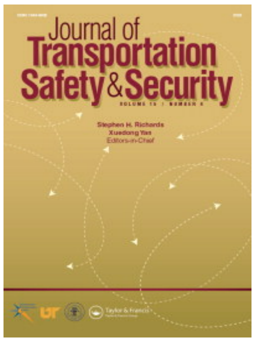

Taylor & Francis
Taylor & Francis Group

Journal of Transportation Safety & Security

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utss20

Multiscale spatial analysis of macro-level determinants of bicycle crash frequencies in the Phoenix metro region

Steven R. Gehrke, Brendan J. Russo, Michael P. Huff & Edward J. Smaglik

To cite this article: Steven R. Gehrke, Brendan J. Russo, Michael P. Huff & Edward J. Smaglik (2023): Multiscale spatial analysis of macro-level determinants of bicycle crash frequencies in the Phoenix metro region, Journal of Transportation Safety & Security, DOI: 10.1080/19439962.2023.2201176

To link to this article: https://doi.org/10.1080/19439962.2023.2201176

Published online: 19 Apr 2023.

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
https://doi.org/10.1080/19439962.2023.2201176

Multiscale spatial analysis of macro-level determinants of bicycle crash frequencies in the Phoenix metro region

Steven R. Gehrke[a], Brendan J. Russo[b], Michael P. Huff[a], and Edward J. Smaglik[b]

[a]Geography, Planning, and Recreation, Northern Arizona University, Flagstaff, Arizona, USA; [b]Civil Engineering, Construction Management, and Environmental Engineering, Northern Arizona University, Flagstaff, Arizona, USA

ABSTRACT

The realization of the many benefits of bicycling will not be achieved in American regions until safer bike infrastructure and bicycling conditions are presented to a more general population. The Phoenix region—one of the nation's most populous—has sought policies and programs to increase bicycling rates. Yet, the region continues to have a small mode share, underscoring a need to motivate population-level bicycling adoption. This study examines 2015–2019 bicycle-vehicle crash data to identify those macro-level factors associated with bicycle-vehicle crashes and a subset of crashes where a serious injury or fatality occurred. Specifically, the effects of a robust set of socioeconomic and built environment factors, measured at three hexagon spatial extents, in negative binomial and spatial Durbin models were estimated for the two crash outcomes. Results show denser zones with a traditional network design experienced more bicyclist-involved crashes, as did zones with a higher percentage of low-income households and working-age adults. Findings, which also found spatial clustering of total and severe bicyclist-involved crashes, suggest that the targeted provision of safer bike infrastructure and a more complete network in zones exhibiting certain macro-level attributes holds promise in creating bike-friendly conditions that generate more utilitarian and recreational bicycling throughout the region.

KEYWORDS
bicycling; bicycle safety; bicycle crashes; spatial Durban model; built environment

1. Background

Bicycling as a mode choice provides several potential benefits, including reduced congestion, lower transport-related emissions, and improved health (Brown et al., 2016, de Hartog et al., 2010). Also, after the outbreak of the COVID-19 pandemic, bicycling provided a socially distanced alternative to shared travel modes as well as a desirable recreational activity. However,

CONTACT Steven R. Gehrke steven.gehrke@nau.edu Geography, Planning, and Recreation, Northern Arizona University, Flagstaff, Arizona, USA.

bicyclists are vulnerable roadway users who are not protected by an enclosed vehicle compartment and, as such, are more likely than motor vehicle occupants to sustain injuries in the event of a crash (NTSB, 2019). In 2019, 871 bicyclists were killed in traffic crashes across the United States (US) and 49,000 bicyclists were injured (representing a 4.3% increase compared to 2018) (NHTSA, 2020), highlighting bicyclist safety as an immediate and important public health issue.

In response, several studies have recently analyzed factors associated with bicyclist injury severity in crashes involving vehicles (Bahrololoom et al., 2020; Chandia-Poblete et al., 2021; Lin & Fan, 2021; Robartes & Chen, 2017; Zhu, 2021). Although studies focused on the factors associated with crash severity are needed, their preventive utility is limited because the investigation of injury outcomes is only after an observed crash. To stop bicycle-vehicle crashes from occurring in the first place, it is imperative to examine and help understand those factors associated with the frequency or occurrence of these crashes to effectively plan preventive countermeasures. While it is common to analyze motor vehicle crashes at the segment or intersection level, bicyclist-involved crashes occur much less frequently, and it may be difficult to obtain meaningful results at these micro-levels. Accordingly, bicyclist-involved crashes are often investigated at a macro level (e.g., traffic analysis zone [TAZ], US Census block group), which allows for consideration of potential associations with zonal characteristics describing an area's sociodemographic and economic composition, overall street network, and built environment.

Past studies that have analyzed bicycle-vehicle crashes at a macro-level geography include Osama and Sayed (2017), who analyzed factors associated with bicycle-vehicle crashes in Vancouver, Canada, using TAZ geographies with generalized linear regression and full Bayesian techniques. In finding that several zonal attributes were associated with bicycle-vehicle crash frequency, including exposure-related variables (e.g., bike and vehicle kilometers traveled) and built environment variables such as traffic signals, transit stops, land use, and roadway type, the authors concluded that there was significant spatial correlation and an importance of accounting for such effects (Osama & Sayed, 2017). Guo et al. (2018) also developed macro-level bicyclist crash frequency models at the TAZ level in Vancouver (1,700 total crashes). In this study, the authors compared Poisson lognormal model (PLN), random intercepts PLN, random parameters PLN, and spatial PLN model estimates, discovering that the spatial model provided the best fit and that several exposure, roadway network, and built environment variables were associated with TAZ-level bicyclist crashes (Guo et al., 2018).

Cai et al. (2016) also investigated factors associated with bicycle-vehicle crashes at the TAZ level using data from Florida. By estimating a series of negative binomial (NB), zero-inflated NB, and Hurdle NB models, the authors found several traffic, roadway, and socioeconomic characteristics were associated with bicycle-vehicle crashes and declared the importance of accounting for their spatial spillover effects (Cai et al., 2016). Amoh-Gyimah et al. (2016) estimated NB, conditional autoregressive NB, and random parameters NB (RPNB) models to investigate spatial factors associated with total, minor injury, and serious injury bicycle-involved crash frequencies at Statistical Area level 2 (spatial units with an average population of 10,000 persons) in Melbourne, Australia. The authors determined that the RPNB model performed best and that several exposure-, sociodemographic- and land use–related variables were associated with bicycle-involved crash frequency, though only four predictors were significant in respect to serious injury crashes (Amoh-Gyimah et al., 2016). Ji et al. (2021) conducted a similar analysis at Statistical Area level 2 units in greater Melbourne, Australia, using a semi-parametric geographically weighted Poisson regression model; finding exposure and street network variables as well as other spatial factors were associated with bicycle crashes.

Chen et al. (2018) investigated the impacts of TAZ-level built environment measures on bicycle-vehicle crash frequency in Beijing, China, through estimation of a Poisson lognormal random effects model (PLREM). The authors concluded that exposure, roadway network, and built environment characteristics were all associated with bicycle-vehicle crash counts. Chen (2015) also employed a PLREM to investigate bicycle-vehicle crashes at the TAZ level in Seattle, Washington, and found exposure measures in addition to land use and road network characteristics to be positively associated with bicycle-vehicle crashes, and that TAZ-based crashes were spatially correlated. There have been other macro-level analyses of bicycle crashes, including one study conducted in Greater London that aggregated spatial information to a Super Output Area (Ding et al., 2020), one conducted with Census block groups in Florida that included separate analyses of both total and severe/fatal injury bicycle crashes (Saha et al, 2018), one conducted with Census block groups that investigated the count of minor, severe, and fatal injury bicycle crashes in Austin, Texas (Sener et al., 2021), and another study considering factors at the TAZ-level in Hillsborough County, Florida (Wang et al., 2017).

While the reviewed studies demonstrate that research investigating bicycle-vehicle crashes at the macro level exists, an extensive review by Merlin et al. (2020) pronounced a need for further investigation of macro-level factors in bicyclist-related crash analyses due to inconsistencies in previous results. This study answers that call by presenting a macro-level analysis of sociodemographic and economic, built environment, and street network factors—measured at three different scales—in a large, southwestern US metro region. The depth of predictors tested here expands on those explored in past studies and includes important variables such as level of traffic stress classification, presence of alcohol-selling establishments, and several street network connectivity measures, as well as socioeconomic context variables related to the representation of residents and workers in different wage categories. Additionally, the measurement of these variables at three spatial extents (one-half-, one-, and two-mile diameter hexagons), opposed to asymmetric geographic scales typically adopted because of user convenience or data availability, permits insights into the consistency and relative magnitude of a macro-level determinants of bicyclist-involved crashes. More specifically, the adoption of hexagon sampling zones offers a desirable property that the centroids of all neighboring hexagons have identical Euclidean distances between them, while the choice of multiple spatial extents allows an investigation of geographic scale sensitivity related to the modifiable areal unit problem (Fotheringham & Wong, 1991). Finally, this study accounts for potential spatial correlation through the estimation of spatial Durbin models. In all, study findings are intended to offer evidence for transportation planners, engineers, and decision makers aiming to implement policies and countermeasures that improve bicyclist safety.

2. Methods

2.1. Study area

For this study, the Phoenix metro region refers to the Census-defined urbanized area of Phoenix and Mesa positioned within Maricopa County, Arizona. Per 2015–2019 American Community Survey estimates, this definition of the Phoenix region houses over 3.92 million residents, with Phoenix accounting for over two-fifths of the region's population (1.63 million residents), followed by the City of Mesa (499,720 residents) bordering to the east. The next four largest cities in the region (Chandler, Scottsdale, Glendale, and Gilbert) are comparable in population, ranging from 252,692 to 243,254 residents. Tempe, home to Arizona State University's main campus, is the seventh largest city in the region with 187,454 residents and the densest, with 7.29 residents per acre.

The percent of workers older than 16 years old in the Phoenix metro region who commute by automobile (alone or pooled) generally exceeds the national average (85.3%), with 87.2% of both Phoenix and Mesa workers commuting by car, truck, or van. In Phoenix and the next five largest cities within the region, the percent of workers who commute by bicycle is less than one, while the bicycle commute mode share in Tempe is 3.17%.

2.2. Data sources

The Arizona statewide police-reported crash data set analyzed in this study was obtained for the years 2015 through 2019 from the Arizona Department of Transportation (ADOT). This five-year data source was chosen to increase the statistical value of data on crashes, which are relatively rare events; represent a multiyear period that was not impacted by travel disruptions associated with the COVID-19 outbreak, which was declared a pandemic in March 2020; and temporally align with publicly available macro-level data on socioeconomic context. These data were then filtered to include only those crashes in the Phoenix metro region that involved at least one bicyclist. This resulted in a total of 4,875 bicyclist-involved crashes identified for inclusion in this study. The injury severity of each crash-involved person is reported as one of five discrete categories per the *Arizona Crash Report Forms Manual* (ADOT, 2017):

- K-Injury (fatal injury): Any injury that results in death within a 30-day time period after the crash occurred.
- A-Injury (suspected serious Injury): Any injury other than fatal which results in one or more of the following: Severe laceration resulting in exposure of underlying tissues/muscle/organs or resulting in significant loss of blood; broken or distorted extremity (arm or leg); crush injuries; suspected skull, chest, or abdominal injury other than bruises or minor lacerations; significant burns; unconsciousness when taken from the crash scene; or paralysis.
- B-Injury (suspected minor injury): A minor injury is any injury that is evident at the scene of the crash, other than fatal or serious injuries. Examples include lump on the head, abrasions, bruises, minor lacerations (cuts on the skin surface with minimal bleeding and no exposure of deeper tissue/muscle).
- C-Injury (possible injury): An injury reported or claimed which is not a fatal, suspected serious, or suspected minor injury. Examples include momentary loss of consciousness, claim of injury, limping, or complaint of pain or nausea. Possible injuries are those which are reported by the person or are indicated by his/her behavior, but no wounds or injuries are readily evident.
- O-No Injury (property damage only): No apparent injury is a situation where there is no reason to believe that the person received any bodily harm from the motor vehicle crash. There is no physical evidence of injury and the person does not report any change in normal function.

The severity of each crash was defined based on the most severely injured crash-involved person (almost always the bicyclist). Of the 4,875 total bicyclist-involved crashes included in the analysis, 281 (5.8%) were no-injury crashes, 1,570 (32.2%) were C-level injury crashes, 2,354 (48.3%) were B-level injury crashes, 589 (12.1%) were A-level injury crashes, and 81 (1.7%) were fatal-injury crashes.

These bicycle crash data and macro-level metrics of socioeconomic context and built environment were summarized to three systems of hexagons with one-half-, one-, and two-mile edge-to-edge diameters that were cast across the study area. Socioeconomic context metrics were calculated for each hexagon using Census tract data from the 2015–2019 American Community Survey's five-year estimates and an area-based apportionment process. These macro-level measures describing various attributes of the residents in each hexagon, including person-level characteristics about sex, age, education, race/ethnicity, and household attributes related to income, tenure, and vehicle ownership. Additional measures about the share of residents in a hexagon employed in low-, mid-, or high-wage occupations were obtained by using the 2018 Longitudinal Employer-Household Dynamics data set.

Common built environment measures of population and employment density, which may also serve as proxy variables for exposure, were also produced for each hexagon using the two aforementioned data sets, as were density measures of activity (sum of persons and jobs) and workplaces in low-, mid-, or high-wage categories and the land use mix measure of jobs-population ratio. A set of street network and design characteristics were computed for every hexagon using OpenStreetMap (OSM) data. Intersection density, beta index, and connected node ratio (Gehrke & Welch, 2019) indicate an area's overall street network connectivity and were created using information on street nodes and links, whereas the percentage of roads in a hexagon categorized as primary, secondary, tertiary, and residential was calculated using the OSM highway tag. In the absence of bicyclist trip data, the sum of these street network lengths was calculated as another proxy measure of exposure because each of these facilities could be used by a potential bicyclist. To help identify any association between bicycle crash frequency and bicycling infrastructure, the percentage of bike facilities in a hexagon defined as having either a low or high level of traffic stress by People for Bikes (2021) Bicycle Network Analysis method was also measured. Finally, the percent of a hexagon's area within a one-half-mile buffer of a Valley Metro light rail transit station was calculated using OSM data.

2.3. Analytic approach

The count of total bicycle crashes and fatal or serious injury-related events, as well as the socioeconomic context and built environment characteristics, which were measured at three hexagonal spatial extents, were analyzed by estimating three aspatial negative binomial (NB) models. NB model specifications were chosen to assess the macro-level factors associated with total and KA-only (K-Injury and A-Injury) bicycle crash frequency in the Phoenix metro region, provided the nonnegative integer and likely over-dispersion nature of these two dependent variables. The relaxation of the equi-dispersion assumption in a Poisson count model that indicates equality in the conditional mean and variance functions is a major advantage of the NB model, which will default to the former model structure if over-dispersion is not present. The structure for this first set of NB models is presented as:

$$\lambda_i = \exp(\beta x_i + \varepsilon_i), \tag{1}$$

where, x_i is a set of various socioeconomic and built environment predictors operationalized at hexagon zone i and ε_i is a Gamma-distributed error term with a mean equal to one and a variance of α^2. This error term permits the variance to differ from the conditional mean:

$$var[y_i] = E[y_i] + \alpha E[y_i]^2 \tag{2}$$

When analyzing crash count data aggregated to any macro-level spatial unit, unobserved spatial correlations that could bias the model results may appear. A commonly accepted approach to assess whether spatial autocorrelation exists in the data set is to estimate the global Moran's I statistic (Anselin, 2010):

$$I = \frac{\sum_i \sum_j w_{ij} z_i \times z_j / S_0}{\sum_i z_i^2 / n}, \tag{3}$$

where, w_{ij} represents the elements of a spatial weight matrix, $S_0 = \sum_i \sum_j w_{ij}$ is the sum of the weights, and n is the number of hexagons (observations). This formulation shows Moran's I to be a cross-product statistic between a variable and its spatial lag, with the variable expressed in terms of deviations from the mean. In testing a null hypothesis that spatial randomness exists, a positive spatial autocorrelation was found to be present within the two dependent variables: count of total bicycle-related crashes ($I = 0.574$, $p < 0.001$) and count of bicycle-related crashes resulting in a fatality or serious injury ($I = 0.499$, $p < 0.001$). A significant finding for each outcome indicating this unobserved correlation should be accounted for with a spatially lagged extension of the above NB model specifications. Additional examination using likelihood ratio tests of the aspatial NB models and their complementary specifications with the addition of spatially lagged terms supported the estimation of a spatial Durbin modeling approach.

Spatial Durbin models (SDM), which specify a spatially lagged outcome and spatially lagged independent variables, produce unbiased coefficient estimates by considering the spatial dependence existing in the endogenous and exogenous relationships of the hexagons and their neighboring units (LeSage & Pace, 2009). The three SDMs (one for each spatial extent) in this study takes the following general structure:

$$y = \rho W_y + \beta + \Theta W x_i + \alpha l_n + \varepsilon_i,$$

where, ρ is the spatial autocorrelation coefficient, W is the spatial weight matrix, x_i is the socioeconomic and built environment predictors of hexagon i, and l_n represents an $n \times 1$ vector of ones. The coefficient estimates are denoted by α, β, and Θ, with ε_i as the error term. Aside from the spatial lag of the outcome variable (Wy), the SDM also presents spatially lagged independent variables (Wx_i) that adopted the queen-contiguous spatial weight matrix in which any immediately neighboring hexagon was given a value of one.

The result of this analytic approach was the initial estimation of two sets of aspatial NB models, followed by the estimation of two sets of SDMs. NB models were specified using a backward elimination process, where all predictors in the final specification were marginally significant ($p < 0.10$), while the SDMs of total and KA-only crash frequency contains the spatial lag of the modeled crash outcome and the spatial lag terms for each marginally significant predictor in the aspatial NB model specification.

3. Results

3.1. Descriptive overview of bicycle crashes and spatial factors

Figure 1 provides a visualization of all reported bicyclist-involved crashes in the Phoenix metro region from 2015 to 2019 at the one-mile hexagon extent. Examining this map shows a concentration of bicycle-vehicle crashes has occurred in midtown Phoenix, as perhaps expected given its clustering of activity centers, but that some measure of bicyclist-involved crashes can be found across the study area. Other clusters of hexagons with at least 10 bicycle-vehicle crashes per year over the study period are located in Mesa and Tempe. Akin to Phoenix, the pattern of crashes for Mesa is likely related to the population of the municipality, its larger aggregate pool of prospective bicyclists, and potentially unsafe bicycling conditions given that the community's bicycling mode share is minimal. However, in Tempe, where the bicycling mode share is relatively high compared to other places in the region, a hexagon exists near Arizona State University

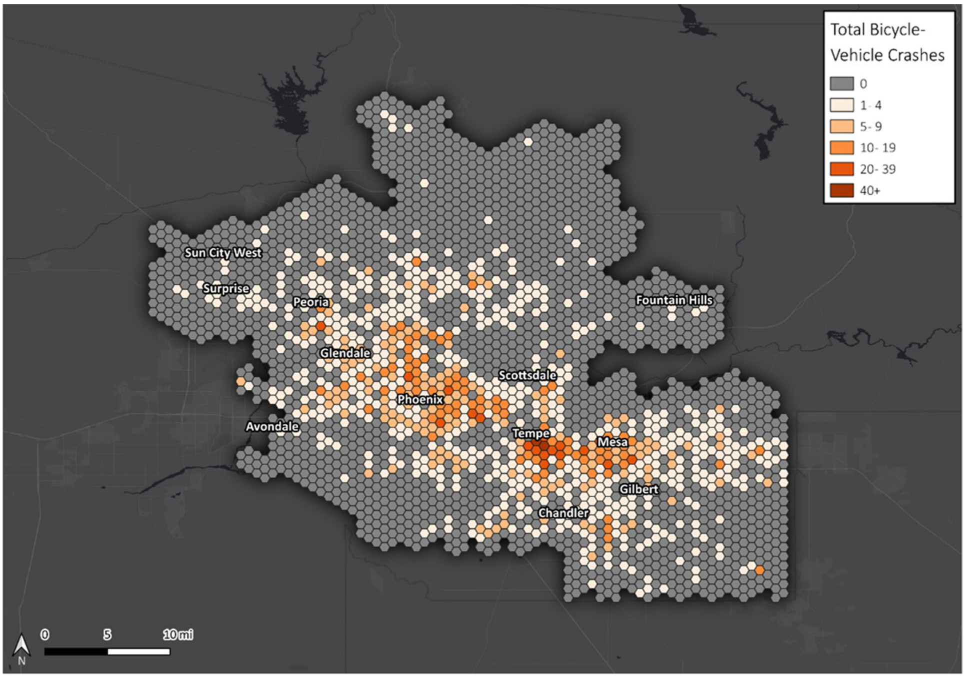

Figure 1. Bicyclist-involved crashes with motorists in the Phoenix metro region (2015–2019) at one-mile hexagons.

with the highest frequency of bicyclist-involved crashes (over 30 bicycle-vehicle crashes per year). This finding suggests the need for improved bike infrastructure and safer bicycling conditions, as other circumstances seem to support the adoption of bicycling for travel in the area.

As for those bicycle-vehicle crashes with a more severe health outcome, Figure 2 shows the spatial pattern of bicyclist-involved crashes within the region—aggregated to a one-mile hexagon zonal system—that resulted in a fatality or serious injury over the five-year study period. The overall spatial distribution of these events reveals that most hexagons do not have a recorded bicycle-vehicle crash with a K- or A-level injury, but that zones with a higher count of this more severe crash outcome are similarly found in Phoenix, Mesa, and Tempe. Of note, most hexagons with a greater frequency of crashes with a fatal or serious injury outcome are also the zones with a higher count of total bicycle-vehicle crashes, with some exceptions.

Table 1 further describes these two bicycle-vehicle crash outcomes as well as the socioeconomic context and built environment variables tested in the aspatial and spatial models. On average, 1.82 bicycle-vehicle crashes were observed per one-mile hexagon in the Phoenix metro region from 2015–2019, with a mean of 0.25 K- or A-level injuries occurring in this zonal geography. The average count of total bicycle-vehicle crashes ranges from 0.47 at the one-half-mile extent to 5.47 crashes at the two-mile spatial

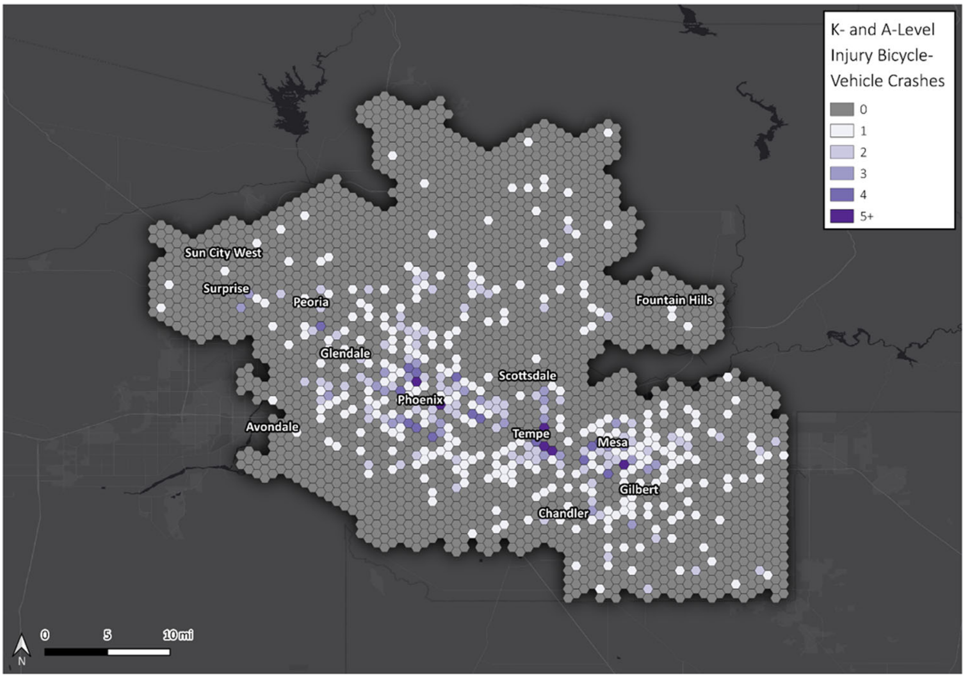

Figure 2. Bicyclist-involved crashes with motorists that resulted in a death or serious injury in the Phoenix metro region (2015–2019) at one-mile hexagons.

extent, while more severe bicyclist-involved crash totals range from 0.06 when averaged at a one-half-mile extent to 1.21 crashes at the largest spatial extent. As for regional bike infrastructure, the share of facilities in most hexagons of any spatial extent were classified as low stress, suggesting that an opportunity exists for growth in overall bicycle mode share under current conditions; however, with about one quarter of bike facilities in each spatial extent classified as high stress, the bike network in most places is likely to have significant safety gaps.

3.2. Spatial factors associated with total bicycle crashes

Analyzing the socioeconomic context and built environment of one-half-, one-, and two-mile hexagons across the Phoenix metro region, Table 2 shows the estimation results for three NB models of total bicycle crash frequency. In terms of macro-level socioeconomic characteristics, an increase in the share of older adults in a hexagon was associated with a decrease in observed bicycle crashes, highlighting that bicyclist exposure is likely lowest in these communities. In turn, an increase in the percentage of residents identified as White, adults with an advanced college degree, and those households reporting an annual income below $75,000 in any of the hexagon spatial extents was associated with a higher frequency of bicyclist-

Table 1. Descriptive statistics for hexagon-level dependent and independent variables.

Hexagon Diameter:	One-half mile		One mile		Two miles	
Variable	Mean	*SD*	Mean	*SD*	Mean	*SD*
Bicyclist-Related Crash Frequency						
Count of total bicycle crashes	0.47	1.43	1.82	4.03	5.47	10.22
Count of fatal (K) or serious injury (A) crashes	0.06	0.29	0.25	0.66	1.21	2.09
Socioeconomic Context						
Sex: Male	0.49	0.07	0.49	0.05	0.49	0.02
Sex: Female	0.50	0.07	0.50	0.05	0.51	0.02
Age: 18-24 years old	0.20	0.11	0.21	0.10	0.23	0.08
Age: 25-34 years old	0.21	0.12	0.21	0.10	0.21	0.10
Age: 35-44 years old	0.13	0.10	0.12	0.04	0.13	0.04
Age: 45-64 years old	0.27	0.11	0.26	0.07	0.26	0.06
Age: 65 years old or more	0.17	0.16	0.18	0.05	0.18	0.15
Education: High school or less	0.61	0.21	0.62	0.19	0.61	0.17
Education: Bachelors or some college	0.23	0.11	0.23	0.11	0.24	0.10
Education: Masters of PhD	0.14	0.09	0.14	0.09	0.15	0.08
Race/Ethnicity: Asian	0.04	0.04	0.04	0.04	0.04	0.04
Race/Ethnicity: Black/African American	0.04	0.05	0.04	0.05	0.05	0.04
Race/Ethnicity: Hispanic/Latinx	0.17	0.19	0.17	0.16	0.17	0.15
Race/Ethnicity: White, Non-Hispanic	0.62	0.27	0.62	0.26	0.63	0.24
Immigrant Status: Population foreign-born	0.12	0.08	0.12	0.07	0.13	0.06
Household Income: Less than $35,000	0.20	0.15	0.21	0.13	0.21	0.12
Household Income: $35,000–$74,999	0.27	0.12	0.28	0.10	0.28	0.09
Household Income: $75,000–$149,999	0.29	0.12	0.30	0.10	0.31	0.08
Household Income: $150,000 or more	0.19	0.16	0.21	0.17	0.20	0.14
Work Status: Unemployed	0.02	0.02	0.02	0.02	0.02	0.01
Employment: Share of low-wage workers	0.18	0.05	0.19	0.03	0.19	0.03
Employment: Share of mid-wage workers	0.30	0.13	0.30	0.10	0.31	0.10
Employment: Share of high-wage workers	0.50	0.15	0.50	0.12	0.50	0.11
Tenure: Homeowners	0.69	0.24	0.70	0.21	0.70	0.19
Tenure: Renters	0.28	0.21	0.29	0.20	0.30	0.19
Car Ownership: 0	0.02	0.03	0.02	0.03	0.02	0.02
Car Ownership: 1	0.18	0.12	0.19	0.11	0.19	0.09
Car Ownership: 2	0.43	0.13	0.44	0.11	0.45	0.07
Car Ownership: 3 or more	0.35	0.16	0.35	0.13	0.35	0.10
Built Environment						
Persons per acre	4.24	4.08	4.15	3.88	4.29	2.12
Jobs per acre	2.09	4.82	2.04	4.40	2.12	3.60
Persons and jobs per acre	6.34	6.85	6.20	6.50	6.41	5.78
Liquor stores	0.04	0.26	0.17	0.61	3.95	4.59
Share of low-wage workplaces	0.22	0.11	0.22	0.10	0.23	0.11
Share of mid-wage workplaces	0.35	0.12	0.35	0.10	0.37	0.10
Share of high-wage workplaces	0.38	0.16	0.40	0.15	0.40	0.15
Intersections per acre	0.76	0.66	0.12	0.09	0.15	0.09
Connected node ratio	0.81	0.34	0.89	0.25	0.78	0.15
Beta index	1.02	0.44	1.04	0.29	0.10	0.15
Share of park space	0.03	0.12	0.03	0.10	0.08	0.10
Share of primary roads	0.09	0.23	0.11	0.21	0.11	0.16
Share of secondary roads	0.13	0.23	0.13	0.18	0.11	0.10
Share of tertiary roads	0.10	0.19	0.11	0.15	0.10	0.09
Share of residential roads	0.52	0.37	0.57	0.31	0.69	0.20
Street network length	1.62	1.20	6.35	4.14	32.61	16.01
Share of BNA low-stress bike facilities	0.61	0.37	0.65	0.31	0.72	0.17
Share of BNA high-stress bike facilities	0.24	0.29	0.26	0.25	0.28	0.17
Half-mile light rail transit shed	0.02	0.14	0.02	0.13	0.02	0.10

involved crashes. Although these are not person-level characteristics, together, the positive relationships underscore the possibility that individuals who bicycle because of the mode's lower relative economic costs or lifestyle preferences may be more susceptible to crash involvement. Regarding

Table 2. Negative binomial model estimates of total bicyclist-involved crashes.

Hexagon Diameter:	One-half mile		One mile		Two miles	
Variable	β	*SE*	β	*SE*	β	*SE*
Intercept	−8.92	0.55	−8.15	0.85	−3.84	0.61
Exposure Control Measures						
Persons per acre	0.14	0.01	0.13	0.01	0.08	0.02
Jobs per acre	0.03	<0.01	0.05	0.01	0.04	0.01
Street network length	0.47	0.03	0.15	0.01	0.03	<0.01
Socioeconomic Context						
Age: 65 years or older	−1.89	0.26	−2.01	0.31	−1.94	0.44
Education: Master's or PhD	2.15	0.47	2.25	0.61	1.96	0.76
Race/Ethnicity: White, Non-Hispanic	1.04	0.17	0.94	0.21	0.58	0.25
Household Income: Less than $35,000	3.66	0.26	3.81	0.33	2.60	0.46
Household Income: $35,000–$74,999	1.98	0.34	2.50	0.45	2.41	0.66
Work Status: Unemployed			−4.99	2.35		
Built Environment						
Liquor stores	0.31	0.06	0.12	0.03	0.08	0.01
Share of low-wage workplaces	0.69	0.28	0.75	0.37		
Share of mid-wage workplaces			1.28	0.41		
Connected node ratio	3.96	0.50	3.76	0.80	1.50	0.46
Share of park space	−0.65	0.30	−0.86	0.43	1.57	0.38
Share of secondary roads	1.57	0.11	1.13	0.20	1.07	0.38
Share of residential roads	−0.92	0.13	−0.82	0.19		
Share of BNA low-stress bike facilities			0.59	0.16		
Model summary						
Number of hexagons	10,409		2,673		553	
Log-likelihood	−6,771.08		−3,442.59		−1,199.50	
Theta (*SE*)	0.73 (0.04)		1.55 (0.10)		5.57 (0.72)	

Note: Estimates of significant ($p < 0.05$) independent variables only shown.

local land uses, areas with more liquor stores were found to be positively associated with total bicycle crash frequency; a result underlining the prospect that neighborhoods with an increased presence of impaired travelers may be linked to more observed bicyclist-involved crashes. Regardless of hexagon size, zones with a higher connected node ratio and share of secondary roads were positively associated with bicyclist-involved crash frequency. For the former environmental factor, this finding may indicate that areas with a traditional grid street pattern, where bicyclists may encounter more potential conflict points with motorists, produce a higher count of bicycle-related crashes, whereas the relationship between secondary roads and crash frequency suggests increased motorist speeds may factor into bicyclist safety. These described socioeconomic and built environment predictors of total bicyclist crash frequency were statistically significant after controlling for the three exposure measures of population density, employment density, and street network length, each which had a positive and significant relationship to the modeled outcome that is consistent with previous research (Amoh-Gyimah et al., 2016; Cai et al., 2016; Chen et al., 2018; Ding et al., 2020).

An extension of this set of aspatial NB models of total bicycle crashes is the estimation of SDMs of total vehicle-bicycle crash frequency, with an identical base specification and the addition of spatially lagged variables for the count of bicyclist-involved crashes and each of the previously described

Table 3. Spatial Durbin model estimates of total bicyclist-involved crashes.

Hexagon Diameter:	One-half mile		One mile		Two miles	
Variable	β	dy/dx	β	dy/dx	β	dy/dx
Intercept	−9.19		−7.29		−1.63	
Spatial Lag: Count of total crashes	0.29	0.02	0.06	0.03	0.02	0.08
Exposure Control Measures						
Persons per acre					0.05	0.18
Spatial Lag	0.14	0.01	0.14	0.07		
Jobs per acre	−0.01	−0.01			0.02	0.07
Spatial Lag	0.05	0.01	0.04	0.02		
Street network length	0.30	0.03	0.11	0.06	0.03	0.11
Spatial Lag						
Socioeconomic Context						
Age: 65 years old or more			−1.13	−0.60	−1.39	−4.71
Spatial Lag						
Household Income: Less than $35,000			1.12	0.60	1.39	4.70
Spatial Lag						
Household Income: $35,000-$74,999			1.46	0.78		
Spatial Lag						
Built Environment						
Liquor stores	0.25	0.02	0.09	0.05	0.07	0.24
Spatial Lag						
Connected node ratio	3.38	0.29	2.73	1.45	1.14	3.88
Spatial Lag	3.16	0.27	2.30	1.23		
Share of park space	−0.91	−0.08	−2.20	−1.17	1.51	5.12
Spatial Lag			3.39	1.80		
Share of secondary roads	1.60	0.14	0.93	0.50	0.94	3.18
Spatial Lag	−1.65	−0.14			2.38	8.05
Share of residential roads	−0.53	−0.05	−0.70	−0.37		
Spatial Lag	−0.50	−0.04	0.70	0.37		
Share of BNA low-stress bike facilities	−1.17	−0.10				
Spatial Lag	2.01	0.17				
Model summary						
Number of hexagons	10,409		2,673		553	
Log-likelihood	−6,566.87		−3,361.57		−1,231.14	
Theta (SE)	0.97 (0.05)		1.97 (0.14)		5.95 (0.78)	

Note: Estimates of significant ($p < 0.05$) independent variables only shown.

predictors. Table 3 shows the estimation results of the full SDM specifications for total bicycle crash frequency at the three different hexagon spatial extents. In these full SDM specifications, the exposure control measure of persons per acre at the two smaller zonal systems and jobs per acre at the one-mile hexagon zonal system were no longer statistically significant, while street network length remained positively associated with the outcome variable measured at all three spatial extents. The percentage of residents who identified as White and adults with higher education attainment no longer significantly predicted bicyclist crash frequency at any spatial extent in the SDMs, while the macro-level measures of low- and mid-wage workplaces were also no longer significantly associated with total bicyclist-involved crashes. The next discussion describes the marginal effects (dy/dx) of the statistically significant predictors in the SDM of total vehicle-bicycle crash frequency at a one-mile spatial extent; interpreted as the expected change in bicyclist-involved crashes per a one-unit change in the explanatory variable, all else being equal.

Focusing on significant socioeconomic context predictors, which were also significant at the two-mile spatial extent, a one-percent increase in the percentage of older adults within a Phoenix metro region one-mile hexagon was associated with an average expected decrease of 1.13 bicyclist-involved crashes, while a one-percent increase in the percentage of households in the lowest annual income cohort was related to an average increase of 1.12 bicyclist-involved crashes. The share of households with a reported annual income between $35,000 and $74,999 was only significantly associated with bicyclist-involved crashes at the one-mile spatial extent, with a one-percent increase in this determinant related to an average increase of 1.46 bicyclist-involved crashes. A handful of built environment predictors in the base NB models were also found to be significant in the SDMs, with three measures significant at all three hexagon geographies (liquor stores, connected node ratio, and share of secondary roads). Connected node ratio was the only one of these built environment measures to have a spatially lagged variable that was significantly predictive of the total crash outcome at a one-mile spatial extent, revealing that the presence of a grid street pattern (characteristic of more urban areas) has a spillover effect on bicyclist-involved crash counts. Turning to the spatially lagged dependent variable, which revealed significant spillover effects, an increase of 100 total bicycle crashes in any study area hexagon was associated with an expected average increase of 6.00 total bicyclist-involved crashes in the immediately neighboring one-mile hexagons.

3.3. Spatial factors associated with fatal and serious injury bicycle crashes

The estimates of the NB models assessing the socioeconomic context and built environment predictors of the count of bicyclist-related crashes resulting in a fatal or serious injury per one-half-, one-, and two-mile hexagons are shown in Table 4. Similar to the previous NB models of total crash frequency, an increased share of adults older than 65 years of age was found to be negatively associated with an increase in K- and A-level injuries over the five-year study period, while an increase of households within the lowest annual income cohort was found to be positively associated with an increase in K- and A-level injuries. Yet, unlike in the previous NB crash model results, an increased share of unemployed working-age adults per hexagon was found to be related to a decrease in bicyclist-related crashes of the two highest severity classifications. While inconclusive from any macro-level assessment, this relationship could be a consequence of employed adults in the study area being less likely to commute via bicycle. Similar to the base NB model specification for total bicyclist-related crashes, the three exposure measures in these base NB models of bicycle

Table 4. Negative binomial model estimates of bicyclist-involved crashes with fatality or serious injury.

Hexagon Diameter: Variable	One-half mile β	One-half mile *SE*	One mile β	One mile *SE*	Two miles β	Two miles *SE*
Intercept	−10.21	1.22	−6.69	1.27	−2.41	0.56
Exposure Control Measures						
Persons per acre	0.10	0.01	0.10	0.01	0.12	0.02
Jobs per acre	0.03	<0.01	0.05	0.01	0.05	0.01
Street network length	0.21	0.05	0.07	0.02	0.02	0.01
Socioeconomic Context						
Age: 65 years old or more	−1.79	0.42	−1.78	0.43	−1.65	0.55
Household Income: Less than $35,000	2.67	0.34	2.89	0.40	3.37	0.59
Household Income: $35,000-$74,999			1.74	0.61		
Work Status: Unemployed	−7.08	3.43	−15.64	4.12	−13.18	6.65
Built Environment						
Connected node ratio	6.13	1.25	3.62	1.31		
Share of park space			−2.14	0.96		
Model summary						
Number of hexagons	10,409		2,673		553	
Log-likelihood	−2,095.81		−1,336.92		−634.97	
Theta (*SE*)	0.68 (0.11)		1.40 (0.24)		3.76 (0.92)	

Note: Estimates of significant ($p < 0.05$) independent variables only shown.

crashes resulting in a fatality or severe injury were statistically significant and positive in association at all spatial extents. While no built environment metric was significant in each NB model, an increase in the connected node ratio was found to have a significant, positive link to the count of K- and A-level injuries observed in one-half- and one-mile hexagons.

Table 5 provides the results for three SDMs of bicyclist-involved crashes resulting in a fatality or serious injury measured at the three spatial extents. Interestingly, the spatially lagged dependent variable in each of the full SDM specifications is positive and significant, indicating that this subset of more serious bicyclist-related injuries, which are rare in nature, are also spatially clustered. However, none of the other predictors specified in these SDMs are significant at each of the three extents. In terms of socioeconomic context, the percentage of adults older than 65 years of age was the only significant variable in the one-half-mile hexagon model, with a one-percent increase in the share of older adults associated with a decrease of 3.62 bicyclist-related crashes classified by one of the two highest severity levels. Measured at a one-mile spatial extent, the percentage of households reporting an annual income between $35,000 and $74,999 was positively associated with the more severe category of bicyclist-related crashes, with a one-percent increase of households in this income cohort associated with a 3.21 increase in vehicle-bicycle crashes with a fatality or severe injury. In terms of the built environment, an increase in the average connected node ratio of a one-half or one-mile hexagon was positively associated with the reported frequency of bicyclist-related crashes classified by the two highest

Table 5. Spatial Durbin model estimates of bicyclist-involved crashes with fatality or serious injury.

Hexagon Diameter:	One-half mile		One mile		Two miles	
Variable	β	dy/dx	β	dy/dx	β	dy/dx
Intercept	−11.35		−8.60		−1.92	
Spatial Lag: Count of KA crashes	0.99	0.01	0.48	0.05	0.20	0.13
Exposure Control Measures						
Persons per acre					0.15	0.10
Spatial Lag	0.13	0.01	0.17	0.02		
Jobs per acre					0.05	0.03
Spatial Lag	0.06	0.01	0.05	0.01		
Street network length	0.14	0.01			0.01	0.01
Spatial Lag						
Socioeconomic Context						
Age: 65 years old or more	−3.62	−0.05				
Spatial Lag	2.89	0.04				
Household Income: $35,000–$74,999			3.21	0.32		
Spatial Lag			−3.23	−0.32		
Built Environment						
Connected node ratio	7.20	0.10	3.74	0.37		
Spatial Lag			2.17	0.21		
Share of park space			−3.77	−0.37		
Spatial Lag			3.07	0.30		
Model summary						
Number of hexagons	10,409		2,673		553	
Log-likelihood	−2,096.59		−1,305.52		−639.24	
Theta (*SE*)	0.68 (0.11)		2.11 (0.48)		3.71 (0.94)	

Note: Estimates of significant ($p < 0.05$) independent variables only shown.

severity levels, while an increase of park space within a one-mile hexagon was associated with a decrease in bicyclist-related crashes with a fatality or severe injury. The spatially lagged variables of these latter two predictors when measured at a one-mile hexagon were positive and significant, revealing that an increase in the presence of a more traditional street network or park space in neighboring hexagons may relate to an increase in bicyclist-related crashes with a K- or A-level injury.

4. Conclusion

This study presented a spatial analysis of factors associated with bicycle-vehicle crash frequencies (total and severe/fatal injury crashes) in the Phoenix metro region. Five years of crash data were analyzed along with several macro-level measures of exposure, roadway network characteristics, sociodemographic and economic characteristics, and built environment features. These data were summarized at three uniform hexagon spatial extents (as opposed to single zone systems of TAZs, census block groups, or statistical area levels considered in prior research). This study provides new insights into the relationship between a robust set of macro-level characteristics and bicycle-vehicle crashes, which may be useful to planners, engineers, and researchers seeking to implement strategies and countermeasures aimed at improving bicyclist safety. Findings from traditional NB model estimates, which failed to account for spatial spillover effects, were advanced by estimating spatial Durbin models that tested for the significance of spatial-lagged predictors of the dependent and independent variables that were solely explored in the aspatial NB models.

With respect to notable findings of this study, both total crashes and crashes resulting in a K- or A-level injury were more likely to occur in zones with higher activity densities (population and employment), which likely reflects greater opportunity for individuals to travel on a bicycle, and a more traditional grid street network that without proper treatments may introduce more conflict points between bicyclists and motorists (i.e., increased exposure). Spatial spillover effects were found for both the outcome of total bicycle-vehicle crashes as well as the count of severe/fatal injury bicycle-vehicle crashes, highlighting the need for a more holistic approach to addressing missing links in the region's network of low-stress facilities. Zones with a higher share of secondary roads were also more likely to have increased bicyclist-related crash frequencies, further showing the importance of creating more bike-friendly facilities that separate bicyclists from faster-moving vehicles. Finally, the increased presence of liquor stores within a hexagon was associated with total vehicle-bicycle crash frequencies, which suggests a need for better protection for bicyclists in these areas via infrastructure treatments and perhaps education and/or enforcement strategies aimed at motorists.

There are some limitations to this study, most of which can also be considered directions for future research. First, this study utilized a police-reported crash data set, and it is possible some bicyclist-involved crashes were not reported, which is a common limitation when analyzing almost any crash data set. Second, though a robust set of socioeconomic context and built environment predictors were examined, a relatively small set of these macro-level predictors demonstrated significant spillover effects on bicyclist-involved crashes, pointing to the need to explore other environmental predictors or zonal systems. Additionally, the study's analysis of macro-level factors of socioeconomic context somewhat assumes that bicyclist-involved crashes are experienced by the residents of a particular zone, which may not be the case for longer bicycle trips or those near an out-of-home location, and also suffers from a contextual fallacy in which it is assumed that the zonal characteristics are reflective of the individual involved in the crash event. Future studies could also consider a more bicyclist-specific exposure measure (e.g., bicyclist volumes); while not available for this study, agencies are developing more robust bicycle count programs and these data may be more readily available for future studies. Finally, future work could include incident-level analyses of bicyclist injuries that account for person- and event-specific attributes as well as important macro-level factors to provide a more complete picture of how these spatial attributes relate to vehicle-bicycle crashes and injury outcomes.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Steven R. Gehrke http://orcid.org/0000-0001-9355-5571

References

Amoh-Gyimah, R., Saberi, M., & Sarvi, M. (2016). Macroscopic modeling of pedestrian and bicycle crashes: A cross-comparison of estimation methods. *Accident; Analysis and Prevention*, *93*, 147–159. doi:10.1016/j.aap.2016.05.001

Anselin, L. (2010). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93–115. doi:10.1111/j.1538-4632.1995.tb00338.x

Arizona Department of Transportation (ADOT). (2017). *Arizona's crash report forms instruction manual* (11th ed.). Arizona Department of Transportation.

Bahrololoom, S., Young, W., & Logan, D. (2020). Modelling injury severity of bicyclists in bicycle-car crashes at intersections. *Accident; Analysis and Prevention*, *144*, 105597. doi:10.1016/j.aap.2020.105597

Brown, V., Diomedi, B. Z., Moodie, M., Veerman, J. L., & Carter, R. (2016). A systematic review of economic analyses of active transport interventions that include physical activity benefits. *Transport Policy*, *45*, 190–208. doi:10.1016/j.tranpol.2015.10.003

Cai, Q., Lee, J., Eluru, N., & Abdel-Aty, M. (2016). Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. *Accident; Analysis and Prevention*, *93*, 14–22. doi:10.1016/j.aap.2016.04.018

Chandia-Poblete, D., Hill, R. L., Aguilar-Farias, N., & Heesch, K. C. (2021). Individual and contextual factors associated with bicyclist injury severity in traffic incidents between bicyclists and motorists in Chile. *Accident; Analysis and Prevention*, *154*, 106077. doi:10.1016/j.aap.2021.106077

Chen, P. (2015). Built environment factors in explaining the automobile-involved bicycle crash frequencies: A spatial statistic approach. *Safety Science*, *79*, 336–343. doi:10.1016/j.ssci.2015.06.016

Chen, P., Sun, F., Wang, Z., Gao, X., Jiao, J., & Tao, Z. (2018). Built environment effects on bike crash frequency and risk in Beijing. *Journal of Safety Research*, *64*, 135–143. doi:10.1016/j.jsr.2017.12.008

Johan de Hartog, J., Boogaard, H., Nijland, H., & Hoek, G. (2010). Do the health benefits of cycling outweigh the risks? *Environmental Health Perspectives*, *118*(8), 1109–1116. doi:10.1289/ehp.0901747

Ding, H., Sze, N. N., Li, H., & Guo, Y. (2020). Roles of infrastructure and land use in bicycle crash exposure and frequency: A case study using Greater London bike sharing data. *Accident; Analysis and Prevention*, *144*, 105652. doi:10.1016/j.aap.2020.105652

Fotheringham, A. S., & Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A: Economy and Space*, *23*(7), 1025–1044. doi:10.1068/a231025

Gehrke, S. R., & Welch, T. F. (2019). A bikeshare station area typology to forecast the station-level ridership of system expansion. *Journal of Transport and Land Use*, *12*(1), 221–235. https://www.jstor.org/stable/26911265 doi:10.5198/jtlu.2019.1395

Guo, Y., Osama, A., & Sayed, T. (2018). A cross-comparison of different techniques for modeling macro-level cyclist crashes. *Accident; Analysis and Prevention*, *113*, 38–46. doi:10.1016/j.aap.2018.01.015

Ji, S., Wang, Y., & Wang, Y. (2021). Geographically weighted poisson regression under linear model of coregionalization assistance: Application to a bicycle crash study. *Accident; Analysis and Prevention*, *159*, 106230. doi:10.1016/j.aap.2021.106230

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Lin, Z., & Fan, W. (David). (2021). Exploring bicyclist injury severity in bicycle-vehicle crashes using latent class clustering analysis and partial proportional odds models. *Journal of Safety Research*, *76*, 101–117. doi:10.1016/j.jsr.2020.11.012

Merlin, L. A., Guerra, E., & Dumbaugh, E. (2020). Crash risk, crash exposure, and the built environment: A conceptual review. *Accident; Analysis and Prevention*, *134*, 105244. doi:10.1016/j.aap.2019.07.020

NHTSA (2020). Overview of Motor Vehicle Crashes in 2019 (DOT HS 813 060). *National Highway Traffic Safety Administration*. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813060

NTSB. (2019). Bicyclist Safety on US Roadways: Crash Risks and Countermeasures (NTSB/SS-10/01 PB2019-101397; p. 100). *National Transportation Safety Board*. https://www.ntsb.gov/safety/safety-studies/Documents/SS1901.pdf

Osama, A., & Sayed, T. (2017). Evaluating the impact of socioeconomics, land use, built environment, and road facility on cyclist safety. *Transportation Research Record: Journal of the Transportation Research Board*, *2659*(1), 33–42. doi:10.3141/2659-04

People for Bikes. (2021). Bicycle network analysis. https://bna.peopleforbikes.org/#/.

Robartes, E., & Chen, T. D. (2017). The effect of crash characteristics on cyclist injuries: An analysis of Virginia automobile-bicycle crash data. *Accident; Analysis and Prevention*, *104*, 165–173. doi:10.1016/j.aap.2017.04.020

Saha, D., Alluri, P., Gan, A., & Wu, W. (2018). Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models. *Accident; Analysis and Prevention*, *118*, 166–177. doi:10.1016/j.aap.2018.02.014

Sener, I. N., Lee, K., Hudson, J. G., Martin, M., & Dai, B. (2021). The challenge of safe and active transportation: Macrolevel examination of pedestrian and bicycle crashes in the Austin District. *Journal of Transportation Safety & Security*, *13*(5), 525–551. doi:10.1080/19439962.2019.1645778

Wang, J., Huang, H., & Zeng, Q. (2017). The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian. *Accident; Analysis and Prevention*, *98*, 223–231. doi:10.1016/j.aap.2016.10.018

Zhu, S. (2021). Analysis of the severity of vehicle-bicycle crashes with data mining techniques. *Journal of Safety Research*, *76*, 218–227. doi:10.1016/j.jsr.2020.11.011