Tasic I, Elvik R, Brewer S. Exploring the safety in numbers effect for vulnerable road users on a macroscopic scale.

Accident Analysis & Prevention. 109 (December), 2017, 36-46. 0001-4575

The article has been published in final form by Elsevier at

http://dx.doi.org/10.1016/j.aap.2017.07.029

The final publication id available in: Accident Analysis and Prevention, 109, 2017, 36-46

10.1016/j.aap.2017.07.029

Exploring the Safety in Numbers Effect for Vulnerable Road Users on a Macroscopic Scale

Ivana Tasic, University of Utah, Department of Civil and Environmental Engineering, 110 Central Campus Drive, Room 1650, Salt Lake City, 84112 UT, United States

Rune Elvik, Institute of Transport Economics

Simon Brewer, University of Utah, Department of Geography

Abstract

A "Safety in Numbers" effect for a certain group of road users is present if the number of crashes increases at a lower rate than the number of road users. The existence of this effect has been invoked to justify investments in multimodal transportation improvements in order to create more sustainable urban transportation systems by encouraging walking, biking, and transit ridership. The goal of this paper is to explore safety in numbers effect for cyclists and pedestrians in areas with different levels of access to multimodal infrastructure. Data from Chicago served to estimate the expected number of crashes on the census tract level by applying Generalized Additive Models (GAM) to capture spatial dependence in crash data. Measures of trip generation, multimodal infrastructure, network connectivity and completeness, and accessibility were used to model travel exposure in terms of activity, number of trips, trip length, travel opportunities, and conflicts. The results show that a safety in numbers effect exists on a macroscopic level for motor vehicles, pedestrians, and bicyclists.

Key words: Safety in numbers, multimodal transportation, vulnerable road users, urban context

1 Introduction

One of the general concerns about investing and planning for more sustainable multimodal transportation infrastructure is the increase in exposure of road users who are vulnerable to crashes. However, as cities grow and develop, robust multimodal systems are required to enable adequate integration of land use and transportation, and provide viable travel options for non-driving populations. This is where the concept of Safety in Numbers emerges, supported by the assumption that more multimodal travel options would lead to more walking and cycling, which would be associated with an increase in crashes, but less than proportional with the increase in walking and cycling (Elvik, 2016).

This paper explores the existence of a Safety in Numbers effect in the context of a major city in the U.S., relying on a detailed dataset on multimodal infrastructure, and using a combination of exposure measures. The goal of the paper is to determine whether the effect of Safety in Numbers exists on a macroscopic level, for vehicular users, pedestrians and bicyclists. The research aims to contribute to the current literature on urban safety and inform the practice of planning for multimodal solutions with consideration of safety effects.

This paper uses data from Chicago aggregated on the census tract level, to estimate the expected number of vehicle-only (vehicular), vehicle-pedestrian (pedestrian), and vehicle-bicyclist (bicyclist) total and injurious (severe) crashes. Generalized Additive Models (GAM) are used in the Statistical Area Safety Modeling (SASM) framework to model these six crash types and address the potential effects of spatial autocorrelation in spatially aggregateddata. Measures of exposure are derived from travel demand model estimates and complemented by proxies for exposure that include the representation of multimodal infrastructure and accessibility.

The following section of the paper reviews literature on the safety in numbers effect, as well as the interaction between it and the provision of infrastructure facilitating multimodal trips.

The data and methods are described in the third section of the paper, while the results and the discussion follow in section four. The final section provides summary of research findings and recommendations for the future research focusing on vulnerable road users on the macroscopic level.

2 Literature Review

One of the first studies that focused on discovering whether the Safety in Numbers effect exists, used data from Oakland, California, to examine the relationship between pedestrian volume and the rate of pedestrian-vehicle crashes (Geyer et al., 2006). That study and many later studies have confirmed the existence of a safety in numbers effect, but one needs to know the mechanisms producing the effect if one aims to exploit it in planning infrastructure to encourage walking or cycling (Bhatia and Wier, 2011). The early studies on Safety in Numbers focused on pedestrians (Geyer, 2006; Bhatia and Wier, 2011)., There are fewer studies focusing on bicyclists (Johnson et al., 2014; Fyhri et al., 2016), and even fewer studies that include both pedestrians and cyclists (Elvik, 2016). More recent studies attempt to demonstrate the safety in numbers effect on a macroscopic scale for potential use in planning and predictions on the zonal level, but again only focus on a single group of road users (Wang and Kockelman, 2014). Previous research generally concludes that a deeper understating and more knowledge on Safety in Numbers effect is required (Bhatia and Wier, 2011; Elvikand Bjørnskau 2017).

This paper focuses on exploring whether a “safety in numbers’ effect exists for pedestrians and cyclists in a major U.S. city, when safety is evaluated at a macroscopic level.

3 Methodology

The City of Chicago served as a case study, and data aggregation on the census tract level helped with capturing the integration of land use mixture and multimodal transportation system features. Data sources used in this study are provided in Table 1. The measures of exposure used in this study are a combination of measures obtained from city transportation agencies and measures developed by the research team. The SASM framework relied on Generalized Additive Models developed for all six crash types addressed, as this approach was found to be a good alternative based on frequentist statistical inference in earlier studies (Tasic et al., 2016). The results of the SASM served to explore whether there is a Safety in Numbers effect present for multimodal road users.

Table 1 Data Sources, Descriptions, and Formats

Data	Source	Year	Format
Crash records	Illinois DOT, Chicago Crash Browser	2005-2012	csv
Socio-economic characteristics	U.S. Bureau of Census, ACS 5-Year Estimates	2008-2012	csv
Land use	Chicago Metropolitan Agency for Planning	2010	shp
Road network	City of Chicago	2012	shp
Travel demand model	Chicago Metropolitan Agency for Planning	2010	csv, shp
Other traffic volume data	Illinois DOT	2014	csv
L Train lines, stops and ridership	Chicago Transit Authority	2012	csv, shp
Bus lines, stops and ridership	Chicago Transit Authority	2012	csv, shp
Bike lanes and bike racks	City of Chicago	2012	shp
Sidewalk	City of Chicago	2012	shp
Commuter trips to work by means	U.S. Bureau of Census, ACS 5-Year Estimates	2008-2012	csv
Spatial units of analysis	City of Chicago	2012	shp

Data Collection

As shown in Table 1, the combination of official transportation agency data sources as well as open source data platforms enabled the development of a dataset consisting of roughly one hundred variables that represent the variety of factors potentially influencing safety in major U.S. cities. Data collected included crashes, multimodal transportation features (i.e. features facilitating the use of more than one mode of transportation on a given trip), road network features and traffic conditions, land use data, socio-economic characteristics, and spatial features supporting the selection of appropriate spatial units of analysis. Data were obtained from the Illinois Department of Transportation (DOT), Chicago Metropolitan Agency for Planning

(CMAP), Chicago Transit Authority (CTA), City of Chicago, U.S. Bureau of Census, as well as the available open data platforms supported by the City of Chicago.

Variables and Measures

The variables and measures developed from the data collected for this study can be divided into crash-related variables, exposure variables, surrogates for exposure, and variables that represent area-wide effects that influence crashes. This study uses six crash-related variables to model vehicle-only (vehicular) total crashes, vehicular fatal and injury (severe) crashes, pedestrian total crashes, pedestrian severe crashes, bicyclist total crashes, and bicyclist severe crashes. Total number of vehicular and nonmotorized trips estimated through the City of Chicago Air Quality Conformity Study conducted by CMAP served as primary measures of vehicular, pedestrian, and bicyclist exposure. As these are estimates, and additional proxies for exposure were needed to represent locations with a high concentration of activity and potential for conflicts, data on multimodal infrastructure and accessibility measures served as surrogates for exposure. In particular, accessibility measures that reflect the ease of reaching specific destinations via walking, biking, and transit mode were developed using the methods from previously published research on multimodal accessibility (Tasic et al., 2014a; Tasic et al., 2014b). In addition, the complexity of urban environment in major cities was captured by adding the variables on socio-economic characteristics and land use. All these variables were aggregated on the census tract level, as this was found to be the most appropriate unit of analysis for the purpose of this study, due to compatibility with socio-economic data, general data availability, as well as the size of the unit that adequately captures the characteristics of multimodal transportation systems. Table 2 provides the complete list of variables used in the SASM process.

Statistical Area Safety Modeling

Several approaches based on both frequentist and Bayesian statistical inference were explored in the SASM context, particularly making sure that they account for spatial autocorrelation that may appear when the data are spatially aggregated. Spatial safety studies conducted in the past prove that Bayesian Hierarchical Models have the ability to deal with various issues that may arise in spatially aggregated crash data. A more recent study showed that GAM (Generalized Additive Models) are able to provide results comparable to Bayesian models, while dealing with spatial autocorrelation present in the data, and using frequentist methods of statistical inference to interpret the relationships between explanatory and outcome variables. This study used GAM as the primary method for SASM of vehicular, pedestrian, and bicyclist total and severe crashes. GAM is an additive extension to the family of generalized linear models introduced by Hastie and Tibshirani (1990). In addition to the parameters related to explanatory variables, these models also estimate smoothing functions of explanatory variables that have interactions among each other, or other type of effects that may influence the outcome variables estimation, such as spatial or temporal correlation.

Table 2 Descriptive Statistics (801 census tract observations)

Variable	**Description**	**Mean**	**Std. Dev.**	**Min**	**Max**
Crash Variables					
VehCrash	Vehicle-only Crashes	375.176	354.534	5	3920
Veh_KA	Vehicle-only Fatal and Severe Injury Crashes	8.004	8.465	0	71
PedCrash	Crashes Involving Pedestrians	17.750	22.528	0	481
Ped_KA	Fatal and Severe Injury Crashes Involving Pedestrians	2.131	2.555	0	41
BikeCrash	Crashes Involving Bicyclists	9.528	13.178	0	172
Bike_KA	Fatal, and Severe Injury Crashes Involving Bicyclists	0.783	1.293	0	12
Socio-economic variables					
Population	Population Size	3.402	1.741	0.000	15.740
Pop_Dens	Population Density per mile squared	18.203	20.206	0.000	485.019
Employed	Percent of Employed Population	6.759	18.955	0.000	86.000
Unemploy	Percent of Unemployed Civil Population	14.970	9.459	0.000	51.000
PerCapInc	Average Income per Capita	27,786.690	20,029.490	0.000	131,548.000
NoVeh	Households with no Vehicles, %	26.537	15.118	0.000	89.400
Veh1	Households with 1 Vehicle, %	43.589	9.508	0.000	81.300
Veh2	Households with 2 Vehicles, %	22.558	11.544	0.000	59.100
Veh3plus	Households with 3 or more Vehicles, %	6.814	5.648	0.000	26.900
Infrastructure Variables					

Road	Total Length of Roads, miles	6.278	3.910	0.142	30.762
EXPY	Expressways, % of street network	0.219	0.601	0.000	4.302
Art	Arterials, % of street network	0.924	0.790	0.000	7.675
Exp_Art	Expressways and Arterials, % of street network	1.143	1.115	0.000	11.976
Coll	Collectors, % of street network	0.876	0.668	0.000	4.668
Street	Other Streets, % of street network	3.848	2.694	0.000	15.096
Alley	Named Alleys, % of street network	0.000	0.000	0.000	0.000
BikeLane	Total Length of Bike Lanes, miles	0.679	0.723	0.000	6.163
BusRoute	Total Length of Bus Routes, miles	1.541	2.559	0.000	39.980
Ltrain	Total Length of L Train Lines, miles	0.147	0.353	0.000	4.411
Sidewalk	Total Sidewalk Area, feet squared	287.382	198.201	0.000	1,131.373
Intersect	Total Number of Intersections	37.803	27.800	0.000	163.000
Connect	Connectivity Index, intersections/mile of road	5.798	1.531	0.000	16.232
Signal_P	Signalized Intersections, %	0.123	0.141	0.000	1.333
BusStops	Total Number of Bus Stops	13.104	9.099	0.000	75.000
LStops	Total Number of L Train Stops	0.091	0.325	0.000	2.000
BikeRacks	Total Number of Bike Racks	6.446	11.394	0.000	220.000
DVMT	Daily Vehicle Miles Traveled	40,563.580	57,246.750	8.057	522,024.400
Ped	Pedestrian Trips Generated	47.715	103.345	1.191	1581.315
Bike	Bicyclist Trips Generated	2.511	5.439	0.062	83.227
DriveAlone	Drive-alone Trips to Work, %	50.186	15.522	0.000	86.300
Carpool	Carpool Trips to Work, %	9.511	6.560	0.000	39.500
Transit	Transit Trips to Work, %	27.506	12.956	0.000	79.100
Walk	Walk Trips to Work, %	0.603	3.156	0.000	35.000
OtherMeans	Trips to Work by Other Means, %	2.542	2.942	0.000	21.300
WorkHome	Work Home, %	4.058	3.296	0.000	21.300
TT_min	Average Travel Time to Work, minutes	34.019	6.303	0.000	56.500
Connectivity and Accessibility Measures					
NC_Car	Network Serving Cars only (%)	0.030	0.076	0	0.543
NC_Car_W	Network Serving Cars and Pedestrians (%)	0.577	0.286	0	1.000
NC_CarWT	Network Serving Cars, Pedestrians and Transit (%)	0.438	0.930	0	1.000
NC_Car_WB	Network Serving Cars, Pedestrians and Bicyclists (%)	0.061	0.078	0	0.528
NC_Car_WTB	Network Serving All Modes (%)	0.085	0.143	0	1.000
Ped_D15	No. of destinations accessible within 15 minute walk time	34.491	67.816	0	783
Ped_A	Weighted pedestrian accessibility	21.301	28.609	0	310.65
Bike_A	Weighted bicyclist accessibility	1322.587	1584.606	0	11085.87
Transit_A	Weighted transit accessibility	73.694	52.355	0	342.84
Land Use Variables					
LUDiv	Total number of land uses	5.916	1.012	1.000	8.000
Entropy	Land use entropy	0.472	0.123	0.015	0.802

The inclusion of smoothing functions gives GAM more flexibility in terms of model specification and description of the relationships between the explanatory variables, than the generalized linear models allow for. In the case of this study, GAM is used to incorporate the smoothing function across the locations, in order to account for spatial autocorrelation present in data aggregated on the census tract level. A two-dimensional spatial trend function included in GAM serves to capture these effects in the following manner (Wood, 2006):

$$log(\theta_i) = \beta_0 + \sum_j \beta_j\, x_{ij} + f_i(lat_i, lon_i) + \varepsilon \qquad \text{Equation (1)}$$

Where:

θ_i - expected number of crashes for census tract "i"

β_0- intercept

β_i - coefficients quantifying the effect of the "j" explanatory variables characterizing spatial unit "i" on θ_i

x_{ij} – a set of "j" explanatory variables that characterize census tract "i" and influence θ_i

ε_i - disturbance term corresponding to census tract "i"

$f_i(lat_i, lon_i)$ –two-dimensional smooth function for modeling spatial trends in census tract "i"

The GAM parameters are estimated by penalized likelihood maximization (Hastie and Tibshirani, 1990; Wood, 2006). The essential part of parameter estimation in GAM is estimating a smooth function f_i by choosing an adequate basis to represent the smooth function as a linear model. In the case when a smooth function is assumed to be two-dimensional in order to account for spatial dependence as it is in the case of this research, the adequate basis is penalized thin plate regression spline, explained in detail in Wood (2006). This study estimates six outcome variables using GAM, for three user types: vehicles, pedestrians, and bicyclists. The exposure for vehicular crashes is defined using the variables DVMT and length of roadway network in miles. The exposure for pedestrian crashes is defined using the product of DVMT and the estimated number of pedestrian trips generated within the census tract. The exposure for bicyclist crashes is defined using the product of DVMT and the estimated number of bike trips generated within the census tract. The basic assumption with respect to exposure measures was that no crashes of a particular type have occurred within the census tract when at least one of the exposure measures used in each model is zero. The example of basic GAM specification used to estimate vehicular, pedestrian, and bicyclist crashes is provided in Equations 2, 3, and 4:

$$\theta_{veh\_i} = e^{(\beta_0+\beta_1 \ln(DVMT)+\beta_2 \ln(\text{Road})+\sum_j \beta_j x_{ij}+f_i(lat_i,lon_i)+\varepsilon_i)} \quad \text{Equation (2)}$$

$$\theta_{ped\_i} = e^{(\beta_0+\beta_1 \ln(DVMT)+\beta_2 \ln(Ped)+\sum_j \beta_j x_{ij}+f_i(lat_i,lon_i)+\varepsilon_i)} \quad \text{Equation (3)}$$

$$\theta_{bike\_i} = e^{(\beta_0+\beta_1 \ln(DVMT)+\beta_2 \ln(\text{Bike})+\sum_j \beta_j x_{ij}+f_i(lat_i,lon_i)+\varepsilon_i)} \quad \text{Equation (4)}$$

Where:

$\theta_{\mathrm{veh\_i}}$ - expected number of vehicular crashes for census tract “i”

$\theta_{\mathrm{ped\_i}}$ - expected number of pedestrian crashes for census tract “i”

$\theta_{\mathrm{bike\_i}}$ - expected number of bicyclist crashes for census tract “i”

β_0- intercept

β_i - coefficients quantifying the effect of the “j” explanatory variables characterizing spatial unit “i” on θ_i

x_{ij} – a set of “j” explanatory variables that characterize census tract “i” and influence θ_i

ε_i - disturbance term corresponding to census tract “i”

$f_i(\mathrm{lat_i}, \mathrm{lon_i})$ –two-dimensional smooth function for modeling spatial trends in census tract “i”

Statistical model diagnostics

The following values served as an indicator of model goodness of fit for GAM developed for the modeled outcome variables:

- Smooth terms: Two-dimensional smooth function parameters based on penalized thin-plate regression splines. Coefficient estimates of smooth terms provided with standard errors and *p*-values indicate the statistical significance of smooth functions included to account for spatial autocorrelation

- Deviance explained: The percentage of deviance explained, based on the sum of squares of the deviance residuals, as the model deviance, and the sum of squares of the deviance residuals when the covariate effects are set to zero, as the null deviance
- Adj. R2: Adjusted *r*-squared as the proportion of variance explained
- REML: The value of restricted (or penalized) maximum likelihood function

4 Results and Discussion

Vehicle-only total and severe crashes

Table 3 provides the final model specification for vehicle-only total and severe crashes. The results indicate that the expected total crash frequency increases as population density increases, and similar findings have been reported in the literature (Flask and Schneider, 2013; Castro et al., 2013; Noland and Quddus, 2004). Other socio-economic variables were not significantly related to the frequency of vehicular crashes. This could be due to the fact that vehicular trips can be generated through the census tract areas regardless of economic status of the population. The natural logarithm of road mileage and DVMT were used as the exposure variable to estimate the expected number of vehicular crashes on the census tract level. Although DVMT is calculated from the ADT values related to each link in a census tract multiplied by the length of corresponding links, thus incorporating road segments length into this measure, total road mileage is still included in vehicular crash models. Suppose there are two census tracts with the same DVMT. One of the census tracts could have a denser road network and higher road mileage with lower volumes of traffic, while the other census tract could have fewer roads but higher traffic volumes resulting with the same DVMT value. It is expected that these two hypothetical census tract areas would have different number of vehicular crashes, even though their DVMT value is the same, due to differences in the road network structure, and traffic flow intensity and its distribution across the network. This is why the road mileage variable was included as an additional exposure variable in vehicular crash models. As expected, increases in total length of roads in miles and daily vehicle miles traveled, were associated with an increase in expected crash frequency at the ninety nine percent confidence level. Intersection-related variables, such as intersection density and the percentage of signalized intersections, were found to be positively associated with the number of crashes. Previous studies have found some similar relationships between network and intersection densities and crash frequencies (Moeinaddini, 2014; Siddiqui, 2012). Presence of bus stops was associated with an increase in expected vehicle crash frequency. The total area of sidewalk was associated with a decrease in vehicular crash frequency.

As shown in Table 3, the variables that are found to be associated with the expected number of severe vehicular crashes include road mileage, DVMT, percent of signalized intersections, bus stops, sidewalk area, L train stops, and land use entropy. Similar to total vehicular crashes, the estimated model results show that it can be expected that the number of severe crashes is almost proportional to road segment lengths, while the increase in traffic volumes is not associated with a proportional increase in crashes. The presence of traffic signals is positively associated with both total and severe vehicular crashes. The presence of sidewalk was associated with a decrease in severe vehicular crashes. The results of the models estimated for total and severe vehicular crashes indicate that increasing sidewalk area could be considered as a safety countermeasure in urban environments. A limitation that should be remembered is that very few cities would have the available data on sidewalk area coverage, which limits the application of the model for different locations.

Total and Severe Pedestrian Crashes

The estimated statistical models for the expected number of total and severe pedestrian crashes in census tracts are provided in Table 4. The estimated coefficients for the exposure variables indicate that the expected number of crashes does not increase proportionally with the increase in vehicular or pedestrian trips, and this effect will be discussed in one of the following sections. Pedestrian crashes are expected to increase as pedestrian accessibility increases as a function of the number of accessible destinations and travel time to destinations. The total number of destinations that pedestrians are able to reach within a fifteen-minute walk is associated with a decrease in pedestrian crashes. These two variables have different signs, indicating that the concentration of destinations in such a way that it decreases the length of pedestrian travel time could lead to pedestrian crash reduction. Further analyses of the relationships between the accessibility related measures, exposure, and crashes is required seeking to incorporate utility-based measures and match accessibility indicators with pedestrian exposure. Variables that represent functional classification, conflict points, and intersection traffic control are associated with an increase in pedestrian crashes. Street connectivity is associated with a reduction of pedestrian crashes. The presence of signalized intersections is associated with a higher number of pedestrian crashes, and appears to be the major driver of pedestrian crash occurrence among the variables in the pedestrian crash model. Similar effects of the presence of signalized intersections on pedestrian crashes have been reported in previous research (Ukkusuri et al., 2012; Elvik 2016). The product of DVMT and the number of pedestrian trips within census tract estimated from the CMAP trip generation model served as the main indicator of pedestrian exposure to crashes. It was assumed that if either of these two variables (DVMT or the number of pedestrian trips) is equal to zero, no pedestrian crashes would be expected. Additional measures that would serve as a proxy for exposure were considered during the statistical modeling process, including the total road mileage and sidewalk area. In the case where roadway mileage was included in the models, sidewalk area was treated as a form of pedestrian safety countermeasure. A better, more complete measure that indicates pedestrian presence on roadway facilities, particularly in the context of the potential conflicts between other modes of travel, was the indicator of network completeness, expressed as the percentage of network that serves all four modes. Statistical models that serve to estimate pedestrian crash outcome, particularly on a spatially aggregated level, should include some indicators related to pedestrian infrastructure that would complement the measures of exposure. Whether simply road mileage, or sidewalk area, or in this case an indicator of the presence of complete streets in the network is used, will depend primarily on the data availability and the complexity of networks.

The variables that were associated with the expected number of severe pedestrian crashes include DVMT, the number of pedestrian trips within the census tract, weighted pedestrian accessibility, the number of destinations accessible within fifteen minute walking time, and the percent of signalized intersections (Table 4).

Total and Severe Bicyclist Crashes

The estimated statistical models for the expected number of bicyclist crashes in census tracts are provided in Table 5. The variables that were associated with the expected number of bicyclist crashes include the estimated DVMT, the number of bike trips within the census tract, weighted bicyclist accessibility, intersection density, bus stops, bike lanes mileage, CBD, and the presence of L Train lines. The estimated DVMT and the number of generated bike trips were used as the primary measures of exposure. The estimated coefficients for volumes of vehicles and bicyclist trips in census tracts show non-linear relationship with the number of bicyclist crashes. Similar to pedestrian crashes, the expected number of bicyclist crashes is increasing much less than proportional to vehicular and bicyclist volume. Bike lanes mileage, weighted bicyclist accessibility, and intersection density served as approximate measures of the opportunities for conflicts between bicyclists and vehicles. It was found that doubling the mileage of bike lanes is not associated with a proportional increase in bicyclist crashes. This is probably due to the fact that biking may also be present on road segments that do not include bike lanes. Weighted bicyclist accessibility is an indicator of the s concentration of locations accessible by bike in the city, and is included in the model as a statistically significant variable. Intersection density proved to be statistically significant in the total bicyclist crashes model specification. Bicyclists are more exposed to crashes than pedestrians in terms of spatial opportunities for conflicts, as the conflicts may occur anywhere along the roadway segments, which may be the reason why the type of intersection traffic control is less significant. The CBD area, presence of bus stops and the presence of L Train facilities proved to be significantly related to the expected number of bicyclist crashes. The downtown area in Chicago tends to be more oriented towards non-motorized modes, with better defined biking facilities network. However, additional analysis is needed to determine if different types of biking facilities (e.g., protected bike lanes), tend to lead to reduction of biking crashes. Biking trips do have higher concentrations in the downtown area, and given the estimated coefficient that indicates that bicyclist crashes are less likely to occur in CBD area, this could confirm the non-linear relationship between the number of people biking and bicyclist crash outcome. The L Train facilities was associated with fewer bicyclist crashes.. It is important to acknowledge here that this association may not be the result of the presence of train facilities, but the environment that is created due to the particular design of train line and station facilities in Chicago.

The variables that were associated with the expected number of severe bicyclist crashes include the estimated DVMT, the number of bike trips within the census tract, weighted bicyclist accessibility, and bike lanes mileage. All variables that were statistically significant in the estimated severe bicyclist crash model were already included in the total bicyclist crashes model specification.

Figures 2 and 3 show the visualization of GAM models for the six outcome variables modeled in this study. Selected significant variables are included in these visualizations to demonstrate how the expected number of all crash types changes with the change in these variables. The influence of smooth terms representing spatial dependence among the census tract entities is also included in the figures, showing how different outcome variables have different level of sensitivity to spatial autocorrelation. Generally, stronger spatial dependence exists among total crashes than among severe crashes, for all crash types.

Safety in Numbers Effect

Figures 4-6 show the relationships between the exposure variables and associated crash rates for total and severe crashes.

As previously explained, non-motorized crashes were estimated using the combinations of exposure measures for private vehicles (DVMT) and the non-motorized modes (pedestrian trips or bicyclist trips). The relationships between crash rates for all crash types and the related exposure measures are non-linear. The estimated coefficients for the exposure variables in all estimated models have a value lower than one. This means that while the crashes are expected to increase as the exposure increases, the increase in crash rate is expected to be much lower than the rate at which the exposure increases, and this is true for all crash types. This effect is known in the literature as the "Safety in Numbers" effect (Hauer, 1982; Elvik and Bjørnskau 2017).

As the exposure increases for private vehicle users, both total vehicular crashes and severe vehicular crashes increase at almost the same rate, which is much lower than the increase in DMVT, implying that risk goes down as shown in Figure 4. Similar conclusions can be drawn for pedestrian and bicyclist crashes (Figures 5 and 6), as it is expected that with the increase in exposure (both vehicular and non-motorized), the number of crashes increases at a lower rate than increase in exposure.

The estimated crash rates based on the models for severe vehicular and severe pedestrian crashes show that for the same rate of change in the estimated DVMT, while all other variables (including the pedestrian user exposure) remain constant, the expected rate of severe vehicular crashes is two times higher than the expected rate of severe pedestrian crashes. . Severe pedestrian crashes are more sensitive to change in pedestrian exposure than to change in vehicular exposure, while according to the estimated models severe bicyclist crashes are almost equally strongly related to bicyclist and vehicular exposure.

Table 3 Recommended Model for Total & Severe Vehicle-only Crash Estimation

Vehicular Crashes	**Generalized Additive Model**		
Variables	Coeff.	Std. Err.	P>|z|
Population Density	0.0086	0.0009	0.000
ln (Road Mileage)	0.9912	0.0600	0.000
ln (DVMT)	0.2607	0.0199	0.000
Intersection Density	0.0011	0.0003	0.000
Signalized Intersections (%)	1.5351	0.1202	0.000
Bus Stops	0.0071	0.0023	0.002
Sidewalk Area	-0.2282	0.0421	0.000
Intercept	1.1410	0.1737	0.000
Smooth terms	7.7200	8.1730	0.000
Deviance explained	76.30%	*Adj. R2*	0.801
	REML = 4923		

Severe Vehicular Crashes	**Generalized Additive Model**		
Variables	Coeff.	Std. Err.	P>|z|
ln (Road Mileage)	0.9600	0.0925	0.000
ln (DVMT)	0.2641	0.0309	0.000
Signalized Intersections (%)	1.3066	0.1707	0.000
Bus Stops	0.0054	0.0030	0.000
Sidewalk Area	-0.1536	0.0546	0.074
L Train Stops	-0.0856	0.0614	0.005
Intercept	-2.5875	0.2541	0.078
Smooth terms	6.0100	7.0520	0.001
Deviance explained	63.20%	*Adj. R2*	0.721
	REML = 2109		

Table 4 Recommended Model for Total & Severe Pedestrian Crash Estimation

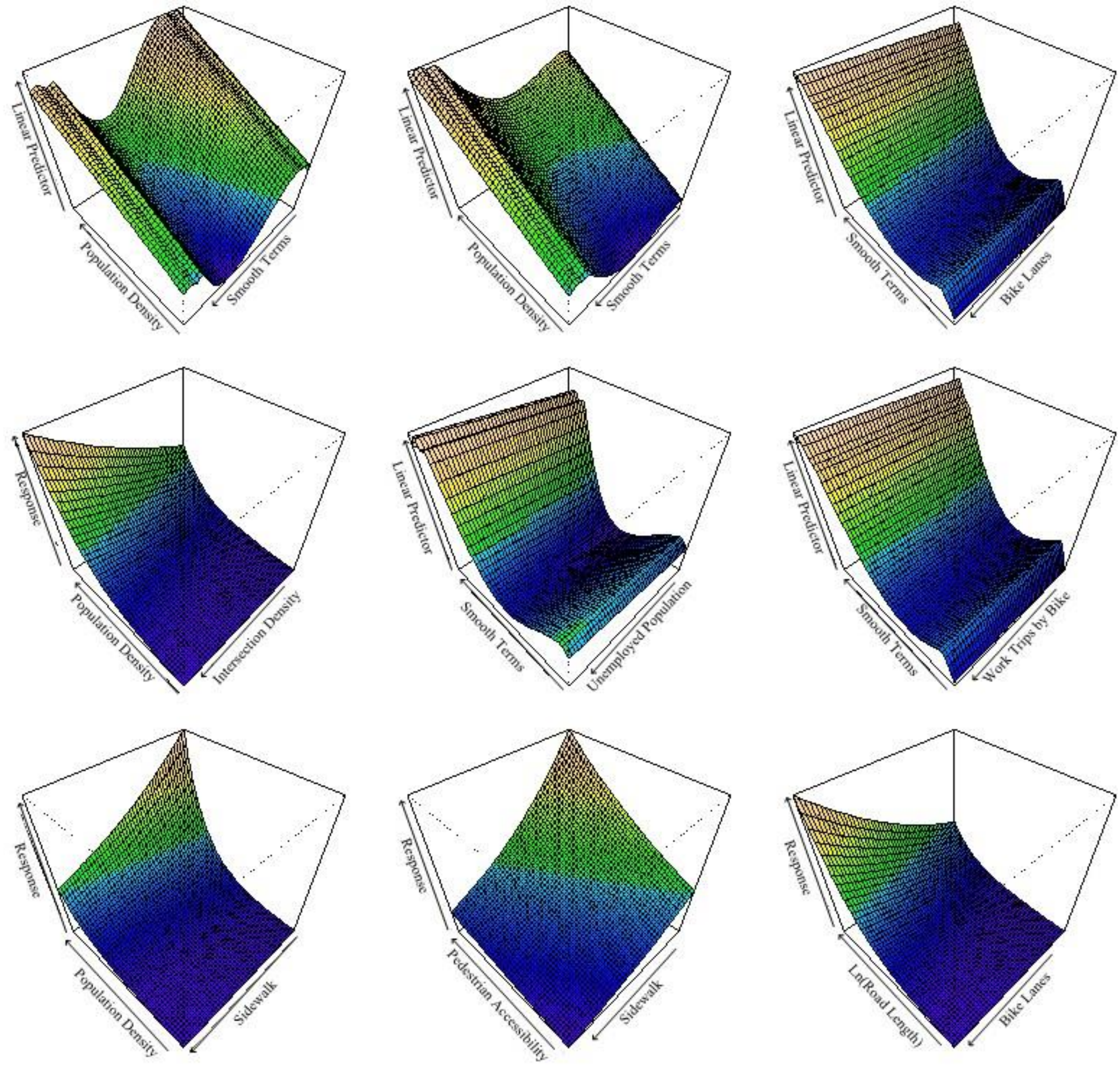
Pedestrian Crashes	**Generalized Additive Model**		
Variables	Coeff.	Std. Err.	P>|z|
ln (DVMT)	0.0493	0.0277	0.075
ln (Pedestrian Trips)	0.2949	0.0363	0.000
Weighted Ped. Accessibility	0.0114	0.0021	0.000
Average Daily Transit Accessibility	0.0045	0.0006	0.000
Destinations within 15-min. Walk	-0.0038	0.0009	0.000
Percentage of Arterials	0.1271	0.0421	0.003
Intersection Density	0.0027	0.0005	0.000
Signalized Intersections (%)	1.1283	0.1909	0.000
Street Connectivity	-0.0834	0.0189	0.000
Network Completeness	0.5031	0.1741	0.004
Intercept	0.6620	0.2694	0.014
Smooth terms	1.003	1.006	0.0011
Deviance explained	43.70%	*Adj. R2*	0.5600
	REML = 2895		

Severe Pedestrian Crashes	**Generalized Additive Model**		
Variable	Coeff.	Std. Err	P>|z|
ln (DVMT)	0.1685	0.0351	0.000
ln (Pedestrian Trips)	0.3491	0.0441	0.000
Weighted Ped. Accessibility	0.0130	0.0027	0.000
Destinations within 15-min Walk	-0.0044	0.0011	0.000
Signalized intersections (%)	1.0195	0.2296	0.000
Intercept	-2.4664	0.3398	0.000
Smooth terms	1.7340	1.9980	0.000
	20.80%	*Adj. R2*	0.281
	REML = 1479		

Table 5 Recommended Model for Total & Severe Bicyclist Crash Estimation

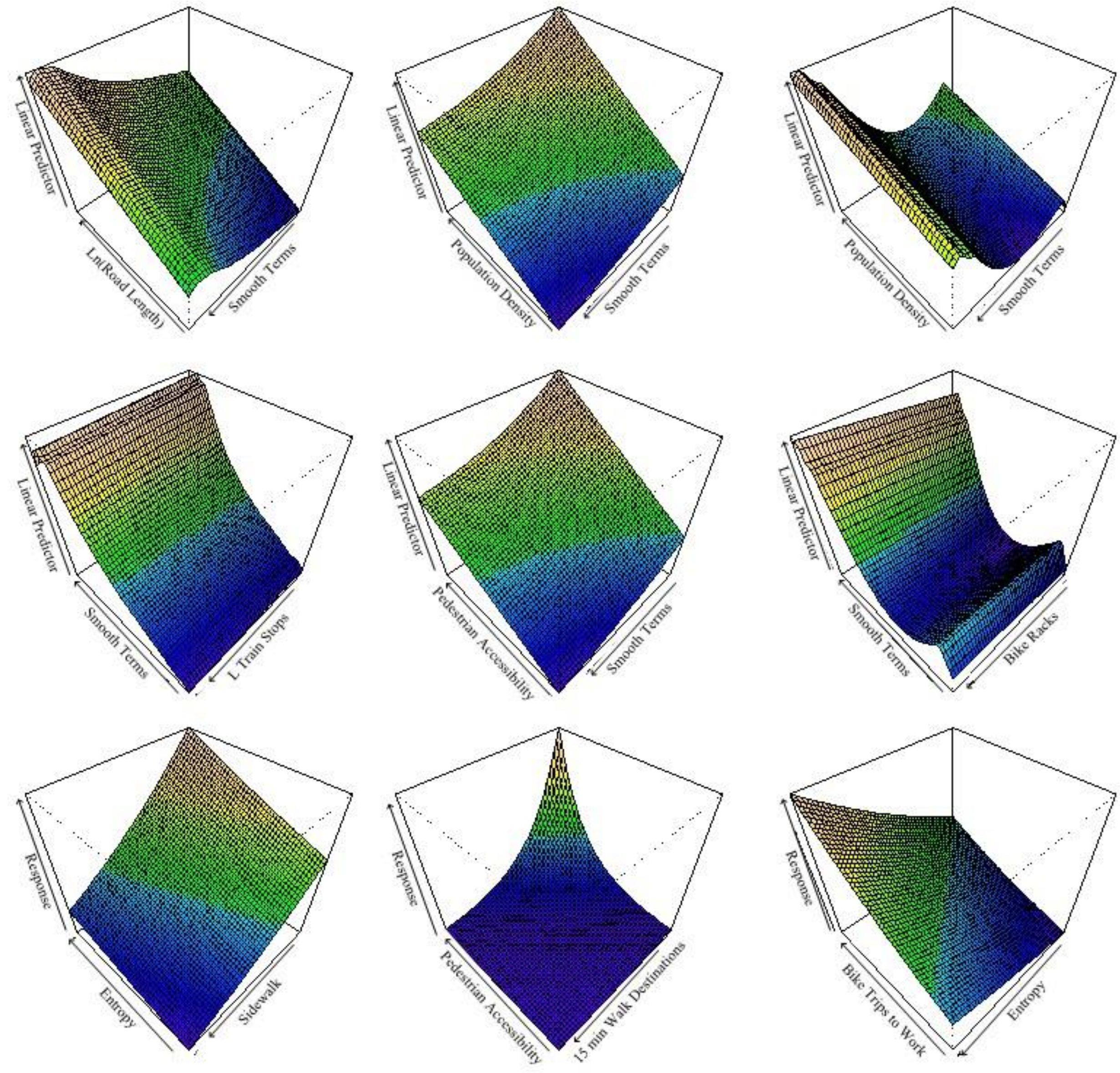
Bicyclist Crashes	**Generalized Additive Model**		
Variables	Coeff.	Std. Err.	P>|z|
ln (DVMT)	0.2183	0.0278	0.000
ln (Bicyclist Trips)	0.4933	0.0449	0.000
Weighted Bicyclist Accessibility	0.0000	0.0000	0.006
Intersection Density	0.0022	0.0004	0.000
L Train Line (miles)	-0.1412	0.0743	0.057
Bike Lanes (miles)	0.2650	0.0365	0.000
Central Business District	-0.4601	0.1656	0.005
Intercept	-1.0690	0.2814	0.000
Smooth terms	7.351	7.955	0.000
Deviance explained	62.20%	*Adj. R2*	0.578
	REML = 2297		

Severe Bicyclist Crashes	**Generalized Additive Model**		
Variable	Coeff.	Std. Err	P>|z|
ln (DVMT)	0.2199	0.0552	0.000
ln (Bicyclist Trips)	0.2684	0.0639	0.000
Weighted Bike Accessibility	0.0001	0.0000	0.046
Bike Lanes (miles)	0.2932	0.0605	0.000
Intercept	-3.1670	0.5374	0.000
Smooth terms	6.5230	7.4470	0.000
	31.70%	*Adj. R2*	0.35
	REML = 874		



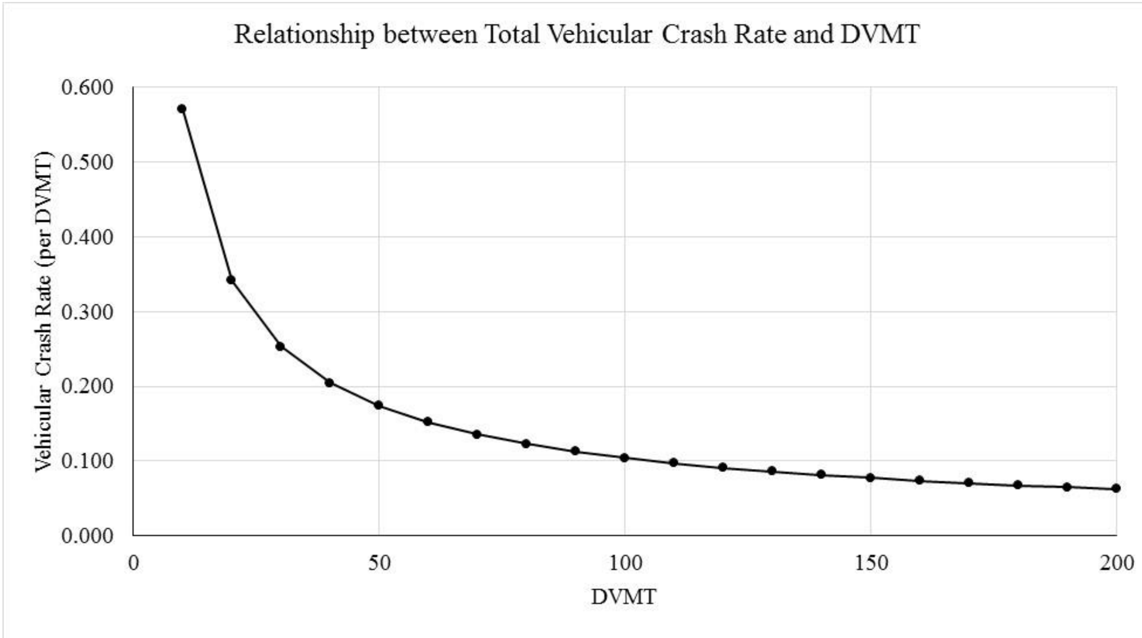
a) Vehicle-only crashes b) Pedestrian crashes c) Bicyclist crashes

Figure 2 Visualization of GAM for some of the Statistically Significant Explanatory Variables in Total Crash Models



a) Vehicle-only KA crashes b) Pedestrian KA crashes c) Bicyclist KA crashes

Figure 3 Visualization of GAM for some of the Statistically Significant Explanatory Variables in Severe Crash Models



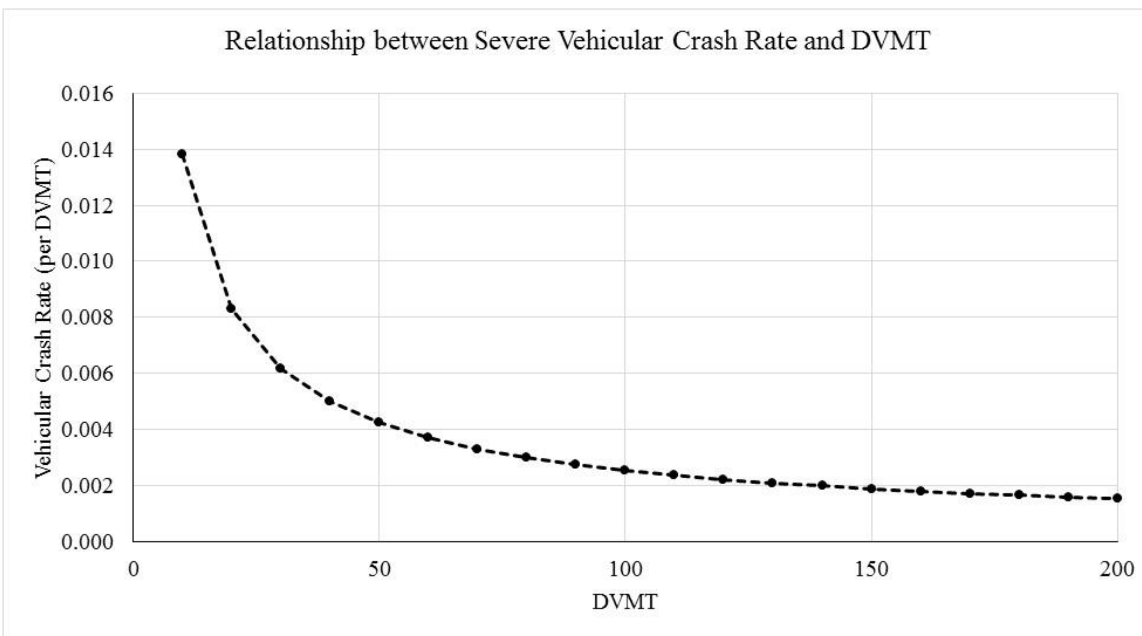
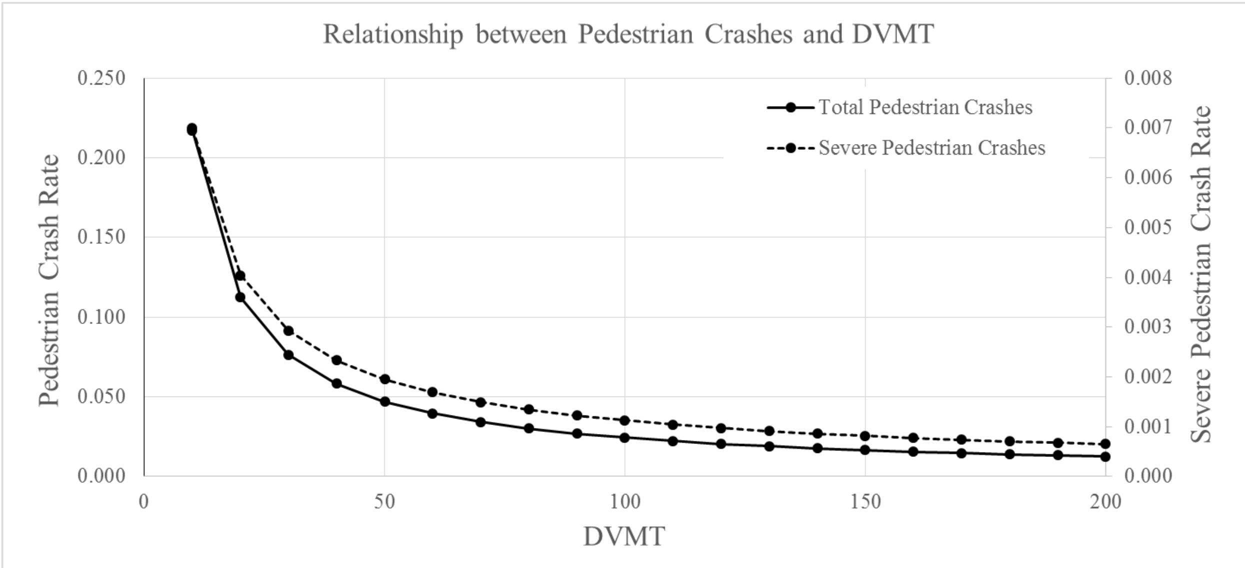


Figure 4 The "Safety in Numbers" Effect for Private Vehicle Users



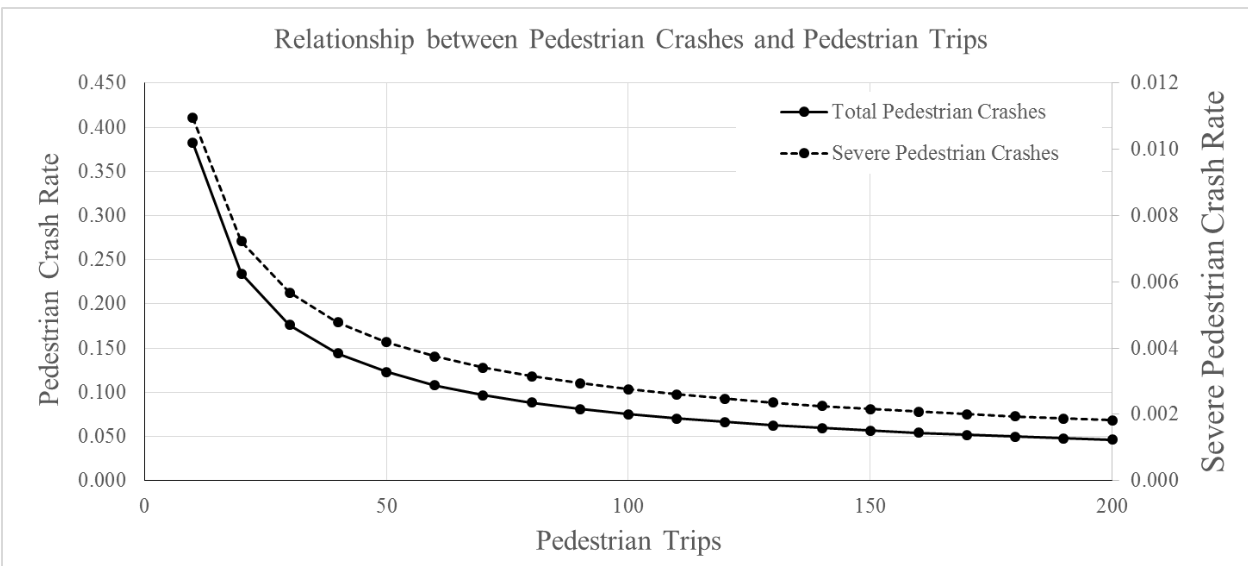
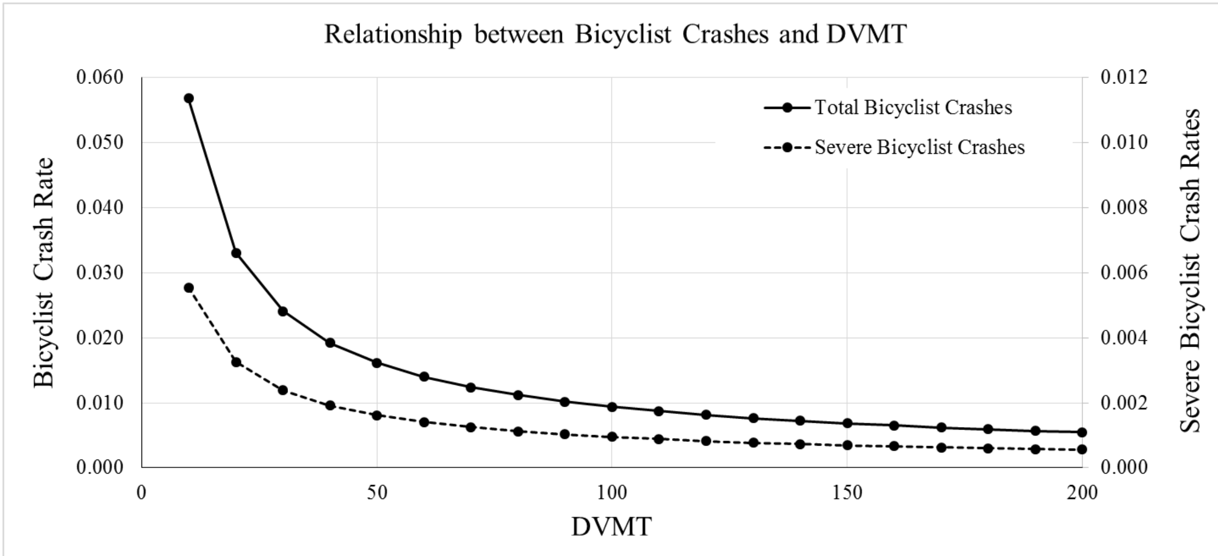


Figure 5 The "Safety in Numbers" Effect for Pedestrian Users



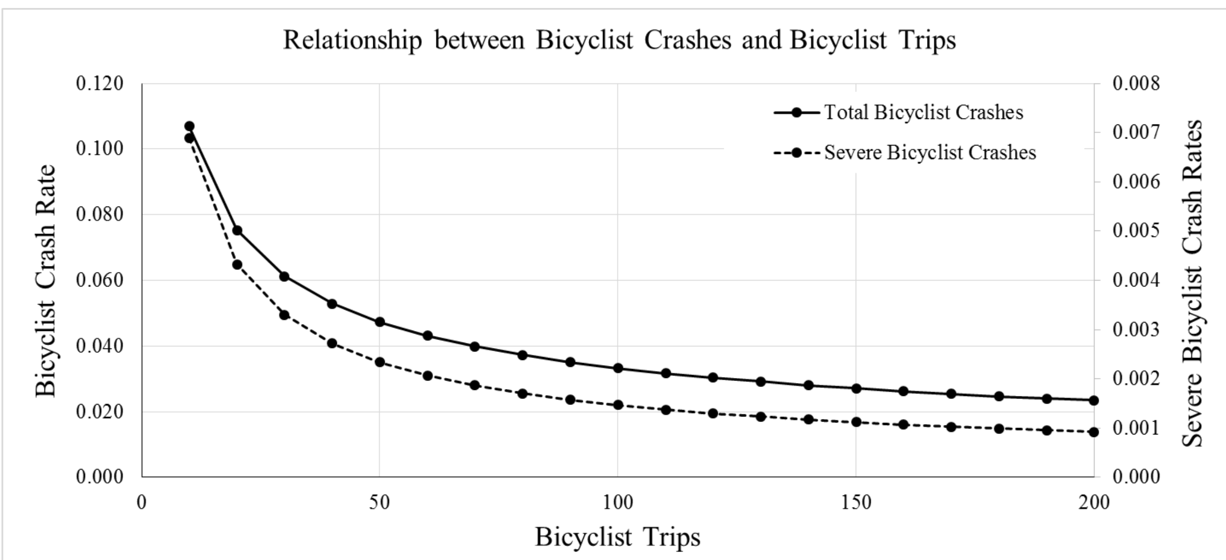


Figure 6 The “Safety in Numbers” Effect for Bicyclist Users

5 Conclusions and Recommendations

There is a general understanding that improved multimodal transportation systems may lead towards resolving multiple long-term issues related to sustainability and efficiency of travel in urban environments. There is also a need to continue to explore the effects of investments in multimodal infrastructure on safety for different road user types, particularly vulnerable road users (pedestrians and cyclists). Walking and cycling is essential in developing sustainable transportation solutions. These modes require less space in terms of right-of-way and parking, increase active travel behavior prevents obesity, contributes to reducing emissions associated with motor vehicle usage, and may contribute to reducing person-level travel delay as a significant portion of these users tend to combine walking or cycling with the use of public transportation.

The research presented in this paper is focused on exploring the safety in numbers effects for pedestrians and cyclists in urban environments at a macroscopic level. The goal of this research was to provide a detailed set of indicators that could represent exposure as well as exposure surrogates in multimodal transportation systems, and relate these indicators to the expected number of crashes for vehicles, pedestrians, and bicyclists.

A safety in numbers effect is found in all estimated SASM models for all crash types. This indicates that within the spatial units of analysis used, an increase in exposure would not be associated with a proportional increase in the expected number of crashes. This applies to all the three groups examined: cars, pedestrian and cyclists. These findings are consistent with previous research on the safety in numbers effect. In addition to exploring the relationships between the available measures of exposure and safety, proxies for exposure such as accessibility to destinations for various users are also explored in this paper. Generally, cumulative accessibility has a stronger association with the expected number of crashes than weighted accessibility, and is negatively associated with pedestrian crashes. This indicates that there is a need to further explore the association between the measures of accessibility and the available indicators of multimodal exposure, and see if perhaps similar surrogates can be used to better represent multimodal exposure in complex urban environments. The finding that the relationship between exposure (and surrogates for exposure) and crashes is mode-dependent should be further explored in the future research efforts.

The results of this research show that crash rate per unit of exposure decreases with the increase of the exposure in the defined spatial unit of analysis. A finding such as this could be used to support long range transportation plans for what is considered to be sustainable infrastructure, while partially addressing concerns about negative safety effects, particularly for pedestrians and cyclists. Using a set of exposure indicators that includes the surrogate measures to properly represent the complexity of urban transportation and land use in major cities, as well as access to destinations that in this case served as a proxy for activity concentration, could strengthen efforts to justify area-wide transportation improvements in both short-term and long-term transportation planning. The SASM framework based on GAM approach has rarely been used in previous safety studies, particularly at a macroscopic level. A comprehensive overview and statistical safety modeling in the existing research, rarely captures more than one or two user types, which could be considered as another contribution of this study. There is an obvious need to continue to improve the way exposure data are collected, particularly for non-motorized users, as this would help to improve the safety estimates in all types of environments. This however should not serve as an impediment to further considerations of investments in multimodal infrastructure, and further development of safety evaluation methods in a multimodal context.

References

R. Bhatia and M. Wier. 2011. "Safety in Numbers" re-examined: Can we make valid or practical inferences from available evidence? *Accident Analysis and Prevention*, 43, 235-240.

Castro, M., Paleti, R., Bhat, C. R. 2013. A spatial generalized ordered response model to examine highway crash injury severity. *Accident Analysis and Prevention, 52*, 188-203.

Elvik, R. 2016. Safety-in-numbers: Estimates based on a sample of pedestrian crossings in Norway. *Accident Analysis and Prevention*, 91, 175-182.

Elvik, R., Bjørnskau, T. 2017.Safety-in-numbers: A systematic review and meta-analysis of evidence. *Safety Science*, 92, 274-282.

Flask, T., Schneider, W. 2013. A Bayesian analysis of multi-level spatial correlation in single vehicle motorcycle crashes in Ohio. *Safety Science, 53*, 1-10.

Fyhri, A., Sundfør, H. B., Bjørnskau, T., Laureshyn, A. 2016. Safety in numbers for cyclists—conclusions from a multidisciplinary study of seasonal change in interplay and conflicts, *Accident Analysis and Prevention*, forthcoming.

Geyer, J., Raford,N., Pham, T., Ragland, D. 2006. Safety in Numbers: Data from Oakland, California. *Transportation Research Record: Journal of the Transportation Research Board,* 1982, 150-154.

Hastie, T., Tibshirani, R. 1990. *Generalized additive models*. London: Chapman and Hall.

Hauer, E. 1982. Traffic conflicts and exposure. *Accident Analysis and Prevention,* 14, 359-364.

Johnson, M.,, Oxley, J., Newstead, S.,, Charlton. J. 2014. Safety in numbers? Investigating Australian driver behaviour, knowledge and attitudes towards cyclists. *Accident Analysis and Prevention*, 70, 148-154.

Moeinaddini, M., Asadi-Shekari, Z., Shah, M. Z. 2014. The relationship between urban street networks and the number of transport fatalities at the city level. *Safety Science, 62*, 114-120.

Noland, R. B., Quddus, M. A. 2004. A spatially disaggregate analysis of road casualties in England. *Accident Analysis & Prevention,* 36, 973-984.

Siddiqui, C., Abdel-Aty, M., Choi, K. 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, 45, 382-391.

Tasic, I., Porter, R. J., Brewer, S. 2016. Applications of generalized additive models and Bayesian hierarchical models for areal safety analysis of urban multimodal transportation systems. *Transportation Research Record, Journal of the Transportation Research Board*, 2601,99-109.

Tasic, I., Musunuru, A., Porter, R. J. 2014. Quantifying accessibility of non-motorized transportation modes in recreational areas: Case study of Mill Creek Canyon, Utah. In *The Journal of Parks and Recreation Administration*, 32(3), 49-73.

Tasic, I., Zhou, X., Zlatkovic, M. 2014. Use of spatiotemporal constraints to quantify transit accessibility. *Transportation Research Record: Journal of the Transportation Research Board*, 2417, 130-138.

Ukkusuri, S., Miranda-Moreno, L. F., Ramadurai, G., Isa-Tavarez, J. 2012. The role of built environment on pedestrian crash frequency. *Safety Science*, 50, 1141-1151.

Wang, Y., Kockelman, K. M, 2013. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis and Prevention*, 60, 71-84.

Wood, S. N.. 2006. *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman & Hall/CRC.