Link-Level Crash Models For Different Road Users: A Practical Tool For Large-Scale Safety Assessment

Qin Zhang
Department of Civil, Geo and Environmental Engineering
Technical University of Munich, Germany
Email: qin.zhang@tum.de

Rolf Moeckel Ph.D.
Department of Civil, Geo and Environmental Engineering
Technical University of Munich, Germany
Email: rolf.moeckel@tum.de

Carlos Llorca Ph.D.
Department of Civil, Geo and Environmental Engineering
Technical University of Munich, Germany
Email: carlos.llorca@tum.de

Mehmet Baran Ulak Ph.D.
Department of Civil Engineering
University of Twente, the Netherlands
Email: m.b.ulak@utwente.nl

James Woodcock Ph.D.
Centre for Diet and Activity Research
University of Cambridge, United Kingdom
Email: jw745@medschl.cam.ac.uk

Word Count: 4928 words + 7 table (250 words per table) = 6678 words

Submitted: 01.08.2020

ABSTRACT

Road safety is one critical transport effect on health. To assess the safety performance of a city or a region, large scale crash estimation models are essential to be integrated into transport models. Most research on crash estimation models focused on improving model predication performance by adding more omitted variables, using random parameters, or more advanced statistical techniques. However, those models are not easy to implement in practice due to the model complexity. This research aims to develop practical crash estimation models within an entire region. This research employed the road accident data from the German Federal Statistic Office to estimate the number of crashes on each link for different road users including car occupants, cyclists, and pedestrians. Results show that magnitudes of the effects of traffic volumes varied from car occupancy model to cyclist and pedestrian model. The effect of motor traffic volume is notable on the number of car occupancy crashes, while bike flow and pedestrian flow have more positive influences on cyclist and pedestrian crashes than motor traffic volume. The models are practical in aiding transportation safety planning. In addition, the models can be applied in practice within existing agent-based transport models to assess the crash exposure rate of each road user.

Keywords: crash estimation model; cyclist crash; pedestrian crash; car occupant crash

INTRODUCTION

Safety of different road users has been closely linked to the public health level. According to the German Federal Statistic Office, 206,129 car occupants were injured in 2019 which accounts for 54% of all injured people in traffic crashes (*1*). Furthermore, cyclists and pedestrians are usually considered as more vulnerable road users compared to car occupants since they are exposed to a higher risk of injury and fatality in urban road accidents (*1*). It is also a fact that traveling on a roadway with poor infrastructure or on a busy network could be detrimental to individual's health. Therefore, crash estimation models are of critical importance to assess the safety performance of roadways which can help inform governments and agencies of how to allocate infrastructure investments and devise policies to maintain safer roadways.

The motivation of this research is to assess transport influences on health aspects in an agent-based simulation environment. Agent-based simulation attempts to more accurately represent individual trips thereby giving a better assessment of health. Road safety is one critical transport effect on health. The goal of this research is to implement practical crash estimation models in agent-based transport simulation tool at a region-wide study area.

For many decades, various crash estimation models and statistical techniques have been developed to better understand the factors that affect the number of crashes, injuries, and fatalities. Conventional Poisson and negative binomial models are the starting points. More advanced models, like random parameter models and random effect models, are widely used to better account for unobserved heterogeneity and to handle temporal and spatial correlations (*2*). These models made a major contribution to transportation safety research. However, Lord and Mannering (*2*) pointed out the trade-off between model prediction performance and practicality. They argued that these models may not be easily transferable to other datasets and they are not easy to implement in transportation safety planning due to the model complexity, which are substantial disadvantages. A common practice in transportation safety planning is to predict crash frequencies and severity using only traffic volumes and link length as explanatory variables (*3*). Nonetheless, researchers argue that such simple models may result in inaccurate predictions. (*2*, *4*)

Given this context, we attempt to balance the tradeoff between model prediction performance and model practicality for future implementation. Considering the practical aspect, we examine this using a conventional zero-inflated Poisson model (ZIP), which is easy to implement in the agent-based simulation. Given the model prediction performance, we test more variables that are usually omitted in the practice of transportation safety planning. We selected explanatory variables that can be easily obtained and/or readily forecasted.

Another highlight of this study is that the practical crash estimation models can be employed at a city-wide or region-wide area for large-scale safety assessment. From the perspective of analysis entity, road safety research can be divided into two main lines: zonal level analysis and link/segment level analysis. Zonal level studies are conducted at different spatial units (transportation analysis zone, block group, census track) (*5*–*9*). With the benefit of agent-based simulation, the traveling routes of individual trips can be specified. By using zonal level crash analysis may diminish the benefit of agent-based simulation. Therefore, such studies are not aligned with the objective of this research. In terms of link/segment level analysis, the majority of the literature focuses on the investigation of the number of crashes on a specific type of road, such as highway segments and arterial roads (*10*, *11*). However, there is scant literature on the analysis of different types of crashes that occurred on residential roads and other minor roads. In this research, all kinds of road links from the motorway to minor roads like pedestrian path and living streets are considered in the estimation model.

This research proposes new link-level crash estimation models for different road users including car occupants, cyclists, and pedestrians. We categorize the road accidents into four different cases 1) car-occupancy-as victim crashes 2) cyclist-as-victim crashes in the bike-motor accidents 3) cyclist-as-victim crashes in the bike-bike accidents 4) pedestrian-as-victim crashes. The number of severe-fatal crashes and the number of light injury crashes are estimated separately for all cases.

The organization of the paper is as follows. Section 2 reviews research related to factors associated with crash frequency and severity, after which section 3 presents the data sources and methodology. Model estimation results are discussed in Section 4. Finally, a summary of the main findings and conclusions are provided in Section 5.

LITERATURE REVIEW

Previous studies have identified many factors that influence different kinds of accidents. While the magnitudes of the effects vary across studies (*5–11*), a common set of factors that affect the number of injury and fatality has been identified in the literature.

The association between land use characteristics and the number of crashes are widely explored in the zonal level safety studies. Amoh-Gyimah et al. (*6*) comprehensively compared different statistical models to estimate pedestrian and cyclist crashes at the aggregated zone level. They found that population and land use diversity have significant and positive correlations with pedestrian and bicycle crash frequency. For instance, a 1% increase in land use mix index will have on total 11.67 increase in pedestrian crashes.

In terms of traffic-related factors, annual average daily traffic (AADT-an indicator of the volume of motorists in the traffic) was found to be consistently significant in the previous studies (*12*, *13*). In addition to the motorist volumes, Aldred et al. (*5*) find that cyclist flow is a significant predictor of bicycle crashes. Their model shows a decrease in cycle injury odds with an increase in cycling flows significantly.

Roadway-related factors also have significant influences on the frequencies of different types of crashes. Kim et al. (*8*) estimated motor vehicle crashes as a function of demographic variables, land use factors, and road characteristics. The key finding is that number of intersections is associated with increases in all kinds of accidents including motor vehicles, cyclists, and pedestrians.

Some studies also pointed out the associations between injury severity and environmental factors. For instance, Pahukula et al. (*13*) explored the influence of time of day on injury severity in truck-involved accidents. Employing the crash data in Texas from 2006 to 2010, they developed random parameter logit models for five-time of period. They found that factors like light condition, traffic volume, and surface condition perform differently in five-time of day models. Also, it is interesting that their results also revealed that clear weather conditions increase the probability of severe injury. On the other hand, it is proved that injury severity is determined by the level of vehicle speed (7, 14). For example, Chen et al. (14) analyzed cyclist injury severity in automobile-involved bicycle crashes and found that speed limit of roadway is positively correlated with the likelihood of severe injury and fatality in cyclist crashes.

DATA AND METHODS

Road accidents data

The study area is the Munich Metropolitan area. Road accident data in this area are obtained from the German Federal Statistic Office based on reports from the police (*1*). Crash information ranging from 2016 to 2018 is used in this study. The severity of accidents is classified into three levels: fatal, severe injury, and light injury. Accidents are classified into these segments based on the most severe injury that occurred in an accident. Fatality is defined as persons who died within 30 days as a result of the accident. A severe injury requires hospital treatment for at least 24 hours, while light level includes all other injuries. Accident locations are recorded at the coordinate level. This study considers all kinds of network links including minor roads like residential roads and living streets. Crash points are assigned (based on geolocation) to the road network obtained from OpenStreetMap (OSM) to aggregate the crash frequency at each network link.

A total of 55,848 accidents were identified in the study area. The data provided the information on which road entities were involved in an accident including pedestrian, cyclists, passenger cars, trucks, public transport vehicles. Unfortunately, the data did not have information to identify which road user was insured. In this study, we assigned the injury or fatality to the weakest road user involved in an accident. We assumed that pedestrians are the most vulnerable road users, while cyclists are relatively more vulnerable than car occupants. With these assumptions, we defined the victim for each accident. To assess the crash risk exposure of different road users, accidents are classified into four cases based on the assumed victim.

- Case 1: car-occupant-as-victim crashes are defined as any crash involving passenger car(s) and/or other types of vehicles (e.g. trucks, bus, tram)
- Case 2: cyclist-as-victim crashes are defined as any crash involving both bike(s) and motor vehicle(s)
- Case 3: cyclist-as-victim crashes are defined as any crash involving only bikes
- Case 4: pedestrian-as-victim crashes are defined as any crash involving pedestrian(s)

In this study, severe injury accidents and fatal accidents are merged into one category. The main reason is that fatal accidents have a small sample size in this dataset. Small sample size may introduce bias into the estimation process so that the results of model estimation will not be accurate and reliable. 3,428 accidents are filtered because they only involve truck/bus/tram, which is out of the scope of this research. Finally, there are 52,420 accidents considered in model estimation. TABLE 1 summarizes the number of severe-fatal and light injury crashes of four cases.

TABLE 1 Number of severe-fatal and light injury crashes of four cases from 2016 to 2018

Case	Truck/PT	Car	Bike	Pedestrian	Victim	Total	Severe-fatal	Light injury
Case 1	√	√			Car-occupant	30516	4262	26254
Case 2	√	√	√		Cyclist	10773	1448	9325
Case 3			√		Cyclist	6486	1352	5134
Case 4	√	√	√	√	Pedestrian	4645	1050	3595

Explanatory variables

Previous studies on road safety factors give us an indication of variable selection. Road safety factors for macro-level analysis are mainly divided into three streams: traffic-related factors, roadway-related factors, and land use factors. Explanatory variables considered in this study are three traffic-related variables: motor traffic volumes, bike flows, and pedestrian flows; and four roadway-related variables: link length, speed limit, road classification, and number of intersections.

An agent-based travel demand model suite created by Moeckel et al. (*15*) is used to derive estimates of motor traffic volumes, bike flows, and pedestrian flows. MITO (microscopic transportation orchestrator, https://github.com/msmobility/mito) generates individual trips for household members of a synthetic population and then pass on to transportation simulation tool MATSim (*16*) for trip assignment. This model suite provides us traffic flows of different modes across the Munich Metropolitan area on a typical weekday. With the benefits of microscopic simulation, traffic flow information is available on both major and minor roads. The agent-based travel demand model also provides information on trip purposes, so the share of commuter trips on each link can be measured. It was initially considered in estimation models but it is removed in the final models due to the consistently insignificance.

In the study area, there are a total of 133,546 links in the car network and a total of 92,748 links in the pedestrian and bike network which excludes motorway and other car-only links. Link length and speed limit data are directly obtained from OSM. The number of intersections on each link is measured in MATSim by counting the number of nodes with more than two legs along the link segment. The network links in OSM have more than twenty classifications of road types. Those links with similar characteristics are grouped into six aggregated road classifications. For example, "trunk" links are merged with "primary" links since they have similar characteristics such as maximum speed and capacity. All minor roads such as "pedestrian", "track" and "living street" are merged into one category because they normally have a lower speed limit and grant pedestrian the right of way over. The six aggregated link types used in the estimation models are motorway, primary road, secondary road, tertiary road, residential road, and other minor roads.

TABLE 2 gives an overview of descriptive statistics of the dependent and explanatory variables. One notable characteristic of crash frequency data is that there is a very large number of zero-crash observations. Especially in severe-fatal crash cases, approximately 99% of all network links have zero severe-fatal crash.

TABLE 2 Descriptive statistics of the set of dependent and explanatory variables

	Total	Mean	Sd.	Min	Max	Zero proportion
Case 1: Car-occupancy-as-victim crashes						
Severe-fatal crash frequency	4262	0.03	0.20	0	11	97.8%
Light crash frequency	26254	0.16	0.70	0	53	90.6%
Case 2: Cyclist-as-victim crashes in bike-motor accidents						
Severe-fatal crash frequency	1448	0.01	0.13	0	5	98.6%
Light crash frequency	9325	0.10	0.42	0	14	93.1%
Case 3: Cyclist-as-victim crashes in bike-bike accidents						
Severe-fatal crash frequency	1352	0.01	0.12	0	4	98.7%
Light crash frequency	5134	0.05	0.27	0	11	95.7%
Case 4: Pedestrian-as-victim crashes						
Severe-fatal crash frequency	1050	0.01	0.11	0	5	99.0%
Light crash frequency	3595	0.04	0.24	0	20	96.9%
Explanatory Variables						
Motor traffic volumes (x1,000)		1.76	3.81	0	54	2.7%
Bike flows (x1,000)		0.47	1.22	0	24	~
Pedestrian flows (x1,000)		0.62	1.24	0	20	~
Speed limit (km/h)		27.50	23.70	3	109	~
Link length (meter)		540.48	837.35	50	17530	~
Number of intersections		2.58	3.16	0	42	18.8%

Zero-inflated Poisson Regression Model

The eight dependent variables in this study are all count variables with non-negative integer values. Two popular non-linear count models that are widely used for estimating this kind of dependent variable: Poisson regression model (P) and negative binomial regression model (NB). The Poisson model assumes that the sample variance is equal to its expected value; however, the crash frequency data is known to have an over-dispersion issue, i.e., the variance of the number of crashes is larger than the mean number of crashes. Nevertheless, the over-dispersion issue was not identified in the crash datasets utilized in this study. Moreover, some researchers argue that the performance of negative binomial models can be negatively influenced by the low sample mean (*2*). As shown in TABLE 2, eight crash frequency variables all have more zero observations than non-zero ones, which indicates that it is necessary to explore whether zero-inflated Poisson model (ZIP) and zero-inflated negative binomial model (ZINB) are more suitable than the ordinary ones.

Therefore, the performance of Poisson, NB, ZIP, ZINB are compared using three measures: Akaike information criterion (AIC), Efron's pseudo $R^{2,}$ and root mean square errors (RMSE). AIC is a common score for comparing the goodness-of-fit of different models using the same dataset. Efron's pseudo R^2 and RMSE are used for evaluating the prediction quality of the models to the observed data. TABLE 3 presents an example of model comparison results for bike crash models. In general, all three measures indicate that zero-inflated models have a better performance than the normal ones. The differences of R^2 and RMSE between ZIP and ZINB are minor. AIC indicates that ZIP fits better for severe-fatal models, while ZINB is more suitable for light injury models. In this study, ZINB does not help much in model prediction accuracy. To keep consistency and take account for the practical side of this study, ZIP model was adopted to examine the number of severe-fatal crashes and light injury crashes in all cases.

The ZIP model (*17*) is also called the two-steps model, as it is essentially a combination of a logistic model and a standard Poisson model. In the first step, a logistic regression model is applied to distinguish zero and non-zero state of each link. If the link has a high probability of generating a crash, then the second step estimates the number of crashes on that link.

TABLE 3 Goodness-of-fit comparison among four different bike crash models

	Severe-fatal				Light injury			
	P	NB	ZIP	ZINB	P	NB	ZIP	ZINB
AIC	12803	12753	*12745*	12756	50698	48447	48647	*48182*
Efron's Pseudo R2	0.024	0.024	0.024	0.024	0.143	0.133	0.147	*0.145*
RMSE	0.127	0.127	0.127	0.127	0.397	0.400	0.397	0.397

RESULTS AND DISCUSSION

To avoid the multicollinearity in the regression model, the Pearson correlation coefficients are calculated for all the explanatory variables as shown in TABLE 4. The speed limit is highly correlated to road classification. Separate individual models were developed for both variables. Road classification is found to be significant and highly correlated to the number of crashes, so it is retained for the final models. It is expected that bike flow and pedestrian flow are correlated to a high degree with a score of 0.92 because the travel demand model simply assigned both cyclists and pedestrians to the shortest path. We were lacking data to describe the differences in bike and pedestrian route choice behavior. After testing bike flow, pedestrian flow, and the sum-up flows in separate individual models, bike flow was retained in the models of Case 2 and 3, while the pedestrian flow was kept in the models of Case 4. Note that, non-log and log-form of traffic flows were tested in separate models and either of them retained in the final models based on the significance and correlation score. After testing, link lengths are transferred to log-form since the log-form results in a better model fit in all models.

In this study, two measurements were applied to evaluate model performance. McFadden's pseudo R^2 is adopted to measure the performance of the final fitted model compared to the intercept only (null) model. Efron's pseudo R^2 is a normalized square error between the predicted value and actual value, which is comparable across models with different datasets. By analyzing the model performance and significance, some of the variables were excluded and specifications of the final models are presented in TABLE 5, 6 and 7.

TABLE 4 Pearson coefficient of correlation for all the explanatory variables

	Motor volume	Bike flow	Pedestrian flow	Log (link length)	Link type	Speed limit	Intersections
Motor volume	1.00						
Bike flow	0.65	1.00					
Pedestrian flow	0.59	*0.92*	1.00				
Log (link length)	-0.26	-0.20	-0.19	1.00			
Link type	-0.54	-0.29	-0.25	0.21	1.00		
Speed limit	0.49	0.26	0.23	-0.18	*0.95*	1.00	
Intersections	-0.16	-0.06	-0.01	0.46	0.11	-0.13	1.00

Car-occupant-as-Victim crash frequency models

The influences of explanatory variables on the car-occupant-as-victim crash are shown in TABLE 5. It was found that significant differences are present between the predictor sets for light injury crashes and severe-fatal crashes. As expected, motor traffic volumes and link length, as two common factors in safety research, are significant in both models. However, number of intersections only shows influences on number of light injury crashes. In the light injury model, the coefficients of link types show consistently decreasing by road hierarchy levels. This indicates that passenger car crashes are prone to happen on motorways and major roads. The interaction coefficients between traffic volumes and link type give us hints that a busy residential road has a higher risk of light injury crashes than a congested highway. This might because of the low capacity and relatively poor road infrastructure on minor roads. It is notable that the coefficient of traffic volume in the light injury model is much smaller than that in the severe-fatal model. The result may be due to the effects of interaction parts in the light injury crashes, which decomponent the influence of traffic volume to link types. In general, we found less model sensitivity in the severe-fatal crashes. Moreover, the large intercept (-6.81) in the severe-fatal model also indicates that there is more unobserved heterogeneity in rare crash types (fatality and severe injury).

TABLE 5 Estimation results of ZIP models for car-occupant-as-victim (Case 1) crash models

	Light injury crashes			Severe-fatal crashes		
	Estimate	Pr(>|z|)		Estimate	Pr(>|z|)	
Step 1: Zero-inflation model coefficients (binomial with logit link)						
Intercept	-2.68	0.00	***	-1.87	0.00	***
Traffic conditions						
Motor traffic volumes in 1,000	0.04	0.00	***	1.72	0.00	***
x motorway	base	base	base	base	base	base
x primary road	0.10	0.00	***			
x secondary road	0.09	0.00	***			
x tertiary road	0.88	0.00	***			
x residential or minor road	4.13	0.00	***			
Roadway function						
is motorway	base	base	base	base	base	base
is primary road	-0.30	0.00	***	-0.33	0.07	.
is secondary road	-0.13	0.19				
is tertiary road	-1.24	0.00	***			
is residential or minor road	-2.33	0.00	***	-1.61	0.00	***
Roadway geometry						
log(link length)	0.36	0.00	***	0.14	0.00	***
Number of intersections	0.06	0.00	***			
Step 2: Count model coefficients (Poisson with log link)						
Intercept	-1.58	0.00	***	-6.81	0.00	***
Traffic conditions						
Motor traffic volumes in 1,000	0.01	0.00	***	0.01	0.00	***
x motorway	base	base	base	base	base	base
x primary road						

x secondary road	0.04	0.00	***			
x tertiary road	0.11	0.00	***			
x residential or minor road	0.18	0.00	***			
Roadway function						
is motorway	base	base	base	base	base	base
is primary road	-0.30	0.00	***	-0.14	0.01	**
is secondary road	-0.68	0.00	***			
is tertiary road	-1.28	0.00	***	-0.78	0.00	***
is residential road	-2.09	0.00	***	-0.93	0.00	***
Roadway geometry						
log(link length)	0.35	0.00	***	0.81	0.00	***
Number of intersections	0.03	0.00	***			
Efron's pseudo R2	*25.6%*			*17.6%*		
McFadden's pseudo R2	*23.9%*			*22.7%*		

Note:
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
Not significant variables were not estimated in the final model

Cyclist-as-Victim crash frequency models

It is assumed that cyclist crashes have different characteristics than the car occupant crashes. Results of light injury crash models (TABLE 6) show that bike flows have a larger influence in bike-motor accidents (Case 2) on both number of crashes and non-zero crash state switching, that is the likelihood of switching from a zero-crash link to a non-zero crash link. Especially in the zero-part model, the coefficient of bike flows is about 23 times larger than that of motor traffic volumes. On the one hand, this large difference results from the inequivalent scales between bike and motor traffic volumes. On average of the study area, daily motor traffic volume of one link is approximately 5 times higher than the daily bike traffic volume. However, bike flows still weigh more than motor traffic volumes after diminishing the scale issue. In Case 3, motor traffic volumes have no significant impact on non-zero state switching. This is reasonable that the accidents in case 3 only involve cyclists. Secondary bike accidents resulting from car accidents seem to be rare in this dataset.

The coefficients of link types are not intuitive at first sight. They give us inconsistent indications in Case 2 and Case 3. In bike-motor accidents, residential roads and minor roads were found to be less risky for cyclists, while major roads including primary, secondary, and tertiary roads are more vulnerable. In bike-bike accidents, tertiary roads have less probability of having a cyclist crash, but it has a higher potential to generate a larger number of crashes on one link. The low probability of having a crash might be due to the low bike flows on the tertiary road. However, some specific tertiary roads tend to be the hot spots of cyclist crashes due to poor road conditions. There might be because of lacking cycle infrastructures but this needs to be explored in further research.

Different from the car occupant models, number of intersections is notable in both Case 2 and Case 3, though the effects are marginal. The influence of number of intersections in bike-motor accidents is slightly higher than that in bike-bike accidents. It is expected that bike-vehicle conflicts tend to be happened at intersections due to the mixed space.

TABLE 7 presents the severe-fatal crashes in cyclist accidents. The model results are quite similar to light injury models. Bike flow is the dominant factor in number of severe-fatal crashes. Differently, tertiary roads do not show negative impacts on generating crashes. Number of intersections is not significantly correlated to the number of severe-fatal crashes in bike-bike accidents. Overall, the severe-fatal models have a poorer performance than the light injury models in terms of R^2. This can be explained that 1) fewer severe and fatal accidents recorded in our dataset, 2) literature points out that severe accidents are highly associated with vehicle factors like driving speed and human factors. For example, aging cyclists appear to be more vulnerable to severe injury or fatality (*14*). However, this information is not available in our dataset, 3) such rare crashes events appear to be more random due to the (fortunately) very small sample size.

TABLE 6 Estimation results of ZIP models for light injury crash models in case 2,3 and 4

<table>
<tr><th></th><th colspan="3">Case 2</th><th colspan="3">Case 3</th><th colspan="3">Case 4</th></tr>
<tr><th></th><th>θ</th><th>Pr(>|z|)</th><th></th><th>θ</th><th>Pr(>|z|)</th><th></th><th>θ</th><th>Pr(>|z|)</th><th></th></tr>
<tr><td colspan="10">Step 1: Zero-inflation model coefficients (binomial with logit link)</td></tr>
<tr><td>Intercept</td><td>-4.14</td><td>0.00</td><td>***</td><td>-2.29</td><td>0.00</td><td>***</td><td>-3.50</td><td>0.00</td><td>***</td></tr>
<tr><td>Motor traffic volumes in 1,000</td><td>0.18</td><td>0.00</td><td>***</td><td></td><td></td><td></td><td>0.05</td><td>0.02</td><td>*</td></tr>
<tr><td>Bike flows in 1,000</td><td>4.91</td><td>0.00</td><td>***</td><td>4.74</td><td>0.00</td><td>***</td><td></td><td></td><td></td></tr>
<tr><td>Pedestrian flows in 1,000</td><td></td><td></td><td></td><td></td><td></td><td></td><td>2.68</td><td>0.00</td><td>***</td></tr>
<tr><td>is primary road</td><td>1.03</td><td>0.00</td><td>***</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>is secondary road</td><td>0.90</td><td>0.00</td><td>***</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>is tertiary road</td><td>1.55</td><td>0.00</td><td>***</td><td>-0.46</td><td>0.00</td><td>***</td><td>0.51</td><td>0.00</td><td>***</td></tr>
<tr><td>is residential or minor road</td><td></td><td></td><td></td><td>0.19</td><td>0.02</td><td>*</td><td></td><td></td><td></td></tr>
<tr><td colspan="10">Step 2: Count model coefficients (Poisson with log link)</td></tr>
<tr><td>Intercept</td><td>-3.24</td><td>0.00</td><td>***</td><td>-4.88</td><td>0.00</td><td>***</td><td>-4.37</td><td>0.00</td><td>***</td></tr>
<tr><td>Motor traffic volumes in 1,000</td><td>0.01</td><td>0.00</td><td>***</td><td>0.01</td><td>0.00</td><td>***</td><td>0.01</td><td>0.00</td><td>***</td></tr>
<tr><td>Bike flows in 1,000</td><td>0.08</td><td>0.00</td><td>***</td><td>0.10</td><td>0.00</td><td>***</td><td></td><td></td><td></td></tr>
<tr><td>Pedestrian flows in 1,000</td><td></td><td></td><td></td><td></td><td></td><td></td><td>0.10</td><td>0.00</td><td>***</td></tr>
<tr><td>is primary road</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>is secondary road</td><td rowspan="2">-0.12</td><td rowspan="2">0.02</td><td rowspan="2">*</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>is tertiary road</td><td>0.32</td><td>0.00</td><td>***</td><td></td><td></td><td></td></tr>
<tr><td>is residential or minor road</td><td>-0.11</td><td>0.08</td><td>.</td><td></td><td></td><td></td><td>-0.24</td><td>0.00</td><td>***</td></tr>
<tr><td>log(link length)</td><td>0.38</td><td>0.00</td><td>***</td><td>0.51</td><td>0.00</td><td>***</td><td>0.46</td><td>0.00</td><td>***</td></tr>
<tr><td>Number of intersections</td><td>0.03</td><td>0.00</td><td>***</td><td>0.01</td><td>0.00</td><td>***</td><td></td><td></td><td></td></tr>
<tr><td>Efron's pseudo R2</td><td>13.9%</td><td></td><td></td><td>7.59%</td><td></td><td></td><td>10.0%</td><td></td><td></td></tr>
<tr><td>McFadden pseudo R2</td><td>13.1%</td><td></td><td></td><td>11.1%</td><td></td><td></td><td>16.6%</td><td></td><td></td></tr>
</table>

Note:
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1
Not significant variables were not estimated in the final model

Pedestrian-as-Victim crash frequency models

The estimation results of pedestrian crashes are presented in TABLE 6 and TABLE 7. Compared to the bike flow, pedestrian flow was found to be highly correlated with pedestrian crashes. In accordance with the previous models, motor traffic volume is also a significant predictor in pedestrian crashes. The model coefficients reveal that for each increase of 1,000 in motor traffic flow, there is a 4.6% increase in the likelihood of having a light injury pedestrian crash. Tertiary roads are associated with the increase in the likelihood of switching from zero crash link to non-zero crash link. Residential and minor roads show a negative relationship with number of light injury pedestrian crashes. One possible reason might be the low motor traffic volume on minor roads resulting in fewer conflict points between vehicles and pedestrians. In terms of severe-fatal pedestrian crashes, tertiary and residential roads remain negatively correlated to the crash generation.

TABLE 7 Estimation results of ZIP models for severe-fatal crash models in case 2, 3 and 4

	Case 2			*Case 3*			*Case 4*		
	θ	Pr(>|z|)		θ	Pr(>|z|)		θ	Pr(>|z|)	
Step 1: Zero-inflation model coefficients (binomial with logit link)									
Intercept	-3.13	0.00	***	-1.94	0.00	***	-2.73	0.00	***
Motor traffic volumes in 1,000	0.23	0.02	*						
Bike flows in 1,000	6.32	0.00	***	5.74	0.00	***			
Pedestrian flows in 1,000							2.80	0.00	***
is primary road	base			base			base		
is secondary road	1.05	0.00	**						
is tertiary road	0.65	0.05	.				-0.37	0.09	.
is residential or minor road	1.14	0.00	**	0.56	0.01	**			
Step 2: Count model coefficients (Poisson with log link)									
Intercept	-5.81	0.00	***	-7.06	0.00	***	-5.99	0.00	***
Motor traffic volumes in 1,000	0.01	0.09	.	0.01	0.11		0.01	0.17	
Bike flows in 1,000	0.08	0.00	***	0.10	0.00	***			
Pedestrian flows in 1,000							0.11	0.00	***
is primary road	base			base			base		
is secondary road	-0.21	0.03	*						
is tertiary road							-0.22	0.08	.
is residential or minor road	-0.43	0.00	***	-0.22	0.04	*	-0.78	0.00	***
log(link length)	0.45	0.00	***	0.62	0.00	***	0.54	0.00	***
Number of intersections	0.04	0.00	***						
Efron's pseudo R2	*2.3%*			*1.6%*			*4.0%*		
McFadden pseudo R2	*9.3%*			*8.1%*			*15.2%*		

Note:
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
Not significant variables were not estimated in the final model

CONCLUSIONS

This study built eight crash estimation models accounting for different road users including car occupants, cyclists, and pedestrians. Conventional statistical approaches were applied to explore the impacts of traffic volume factors and road characteristics factors on the number of severe-fatal crashes and number light injury crashes in different accident types.

One contribution of this study is that it found significant differences between the predictor sets for all cases. Firstly, motor traffic volume, as the most common predictor used in the literature, is identified to be consistently significant. However, magnitudes of the effects varied from car occupancy model to cyclist and pedestrian model. The effect of motor traffic volume is dominant in car occupancy crashes. On the contrary, it slightly influences cyclist and pedestrian crashes. Moreover, different from the typical link/segment crash models, this study accounted for all road link types in the model estimation. It provides the capability to assess the large-scale safety level of a city or a regional area. Although the models have relatively low goodness-of-fit, it can be explained by imperfect crash data and limited data accessibility to important variables like the presence of bike lanes. Rich data consisting of several variables such as the presence of bike lanes, road infrastructure design, and road conditions, are typically collected for the small-scale area. Such detailed information across an urban area or even region is difficult to obtain. Most importantly, with the simple model structure, those crash estimation models are practical in aiding transportation safety planning. Also, it is easier to transfer those models to other study areas due to the less requirement in data collection.

In future research, one practice of these crash estimation models is to be integrated into an agent-based transport model. We plan to assess the number of car occupants, bike and pedestrian crashes in the Munich metropolitan area. With the benefit of agent-based simulation, the model is also able to assess the risk of traffic crash for each road user. Based on the estimated number of crashes resulting from the presented models, the risk exposure rate can be measured for each link in the study area. The specified links where each agent traveled is simulated in the agent-based transport model. Hence, the crash risk/likelihood of each road user can be calculated by cumulating the risk exposure rates of all traveled links. Then Monte Carlo simulation will be applied to determine the occurrence of a crash on each road user. Crash occurrence can be further employed in the transport model. The agents occurred a fatal accident lead to a death event in the simulation, while severely injured agents will have the probability of disability which affects their long-term mobility such as the limited access to car mode and less accessibility to public transport.

ACKNOWLEDGEMENTS

The research was conducted with the European Union GLASST project (Global and Local Health Impact Assessment of Transport), funded by the European Research Council (ERC).

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Qin Zhang, Rolf Moeckel, Mehmet Baran Ulak, Carlos Llorca and James Woodcock; data collection: Qin Zhang; analysis and interpretation of results: Rolf Moeckel, Mehmet Baran Ulak, Carlos Llorca and James Woodcock; draft manuscript preparation: Qin Zhang. All authors reviewed the results and approved the final version of the manuscript.

REFERENCE

1. German Federal Statistic Office. Unfallatlas: Traffic Accidents Data. https://www.destatis.de/EN/Themes/Society-Environment/Traffic-Accidents/_node.html. Accessed Mar. 4, 2020.
2. Lord, D., and F. Mannering. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A: Policy and Practice*, Vol. 44, No. 5, 2010, pp. 291–305. https://doi.org/10.1016/j.tra.2010.02.001.
3. AASHTO. Highway Saftey Manual. *American Association of State Highway and Transportation Officials*. http://www.highwaysafetymanual.org/Pages/default.aspx. Accessed Apr. 1, 2020.
4. So, J., S. Hoffmann, J. Lee, F. Busch, and K. Choi. A Prediction Accuracy-Practicality Tradeoff Analysis of the State-of-the-Art Safety Performance Assessment Methods. *Transportation Research Procedia*, Vol. 15, 2016, pp. 794–805. https://doi.org/10.1016/j.trpro.2016.06.066.
5. Aldred, R., A. Goodman, J. Gulliver, and J. Woodcock. Cycling Injury Risk in London: A Case-Control Study Exploring the Impact of Cycle Volumes, Motor Vehicle Volumes, and Road Characteristics Including Speed Limits. *Accident Analysis & Prevention*, Vol. 117, 2018, pp. 75–84. https://doi.org/10.1016/j.aap.2018.03.003.
6. Amoh-Gyimah, R., M. Saberi, and M. Sarvi. Macroscopic Modeling of Pedestrian and Bicycle Crashes: A Cross-Comparison of Estimation Methods. *Accident Analysis and Prevention*, Vol. 93, 2016, pp. 147–159. https://doi.org/10.1016/j.aap.2016.05.001.
7. Siddiqui, C., M. Abdel-Aty, and K. Choi. Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 382–391. https://doi.org/10.1016/j.aap.2011.08.003.
8. Kim, K., P. Pant, and E. Yamashita. Accidents and Accessibility: Measuring Influences of Demographic and Land Use Variables in Honolulu, Hawaii. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2147, No. 1, 2010, pp. 9–17. https://doi.org/10.3141/2147-02.
9. Liu, C., M. Zhao, W. Li, and A. Sharma. Multivariate Random Parameters Zero-Inflated Negative Binomial Regression for Analyzing Urban Midblock Crashes. *Analytic Methods in Accident Research*, Vol. 17, 2018, pp. 32–46. https://doi.org/10.1016/j.amar.2018.03.001.
10. Mannering, F. Temporal Instability and the Analysis of Highway Accident Data. *Analytic Methods in Accident Research*, Vol. 17, 2018, pp. 1–13. https://doi.org/10.1016/j.amar.2017.10.002.
11. Norman, T. Severity on Two-Lane , Undivided Roadways in North Carolina. *Transportation Research Record*, No. 99, 1990, pp. 78–85.
12. Dumbaugh, E., and W. Li. Designing for the Safety of Pedestrians, Cyclists, and Motorists in Urban Environments. *Journal of the American Planning Association*, Vol. 77, No. 1, 2010, pp. 69–88. https://doi.org/10.1080/01944363.2011.536101.
13. Pahukula, J., S. Hernandez, and A. Unnikrishnan. A Time of Day Analysis of Crashes Involving Large Trucks in Urban Areas. *Accident Analysis and Prevention*, Vol. 75, 2015, pp. 155–163. https://doi.org/10.1016/j.aap.2014.11.021.
14. Chen, P., and Q. Shen. Built Environment Effects on Cyclist Injury Severity in Automobile-Involved Bicycle Crashes. *Accident Analysis and Prevention*, Vol. 86, 2016, pp. 239–246. https://doi.org/10.1016/j.aap.2015.11.002.
15. Moeckel, R., N. Kuehnel, C. Llorca, A. Moreno, and H. Rayaprolu. Microscopic Travel Demand Modeling: Using the Agility of Agent-Based Modeling Without the Complexity of Activity-Based Models. 2019.
16. Horni, A., K. Nagel, and K. W. Axhausen. *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, London, 2016.
17. Greene, W. H. Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models. *NYU Working Paper No. EC-94-10*, Vol. 9, 1994, pp. 265–265. https://doi.org/10.1007/BF00857937.