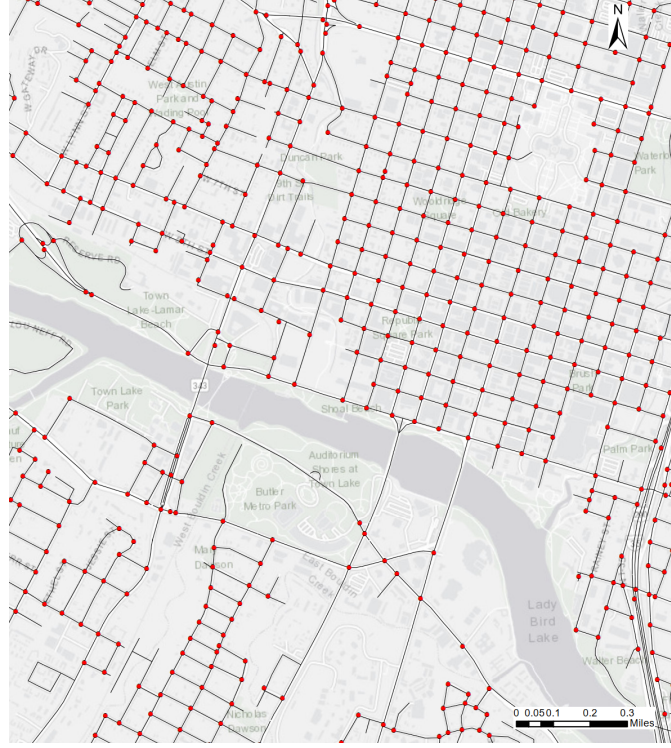
- It has a signal (tag “highway”: “traffic\_signals”). This will also catch signals for mid-block crossings. Or,
- It is met by more than one motorway that has a different type and name combination.
- Nodes serviced by only motorways and motorway links are labeled as a “junction.”
- Nodes that are joined by the ends of just 2 OpenStreetMap roadways are not counted, as they are likely a continuous stretch of roadway.

Next, to combine together candidate intersections that were closely positioned next to each other, the DBScan clustering algorithm (Ester and Kriegel, 1996) available in the PostGIS/PostgreSQL database (PostGIS, 2021) was leveraged to combine intersections that are less than 250 feet in distance from each other. Finally, roadway segments in the 1-mile and 0.1-mile uniform segment sets were associated with the clustered candidate intersection locations through a nearest-proximity search, allowing for candidate intersections to be associated if they are less than 130 feet from the roadway geometry.

This approximation is in support of efforts to perform initial rounds of this research. It had been found that this nearest-neighbor method of matching intersections to geometry by proximity alone still results in erroneous matches, especially around closely-positioned, urban expressway on- and off-ramps. However, because this research has emphasized urban streets and corridors that do not lie along expressways, the success rate for the nearest-neighbor matching approach has empirically been sufficient. It is anticipated in future work that a map-matching strategy (Perrine et al., 2015) can map valid pathways through the OpenStreetMap roadway network to underlying TxDOT Roadway Inventory geometry, allowing intersections to be more successfully tied with only the roadway segments that are truly connected with those intersections.

### **Estimated Geometries**

An example of the estimated geometries is shown in Figure 2. The 0.1-mile roadway segments are shown along with the intersections in a close-up of the City of Austin downtown area. The geometry matches the roadway map. Figure 3 shows the total roadway segments and intersections for the City of Austin and for the entire state. A total of 700 thousand (~1-mile uniform) segments drawn from 575 thousand segments are used to describe the State of Texas, while the City of Austin consists of 20 thousand (0.1-mile uniform) segments and 41 thousand intersections. A detailed description of these geometries is summarized in Tables 1 and 2, showing summary statistics of infrastructure characteristics as well as pedestrian crash counts based on the crash type and map matching process.



**FIGURE 2 Roadway segments and intersections in the City of Austin downtown area.**

## CRASH COUNT MODELING

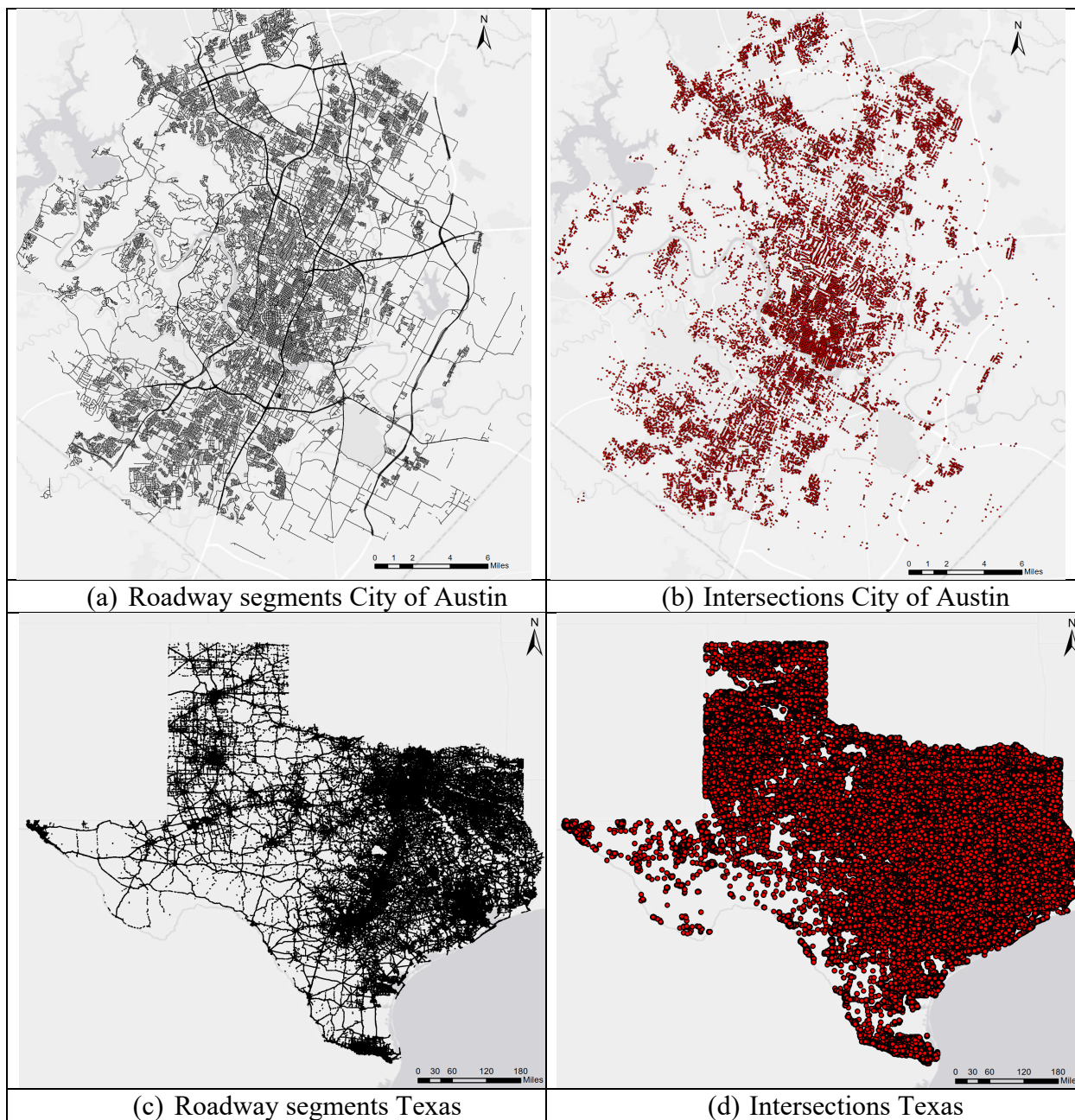
An NB count model was used for pedestrian crash counts. The expected number of pedestrian crash counts  $E(Y_i)$  along the  $i^{th}$  intersection or midblock segment is expressed as follows:

$$E(Y_i) = \exp(\beta_0 + \sum_k x_{ik}\beta_k + \varepsilon_i), \quad (1)$$

where  $\beta_k$  is the  $k^{th}$  covariate,  $\varepsilon_i$  is a random error term which follows a Gamma distribution  $\varepsilon_i \sim \text{Gamma}(\gamma, \gamma)$ ,  $Y_i$  represents the total pedestrian count at intersection  $i$  with mean  $E(Y_i) = \mu_i$  and variance  $\text{Var}(Y_i) = \mu_i + \rho\mu_i^2$ , and  $\rho$  is the dispersion parameter ( $\rho = 0$  for a Poisson model).

Additionally, a sensitivity analysis was applied to the NB estimates to understand the covariates' effects. Specifically, for each covariate, one standard deviation or binary change is applied. The modified variables are passed to the model to calculate the prediction. Then, the difference between the mean of original prediction and permuted prediction is calculated to represent the contribution of that covariate. Because of its appropriateness and suitable fit for modeling count data, this methodology has been applied in other research, including pedestrian crash occurrences at the segment level (Rahman et al., 2021) and e-scooter count models (Dean and Zuniga-Garcia, 2022).





1 **FIGURE 3 Description of the roadway segments and intersections.**



1 **TABLE 1 Summary statistics of variables, Texas**

Variables	Intersections*				Roadway segments (1 mi uniform)			
	Mean	S.D.	Min.	Max.	Mean	S.D.	Min.	Max.
Number of pedestrian crashes	0.020	0.215	0	31	0.062	0.502	0.0	37.0
Signalized intersection indicator	0.021	0.143	0	1				
Number of approaches	3.177	0.672	0	5				
Intersections crossed					2.781	2.693	0	25
Walk miles traveled per pop. dens.	325	453	0	15,339	244	398	0	15,339
Daily vehicle miles traveled	2,458	8,900	0	432,194	1,448	6,935	0	432,194
Speed limit (mph)	56.970	6.460	10	75	59	5	10	75
Number of lanes	2.229	0.713	1	8	2.068	0.401	1	12
Lane width (ft)	10.494	2.113	0	48	9.915	1.439	0	48
Median width (ft)	0.364	6.771	0	519	1.341	10.614	0	710
Design: one-way road indicator	0.009	0.096	0	1	0.011	0.102	0	1
Annual average daily traffic lane	953	1,852	1	142,733	527	1,615	1	100,335
Percentage of trucks	4.803	5.328	0	93	4.914	5.964	0	93
Functional class: local	0.677	0.467	0	1	0.814	0.389	0	1
Functional class: collector	0.178	0.382	0	1	0.110	0.313	0	1
Functional class: arterial	0.145	0.352	0	1	0.075	0.264	0	1
On system roadway indicator	0.150	0.357	0	1	0.137	0.343	0	1
Rural (pop. < 5000)	0.273	0.445	0	1	0.400	0.490	0	1
Small urban (pop: 5000–49999)	0.120	0.324	0	1	0.088	0.283	0	1
Urbanized (pop: 50000–199999)	0.109	0.312	0	1	0.086	0.281	0	1
Large urbanized (pop: 200000+)	0.498	0.500	0	1	0.426	0.495	0	1
Distance to nearest hospital (miles)	5.114	5.167	0	19	6.383	5.639	0	19
Transit stops within 0.25-mi buffer	0.021	0.144	0	1	0.021	0.144	0	1
Number of stops (0.25-mi buffer)	0.066	0.623	0	26	0.079	0.813	0	44
City of Austin indicator	0.027	0.163	0	1	0.023	0.151	0	1

S.D. = standard deviation, Min. = minimum, Max. = maximum

\* Major approach only

**TABLE 2 Summary statistics of variables, City of Austin**

Variables	Intersections*				Roadway segments (0.1 mi uniform)			
	Mean	S.D.	Min.	Max.	Mean	S.D.	Min.	Max.
Number of pedestrian crashes	0.074	0.489	0	13	0.042	0.274	0	9
Signalized intersection indicator	0.042	0.201	0	1				
Number of approaches	3.043	0.659	0	5				
Intersections crossed					0.994	2.113	0	5
Walk miles traveled per pop. dens.	756	808	7	8,180	647	736	7	8,180
Daily vehicle miles traveled	2,627	7,182	0	133,254	4,133	11,904	1	133,254
Speed limit (mph)	56.737	5.698	10	65	57.359	5.001	10	70
Number of lanes	2.259	0.731	1	6	2.266	0.736	1	6
Design: one-way road indicator	0.013	0.114	0	1	0.049	0.216	0	1
Annual average daily traffic lane	1,499	2,753	58	97,049	2,123	4,645	2	97,049
Percentage of trucks	3.294	0.708	0	20	3.398	1.024	0	20
Functional class: local	0.718	0.450	0	1	0.708	0.455	0	1
Functional class: collector	0.165	0.371	0	1	0.161	0.368	0	1
Functional class: arterial	0.117	0.321	0	1	0.131	0.337	0	1
On system roadway indicator	0.039	0.194	0	1	0.123	0.328	0	1
Distance to nearest hospital (miles)	2.092	1.279	0	10	2.258	1.417	0	10
Transit stops within 0.25-mi buffer	0.220	0.414	0	1	0.211	0.408	0	1
Number of stops (0.25-mi buffer)	1.022	2.391	0	18	1.038	2.562	0	21
Population density (per sq. mile)	4,443	3,398	0	64812	3,775	3,282	0	64812
Employment density (per sq. mile)	2,344	9,526	0	419,403	2,196	8,469	0	419,403
Median income (\$10k)	7.193	3.891	0	25	7.228	4.109	0	25
Central business district indicator	0.011	0.105	0	1	0.008	0.089	0	1

S.D. = standard deviation, Min. = minimum, Max. = maximum

\* Major approach only

## RESULTS AND DISCUSION

The results from the NB model are summarized in Table 3 (Texas) and Table 4 (Austin). Two models are estimated for each location; an intersection and midblock-level model is described along with sensitivity analysis summarized in Figures 4 and 5. The dispersion parameter ( $\rho$ ) of the four models is greater than one, indicating that the data is over-dispersed, and an NB model is preferred over a Poisson model.

The Texas model (Table 3) shows a positive correlation between the WMT and the number of pedestrians crashes across intersections and midblock-segment models, likely due to increased exposure levels. However, previous research also found that the relationship between crash exposure and crash rates is non-linear. It has rates falling off dramatically as walk levels rise, presumably due to drivers expecting more pedestrians in high-WMT zones and safer pedestrian environments that encourage walking (Wang and Kockelman, 2013).

The signalized intersection indicator in the intersection model is found to be among the most significant variables in the model. The sensitivity analysis indicates that the number of pedestrian crashes is doubled with signalized intersections when compared with unsignalized intersections with everything else remaining constant. Although signals are relatively safer than other control types for high pedestrian activity areas, higher usage increases the risk of accidents, as found in previous research (Lee et al., 2019; Xie et al., 2018). Similarly, the number of approaches (intersections) and the number of intersections crossed by the roadway segment show a positive, significant effect on the rate of pedestrian crashes. Specifically, one standard deviation of the number of approaches (0.67) leads to an increase of 31% in pedestrian crashes, while the standard deviation of 2.8 intersections crossed increases the crash rate 29%.

Among the highway design variables, the estimates indicate that higher DVMT significantly increases the number of pedestrian crashes. Interestingly, the effect of DVMT is more critical for the midblock segments than for the intersection model, as suggested by the sensitivity analysis, where one standard deviation increase causes 52% more crashes at intersections and 187% more crashes at midblock sections. Other variables such as the number of lanes and lane width also contribute to higher pedestrian crash rates. Higher posted speed limits and longer median widths, in contrast, tend to coincide with a reduction in crash rates. Roadways with a high speed limit are related to higher risk, and pedestrians tend to avoid these areas as there is limited pedestrian infrastructure. However, research findings indicate that, although the crash frequency is lower, the severity is significantly higher in areas with high speed (Bernhardt and Kockelman, 2021; Rahman et al., 2021; Zhao et al., 2021).

One-way roads are related to fewer crashes at midblock segments, while this variable was not significant for intersections at a 95% confidence level. The effect on midblock crashes is likely due to the reduced exposure on segments (less distance to cross). The intersection model suggests that variables from both major and minor approaches have a similar effect. The minor approach has a slightly lesser (but still significant) impact, except for the number of lanes, which shows no significant effect. However, the one-way road (minor) indicator variable showed statistical significance for the model and as found for midblock segments, the effect is negative, suggesting that one-way roads led to fewer pedestrian crashes at intersections as well.

Traffic attributes such as AADT and the percentage of trucks are also highly related to an increase in the crash rate at intersections. The roadway functional class is compared with local and collector roads using the indicator variables for arterials (freeways and highways are not included in the model). The results indicate a positive and significant effect, where arterial roads tend to have a higher number of pedestrian crashes than local and collector roads for both intersections and midblock segments. This effect is consistent in previous research where arterials are categorized as a health problem due to high pedestrian crash frequency, excessive noise and pollution (McAndrews et al., 2017). TxDOT-maintained roads (“on-system roads”) show a negative effect, suggesting that the number of pedestrian crashes is lower on these roads compared to other roadways. However, this finding contrasts previous research analysis of pedestrian crashes that uses the Texas Roadway Inventory geometry (Rahman et al., 2021). This may be because this study developed an analysis at a smaller scale with separations between intersections crashes and midblock crashes, while previous studies aggregated crashes and analyzed only segment-level variables.

1 **TABLE 3 Estimation results of NB for pedestrian crashes, Texas**

	Intersections			Midblock Segments		
	Coeff.	Std. Error	P-value	Coeff.	Std. Error	P-value
(Intercept)	-8.694	0.216	0.000	-8.035	0.098	0.000
WMT per pop dens. (log)	0.335	0.013	0.000	0.305	0.007	0.000
Signalized intersection (ind.)	1.426	0.032	0.000			
Number of approaches	0.398	0.019	0.000			
Intersections crossed				0.093	0.002	0.000
DVM (log) [major]	0.195	0.008	0.000	0.522	0.006	0.000
Speed limit (mph) [major]	-0.020	0.002	0.000	-0.013	0.001	0.000
Number of lanes [major]	0.132	0.012	0.000	0.217	0.010	0.000
Lane width (ft) [major]	0.033	0.004	0.000	0.041	0.003	0.000
Median width (ft) [major]	-0.006	0.001	0.000	-0.014	0.001	0.000
One-way road (ind.) [major]	0.095	0.052	0.068	-0.906	0.048	0.000
DVM (log) [minor]	0.136	0.008	0.000			
Speed limit (mph) [minor]	-0.021	0.002	0.000			
Number of lanes [minor]	-0.004	0.018	0.842			
Lane width (ft) [minor]	0.040	0.005	0.000			
Median width (ft) [minor]	-0.027	0.005	0.000			
One-way road (ind.) [minor]	-0.211	0.063	0.000			
AADT per lane [major]	1.76E-05	4.53E-06	0.000	-7.67E-05	4.19E-06	0.000
Truck percentage [major]	0.020	0.003	0.000	0.003	0.002	0.100
Arterial (ind.) [major]	0.444	0.037	0.000	0.198	0.028	0.000
On system roadway (ind.)	-0.230	0.036	0.000	0.209	0.028	0.000
Rural (ind.)	-0.107	0.087	0.218	-0.339	0.041	0.000
Small urban (ind.)	-0.108	0.055	0.050	0.049	0.034	0.154
Large urbanized (ind.)	0.171	0.037	0.000	0.170	0.025	0.000
Distance to nearest hospital (mi)	-0.023	0.006	0.000	-0.009	0.003	0.002
Transit stops (ind.)	0.525	0.047	0.000	0.526	0.033	0.000
Number of stops	0.042	0.008	0.000	0.049	0.004	0.000
City of Austin (ind.)	0.327	0.047	0.000	-0.392	0.042	0.000
<b>No. of observations</b>	<b>699,954</b>			<b>574,910</b>		
<b>Dispersion Parameter (<math>\rho</math>):</b>	0.393			0.575		
<b>McFadden's R2:</b>	0.483			0.543		
<b>Likelihood ratio test (<math>\chi^2</math>)</b>	32,515			62,980		
<b>Prob &gt; <math>\chi^2</math></b>	0.000			0.000		
<b>2 x log-likelihood</b>	-86,105			-161,539		

2



1 **FIGURE 4 Sensitivity analysis, Texas.**

Population is accounted for in the land use variable where urbanized areas (having a population of 50,000-200,000) are compared to rural, small urban, and large urbanized areas (refer to Table 1 for more details). Compared to urbanized areas, large urbanized areas (population greater than 200,000) show a positive coefficient, suggesting a positive effect. In contrast, small urban areas show a negative effect, suggesting that the rate of crashes is lower in these areas. This effect is consistent with expectations since denser areas have a higher number of pedestrians, and the exposure is higher due to the presence of more vehicles. The distance to the nearest hospital is also analyzed; the coefficient is found to be negative and significant at a confidence level of 95%. This suggests that the number of crashes is lower in areas with hospitals located at a close distance, which is comparable to the finding of higher crashes in denser areas. The sensitivity analysis indicates that one standard deviation of distance (approximately 5 miles) leads to a decrease of 11% (intersections) and 5% (midblock segments) in pedestrian crashes. However, it is important to mention that although the number of crashes is low, the distance to the hospital can be critical for response time and prompt injury treatment.

The presence of public transportation is an indirect measure of pedestrian exposure. In this study, transit information is included in the model through two variables. The first one is an indicator variable of the presence of transit within a buffer of 0.25 miles from the geometry centroid, and the second one is the count of transit stops in this area. As expected, both variables indicate that the number of pedestrian crashes is higher in the presence of transit, with a significant level of 95%. The sensitivity analysis suggests that the crash rate is about 50% higher in comparison to areas with the same characteristics without transit stops.

The Texas model includes a variable to indicate the location of the City of Austin and compare crashes with the rest of the state. Interestingly, the number of intersection crashes in Austin is higher than in the rest of Texas, but the number of midblock crashes is lower. Similarly, Figure 1 suggests that the number of intersection crashes in Austin is comparable to the number of midblock crashes, mainly in recent years (after 2017).

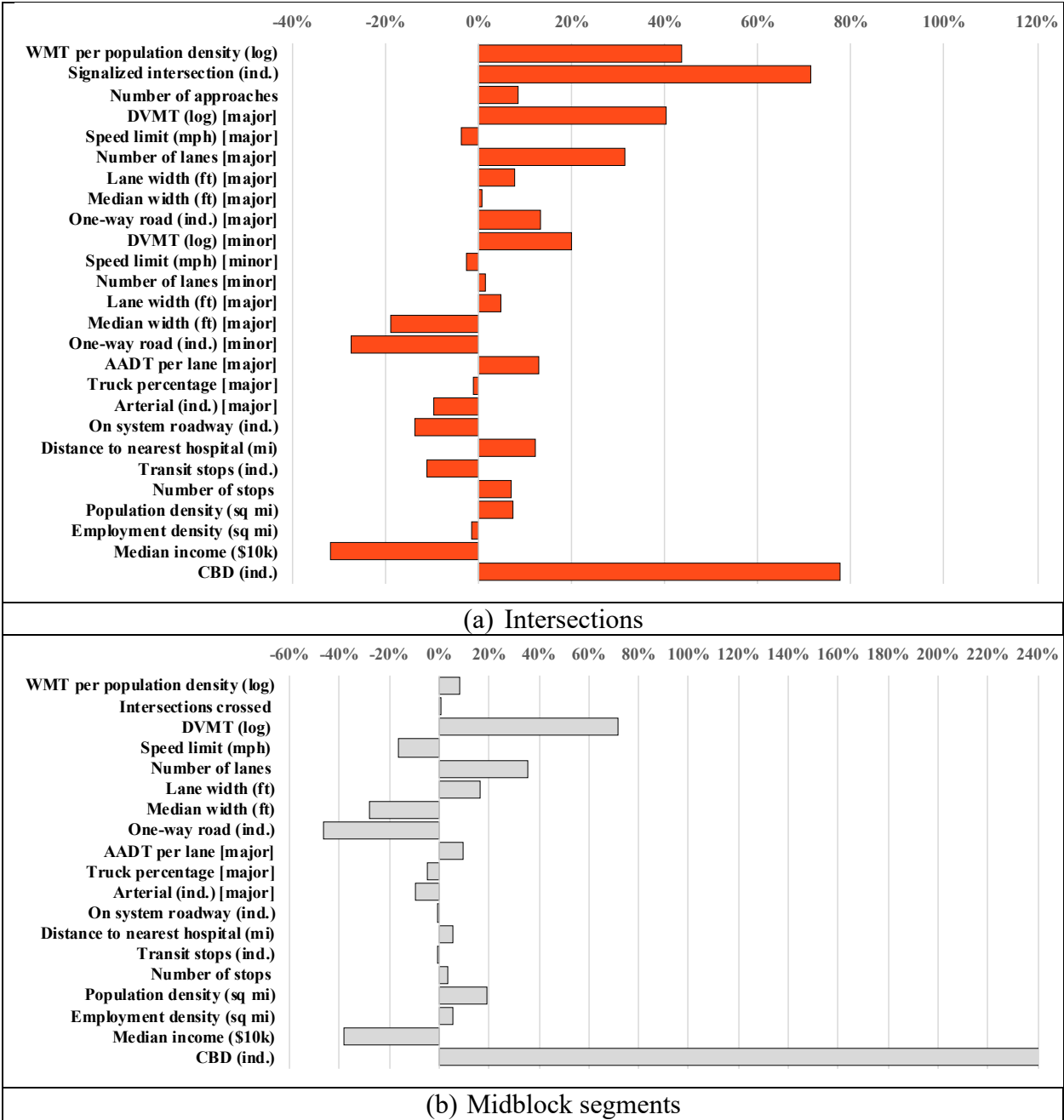
The Austin-specific model in Table 4 shows an intersection model that is less sensitive to WMT and signalized intersections than the Texas model but still shows a significant value. It is likely that pedestrian crashes in non-signalized intersections are more frequent in this area compared to the state of Texas. The number of approaches also shows a positive coefficient. In terms of midblock segments, the number of intersections crossed is not significant, possibly due to the size of the segments. In this case, the segments are 0.1-mile long, and the number of intersections crossed is significantly lower than the case where 1-mile segments were used in the Texas model.

DVMT has a positive significant effect on the number of crashes in the City of Austin. However, the sensitivity analysis suggests that the effect is less than the results obtained from the Texas model. One standard deviation increase of DVMT leads to a 40% (intersection) and 72% (midblock) increase in the number of pedestrian crashes. The effect of the speed limit is only significant for the midblock segments and not for the intersections model, and the sensitivity change is comparable to the rest of the state. The number of lanes also has a positive and significant coefficient. Variables such as lane width, median width, and one-way roads are not significant at a 95% confidence level for the intersections model (in both cases, major and minor approaches).



1 **TABLE 4 Estimation results of NB for pedestrian crashes, City of Austin**

	Intersections			Midblock Segments		
	Coeff.	Std. Error	P-value	Coeff.	Std. Error	P-value
(Intercept)	0.360	0.061	0.000	-5.098	0.473	0.000
WMT per pop. dens. (log)	1.671	0.103	0.000	0.068	0.043	0.114
Signalized intersection (ind.)	0.123	0.067	0.067			
Number of approaches						
Intersections crossed				0.001	0.012	0.904
DVM (log) [major]	0.166	0.028	0.000	0.245	0.024	0.000
Speed limit (mph) [major]	-0.007	0.006	0.248	-0.036	0.004	0.000
Number of lanes [major]	0.375	0.046	0.000	0.411	0.033	0.000
Lane width (ft) [major]	0.030	0.011	0.009	0.081	0.010	0.000
Median width (ft) [major]	0.001	0.002	0.726	-0.017	0.004	0.000
One-way road (ind.) [major]	0.159	0.175	0.735	-1.437	0.160	0.000
DVM (log) [minor]	0.145	0.029	0.000			
Speed limit (mph) [minor]	-0.012	0.009	0.177			
Number of lanes [minor]	0.057	0.063	0.359			
Lane width (ft) [minor]	0.036	0.013	0.006			
Median width (ft) [minor]	-0.049	0.020	0.013			
One-way road (ind.) [minor]	-0.458	0.182	0.012			
AADT per lane [major]	4.45E-05	1.17E-05	0.000	2.02E-05	9.70E-06	0.038
Truck percentage [major]	-0.019	0.037	0.610	-0.049	0.029	0.084
Arterial (ind.) [major]	0.229	0.141	0.105	-0.167	0.085	0.049
On system roadway (ind.)	-0.231	0.131	0.077	-0.010	0.108	0.923
Distance to nearest hospital (mi)	0.089	0.046	0.050	0.035	0.031	0.252
Transit stops (ind.)	0.378	0.116	0.001	0.647	0.091	0.000
Number of stops	0.028	0.016	0.070	0.013	0.012	0.275
Population density (sq mi)	2.11E-05	7.41E-06	0.005	5.30E-05	6.30E-06	0.000
Employment density (sq mi)	-1.59E-06	1.61E-06	0.324	3.33E-05	6.82E-06	0.000
Median income (\$10k)	-0.099	0.015	0.000	-0.119	0.011	0.000
CBD (ind.)	0.738	0.182	0.000	1.453	0.157	0.000
<b>No. of observations</b>	<b>19,194</b>			<b>41,107</b>		
<b>Dispersion Parameter (<math>\rho</math>):</b>	0.821			0.430		
<b>McFadden's R2:</b>	0.616			0.370		
<b>Likelihood ratio test (<math>\chi^2</math>)</b>	2,808			2,718		
<b>Prob &gt; <math>\chi^2</math></b>	0.000			0.000		
<b>2 x log-likelihood</b>	-5,618			-10,567		



**FIGURE 5 Sensitivity analysis, City of Austin.**

In terms of the traffic attributes, AADT is positive and significant for the intersection model (same as Texas). Still, the percentage of trucks is negative in both models (as opposed to the positive effect for intersections in Texas). This suggests that pedestrians in Austin are less likely to be involved in an accident in intersections with a high number of trucks compared to the rest of the state. The functional class indication for arterial roadways is not significant, and no conclusion can be obtained from this variable. In comparison to the Texas model, the City of Austin model also suggests that on-system roads have fewer crashes than other roadways. The distance to hospitals in the Austin model is not relevant, likely due to the small area selected for the model, where multiple hospitals are located across the city. The exposure variable for transit presence is

significant (midblock) and suggests a positive correlation with the number of pedestrian crashes. However, the variables related to transit are not as sensitive in this model compared to the Texas model.

Land use variables such as population density, employment density, and average household income were approximated using the CAMPO data at the TAZ level to analyze the effects on the city's pedestrian crash rate. Population density has a positive and significant effect, as expected. One standard deviation increase (3.3 thousand individuals per square mile) led to a 15% (intersections) and 21% (midblock) increase in crash rates. The employment density, in contrast, does not have an effect on the intersections model. A significant finding suggests that areas with a higher average household income tend to present fewer pedestrian crashes. An increase of \$41,000 in the average household income led to a reduction of 32% (intersections) and 39% (midblock) in pedestrian crash rates. Finally, an indicator variable of the CBD in the city highlights the importance of this area, with midblock crashes being more sensitive (240%) in this area than intersection (78%) crashes, but both models showing a relevant effect.

## SUMMARY AND CONCLUSIONS

In this research, historical pedestrian crash information from police reports in Texas is used to understand the factors associated with crash rates at intersection and midblock levels. Developing micro-level analysis is challenging due to the lack of geographic information and characterization at a statewide scale. Therefore, one of the main contributions of this study is the development of a methodology to spatially model crash locations with respect to known roadways, intersections, and traffic signals. Geometry estimations at intersection and midblock levels are obtained, and information from the roadway infrastructure inventory and other sources is assigned with the objective of characterizing such geometries. Information such as traffic control (signalized intersections), highway design variables, traffic attributes, and land use from multiple sources is combined along with the location of the crashes (separated between intersections and midblock crashes) to provide a comprehensive analysis of the roadway network in the State of Texas.

An NB model is used to identify major factors influencing pedestrian crashes at intersection and midblock levels. Models for the State of Texas are developed along with a case study of the City of Austin, one of the areas with the highest number of crashes in the state, to understand specific factors affecting the city's crash rate. The main results suggest that signalized intersections present a higher pedestrian crash rate, DVMT increases the likelihood of pedestrian crashes, and midblock segments are more vulnerable, where one standard deviation increase causes an increase in crashes at intersections and midblock sections of 52% and 187%, respectively. Variables such as the number of lanes and lane width contribute to higher pedestrian crash rates, while higher posted speed limits and longer median widths tend to coincide with a reduced crash rate. Arterial roads are prone to have a higher number of pedestrian crashes than local and collector roads. Land use variables indicate that areas with a greater population tend to have more crashes. The analysis of the Austin area suggests that the CBD is critical for both models, with midblock crashes being more sensitive (240%) in this area than intersection (78%) crashes. Also, a significant inequity was also found in the area: an increase of \$41,000 in average household income leads to a reduction of 32% (intersections) and 39% (midblock) in pedestrian crash rates.

The pedestrian crash data used in this study, obtained from the TxDOT CRIS database, only includes records from police officers, which limits this study to the analysis of only the recorded crashes. Other limitations of this study include the approximate map-matching process used to obtain the geometry characteristics for both intersections and midblock segments, based on the TxDOT roadway inventory. However, the developed analysis contains detailed intersection and midblock segment information not used in prior research; to the best of the authors' knowledge, the most complete set of intersections currently available at the state level is utilized for this type of analysis. Researchers provide the intersection and uniform roadway segment sets online at <https://github.com/ut-ctr-nmc/peds-midblocks-intersections>.

## AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: writing-original draft preparation: N. Zuniga-Garcia, K. Perrine; conceptualization and design: K. Kockelman, N. Zuniga-Garcia; methodology: K. Kockelman, N. Zuniga-Garcia; data assembly and analysis: Zuniga-Garcia, K. Perrine; writing-reviewing and editing: K. Kockelman, N. Zuniga-Garcia, K. Perrine. All authors have reviewed the results and approved the final version of the manuscript.

## ACKNOWLEDGEMENTS

Funding for this research comes from TxDOT Research and Technology Innovation (RTI) Project 0-7048.

## REFERENCES

- Bernhardt, M., Kockelman, K., 2021. An analysis of pedestrian crash trends and contributing factors in Texas. *Journal of Transport & Health* 22, 101090. <https://doi.org/10.1016/j.jth.2021.101090>
- BTS, U.States.D.O.Transportation.B.O.T.S., 2019. National Transportation Statistics: U.S. Passenger-Miles [WWW Document]. URL [https://rosap.nhtl.bts.gov/gsearch?collection=dot:35533&type1=mods.title&fedora\\_terms1=National+Transportation+Statistics](https://rosap.nhtl.bts.gov/gsearch?collection=dot:35533&type1=mods.title&fedora_terms1=National+Transportation+Statistics) (accessed 5.16.21).
- Dean, M. D., & Zuniga-Garcia, N. (2022). Shared E-Scooter Trajectory Analysis During the COVID-19 Pandemic in Austin, Texas. *Transportation Research Record*. <https://doi.org/10.1177/03611981221083306>
- Diogenes, M.C., Lindau, L.A., 2010. Evaluation of Pedestrian Safety at Midblock Crossings, Porto Alegre, Brazil. *Transportation Research Record* 2193, 37–43. <https://doi.org/10.3141/2193-05>
- Ester, M., Kriegel, H.-P., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise 96, 226--231.
- GHSA, G.H.S.A., 2020. Pedestrian Traffic Deaths by State, Spotlight in Highway Safety.
- Kwayu, K.M., Kwigizile, V., Oh, J.-S., 2019. Evaluation of pedestrian crossing-related crashes at undesignated midblock locations using structured crash data and report narratives. *Journal of Transportation Safety & Security* 0, 1–23. <https://doi.org/10.1080/19439962.2019.1679306>
- Lee, J., Abdel-Aty, M., Shah, I., 2019. Evaluation of surrogate measures for pedestrian trips at intersections and crash modeling. *Accident Analysis & Prevention, Road Safety Data Considerations* 130, 91–98. <https://doi.org/10.1016/j.aap.2018.05.015>

- 1 Lightstone, A.S., Dhillon, P.K., Peek-Asa, C., Kraus, J.F., 2001. A geographic analysis of motor  
2 vehicle collisions with child pedestrians in Long Beach, California: comparing intersection  
3 and midblock incident locations. *Injury Prevention* 7, 155–160.  
4 <https://doi.org/10.1136/ip.7.2.155>
- 5 McAndrews, C., Pollack, K.M., Berrigan, D., Dannenberg, A.L., Christopher, E.J., 2017.  
6 Understanding and Improving Arterial Roads to Support Public Health and Transportation  
7 Goals. *Am J Public Health* 107, 1278–1282. <https://doi.org/10.2105/AJPH.2017.303898>
- 8 Noland, R.B., Klein, N.J., Tulach, N.K., 2013. Do lower income areas have more pedestrian  
9 casualties? *Accident Analysis & Prevention* 59, 337–345.  
10 <https://doi.org/10.1016/j.aap.2013.06.009>
- 11 OpenStreetMap contributors, 2021a. Planet dump retrieved from <https://planet.osm.org> [WWW  
12 Document]. URL <https://www.openstreetmap.org>
- 13 OpenStreetMap contributors, 2021b. Overpass API [WWW Document]. URL  
14 [https://wiki.openstreetmap.org/wiki/Overpass\\_API](https://wiki.openstreetmap.org/wiki/Overpass_API)
- 15 Perrine, K., Khani, A., Ruiz-Juri, N., 2015. Map-Matching Algorithm for Applications in  
16 Multimodal Transportation Network Modeling. *Transportation Research Record* 2537, 62–  
17 70. <https://doi.org/10.3141/2537-07>
- 18 PostGIS, 2021. PostGIS [WWW Document]. URL <https://postgis.net/>
- 19 Pulugurtha, S.S., Sambhara, V.R., 2011. Pedestrian crash estimation models for signalized  
20 intersections. *Accident Analysis & Prevention* 43, 439–446.  
21 <https://doi.org/10.1016/j.aap.2010.09.014>
- 22 Rahman, M., Kockelman, K.M., Perrine, K.A., 2021. Predicting Pedestrian Crash Occurrence and  
23 Injury Severity In Texas. Presented at the 100th Annual Meeting of the Transportation  
24 Research Board. Under review for publication in *Traffic Injury and Prevention*.
- 25 Sandt, L., Zegeer, C.V., 2006. Characteristics Related to Midblock Pedestrian–Vehicle Crashes  
26 and Potential Treatments 1, 113–121.
- 27 Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and  
28 bicycle crashes. *Accident Analysis & Prevention* 45, 382–391.  
29 <https://doi.org/10.1016/j.aap.2011.08.003>
- 30 Texas Department of Transportation, 2020. Crash Data Analysis and Statistics [WWW  
31 Document]. URL <https://www.txdot.gov/government/enforcement/crash-statistics.html>
- 32 Texas Department of Transportation, 2018. Texas Roadway Inventory [WWW Document]. URL  
33 [https://www.txdot.gov/inside-txdot/division/transportation-planning/roadway-](https://www.txdot.gov/inside-txdot/division/transportation-planning/roadway-inventory.html)  
34 [inventory.html](https://www.txdot.gov/inside-txdot/division/transportation-planning/roadway-inventory.html)
- 35 Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built  
36 environment on pedestrian crash frequency. *Safety Science*, First International Symposium  
37 on Mine Safety Science and Engineering 2011 50, 1141–1151.  
38 <https://doi.org/10.1016/j.ssci.2011.09.012>
- 39 Wang, Y., Kockelman, K.M., 2013. A Poisson-lognormal conditional-autoregressive model for  
40 multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident*  
41 *Analysis & Prevention* 60, 71–84. <https://doi.org/10.1016/j.aap.2013.07.030>
- 42 Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of  
43 vehicle-pedestrian injury collisions with implications for land use and transportation  
44 planning. *Accident Analysis & Prevention* 41, 137–145.  
45 <https://doi.org/10.1016/j.aap.2008.10.001>

- 1 Xie, S.Q., Dong, N., Wong, S.C., Huang, H., Xu, P., 2018. Bayesian approach to model pedestrian  
2 crashes at signalized intersections with measurement errors in exposure. *Accident Analysis*  
3 & *Prevention* 121, 285–294. <https://doi.org/10.1016/j.aap.2018.09.030>
- 4 Zhao, B., Zuniga-Garcia, N., Xing, L., Kockelman, K., 2021. Predicting Pedestrian Crash  
5 Occurrence and Injury Severity in Texas Using Tree-Based Machine Learning Models.  
6 Under review for publication in *Transportation Research Record*.  
7