Old Dominion University
ODU Digital Commons

Civil & Environmental Engineering Faculty Publications

Civil & Environmental Engineering

2023

Are Ride-Hailing Services Safer Than Taxis? A Multivariate Spatial Approach with Accomodation of Exposure Uncertainty

Guocong Zhai
Old Dominion University, gzhai001@odu.edu

Kun Xie
Old Dominion University, kxie@odu.edu

Hong Yang
Old Dominion University, hyang@odu.edu

Di Yang
Morgan State University

Follow this and additional works at: https://digitalcommons.odu.edu/cee_fac_pubs

Part of the Automotive Engineering Commons, Computer Engineering Commons, Transportation Commons, and the Transportation Engineering Commons

Original Publication Citation
Zhai, G., Xie, K., Yang, H., & Yang, D. (2023). *Are ride-hailing services safer than taxis? A multivariate spatial approach with accommodation of exposure uncertainty.* SSRN. https://doi.org/10.2139/ssrn.4334669

This Article is brought to you for free and open access by the Civil & Environmental Engineering at ODU Digital Commons. It has been accepted for inclusion in Civil & Environmental Engineering Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

Are ride-hailing services safer than taxis? A multivariate spatial approach with accommodation of exposure uncertainty

Guocong Zhai[a], Kun Xie[a*], Hong Yang[b], Di Yang[c]

[a] Department of Civil & Environmental Engineering, Old Dominion University, 129C Kaufman Hall, Norfolk, VA, 23529, USA

[b] Department of Electrical & Computer Engineering, Old Dominion University, 4700 Elkhorn Avenue, Norfolk, VA, 23529, USA

[c] Department of Transportation & Urban Infrastructure Studies, Morgan State University, 1700 E Cold Spring Ln, Baltimore, MD, 21251, USA

[*] Corresponding author, kxie@odu.edu

Abstract

Despite many research efforts on ride-hailing services and taxis, limited studies have compared the safety performance of the two modes. A major challenge is the need for reliable mode-specific exposure data to model their safety outcomes. Moreover, crash frequencies of the two modes by injury severities tend to be spatially and inherently correlated. To fully address these issues, this study proposes a novel multivariate conditional autoregressive model considering measurement errors in mode-specific exposures (MVCARME). More specially, a classical measurement error structure is used to accommodate the uncertainty of mode-specific exposures estimated, and a multivariate spatial specification is adopted to capture potential spatial and inherent correlations. The model estimation is accelerated by an integrated nest Laplace approximation method. The census tracts in the city of Chicago are set as the spatial analysis unit. The mode-specific

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

exposures (vehicle-mile-traveled) in each census tract are estimated by trip assignments using ride-hailing and taxi trip data in 2019. The modeling results indicate that both ride-hailing crashes and taxi crashes are positively associated with transportation factors (e.g., vehicle-mile-traveled, mode-specific vehicle-mile-traveled, and traffic signal numbers), land use factors (i.e., number of educational and alcohol-related sites), and demographic factors (e.g., median household income, transit ratio, and walk ratio). By comparison, the proposed model outperforms the others (i.e., negative binomial models and multivariate conditional autoregressive model) by yielding the lowest deviance information criterion (DIC), Watanabe-Akaike information criterion (WAIC), mean absolute error (MAE), and root-mean-square error (RMSE). According to the results of *t*-tests, ride-hailing services are found to be prone to a higher risk of minor injury crashes compared with taxis, despite no significant difference between the risks of severe injury crashes. Methodologically, this study adds to the literature a robust safety evaluation approach for comparing crash risks of different modes. At the same time, practically, it provides researchers, practitioners, and policy-makers insights into the safety management of various mobility alternatives.

Keywords: Ride-hailing crashes, Taxi crashes, Integrated nested Laplace approximation, Multivariate conditional autoregressive model, Measurement errors, Safety comparison

1. Introduction

Ride-hailing services (e.g., Uber, Lyft, and Didi Chuxing) and taxis are more convenient than public transit and more economical compared to private vehicles (Yang *et al.* 2022). As indicated by Jiao *et al.* (2020) and Yang *et al.* (2022), ride-hailing services and taxis would also reduce private vehicle use, car ownership, vehicle emissions, etc. Despite many

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

similarities between ride-hailing services and taxis, ride-hailing services tend to be more attractive than taxis (Rayle *et al.* 2016). For example, ride-hailing services outperform taxis by shorter and more reliable waiting times (Rayle *et al.* 2016), which might significantly decrease taxi demands and the corresponding labor incomes of taxi drivers. Ride-hailing users are also found to have lower car ownership and driving frequencies compared to taxi users, more likely to reduce their reliance on private vehicles (Rayle *et al.* 2016).

Some cities, such as Chicago, London, Beijing, and Shanghai, have implemented policies to suspend ride-hailing services temporarily or partially because of aggravated traffic congestion issues, repeated protests from local taxi drivers, and increased crashes and crimes related to ride-hailing services (Yang *et al.* 2022). Take the city of Chicago as an example. The congestion surcharge policy was only issued to suspend ride-hailing services partially from January 2020, mainly due to aggravated traffic congestion issues (Brown 2022). Such policies have raised concerns about whether ride-hailing services and taxis should be regulated and treated differently, especially for the safety management of the two alternatives. For instance, taxi drivers are usually professionally trained and not allowed to use phones while driving. In contrast, ride-hailing drivers typically drive private vehicles and rely heavily on ride-hailing apps to pick up passengers while driving. Phone use while driving will induce driver distractions, thus increasing the likelihood of crash occurrences (Chen *et al.* 2022).

The main challenge in comparing the safety performance of ride-hailing services and taxis is the need for reliable mode-specific exposure data, such as annual average daily traffic

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

(AADT) or vehicle-mile-traveled (VMT). Mode-specific exposures are considered the most indispensable explanatory factor in modeling mode-specific crashes (Ma *et al.* 2019, Xu *et al.* 2022). The ignorance of mode-specific exposures will induce severe omitted variable issues, affecting crash safety inference and management. The critical component for ride-hailing and taxi crash inference is to estimate mode-specific exposures by trip origin-destination (OD) pairs of ride-hailing services and taxis in this study. In addition, the estimated mode-specific exposures at the census tract level might have measurement errors. To be more specific, the estimated mode-specific exposure data are not the ground truth of the mode-specific exposures. For example, the actual routes between the same OD pair might differ slightly for different ride-hailing and taxi drivers spatially and temporally (Liu and Jiang 2022). For simplification, one could assume that all ride-hailing and taxi drivers select the same shortest routes for the same OD pair. However, the exact positions of trip origins and destinations for the two alternatives are usually unknown due to privacy protection. Although the centroids are typically set as good representations of trip origins or destinations in trip assignments at zone levels, measurement errors in mode-specific exposures still exist. More reliable estimates of mode-specific exposures should be incorporated into crash safety modeling to account for potential measurement errors in mode-specific exposures.

Further, conventional crash frequency modeling approaches, such as the Poisson models (Ma *et al.* 2019), Poisson-Gamma models (Zhai *et al.* 2022), and Poisson-lognormal models (Xie *et al.* 2015b), heavily rely on the assumption of independence of crash observations. However, such an assumption is frequently violated due to the potential

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

spatial correlations among different sites and inherent correlations across different types of crashes. Firstly, spatial correlations refer to the case that crashes occurring in one site will also affect crash observations in neighboring sites except for itself (Ziakopoulos and Yannis 2020). For example, crashes occurring at one site might result in rear-end secondary crashes at the corresponding upstream sites due to the disrupted traffic (Yang *et al.* 2018). Secondly, different types of crashes are found to be inherently correlated with each other because unobserved safety factors might simultaneously affect the frequencies of various crash types at one site. For instance, unobserved safety factors (e.g., safer driving behavior, high intensities of traffic law enforcement, good road lighting conditions) would generally reduce severe and minor injury crashes at one site (Xie *et al.* 2019). However, few studies have involved the unobserved safety factors above in modeling crash frequencies by severity (i.e., severe injury and minor injury crashes) because of the data availability. Therefore, traditional crash frequency modeling would get biased estimates if spatial correlations among different observation sites and inherent correlations across different types of crashes are not appropriately considered.

Because previous research that jointly addressed the three research gaps above is relatively limited, this study aims to compare the safety performance of ride-hailing services and taxis by developing a novel multivariate conditional autoregressive model considering measurement errors in mode-specific exposures (MVCARME), using data from Chicago as a case study. First, the shortest path for each OD pair at census tract levels is estimated by the OpenStreetMap-based Routing Service (Giraud 2022). Then, we could get the estimated mode-specific exposures (VMT) by aggregating the shortest paths within each

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

census tract for all ride-hailing or taxi trips in 2019. Secondly, the shortest paths refer to the expected routes on the actual road network between centroids of the OD pairs. Different ride-hailing or taxi drivers might have different preferences for routing choices. To consider the measurement errors between the estimated and actual mode-specific VMT, we integrate the classical measurement error structure into crash safety modeling for ride-hailing and taxi crashes. Thirdly, multivariate conditional autoregressive components are developed to jointly account for spatial correlations among different crash observation sites and inherent correlations across various types of crashes (i.e., severe ride-hailing crashes, minor ride-hailing crashes, severe taxi crashes, and minor taxi crashes). It should be noted that an integrated nest Laplace approximation (INLA) method is used to accelerate parameter estimations in Bayesian inference.

The remainder is constructed as follows. The second section reviews contributing factors for ride-hailing crashes and taxi crashes and commonly used crash frequency modeling approaches. The following steps are to prepare the data and introduce the proposed crash frequency modeling approaches, followed by the modeling results and discussions. The final section concludes the findings and provides corresponding suggestions.

2. Literature Review

This section reviews safety factors for mode-specific crashes and commonly used crash frequency modeling approaches in previous studies. Mode-specific crashes mainly refer to ride-hailing or taxi crashes in this section.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

2.1. Safety factors for mode-specific crashes

Crash exposures (i.e., AADT and VMT) are the most important contributing factors to ride-hailing or taxi crash frequency modeling, especially for mode-specific exposures. Without considering the actual exposures, prior studies used demographic characteristics (e.g., population, number of residents, and number of elderly) as proxy exposures in ride-hailing crash safety inference at the city level (Greenwood and Wattal 2017, Dills and Mulholland 2018, Barrios *et al.* 2020). Of course, some studies considered exposures for all motor vehicles (VMT) in ride-hailing crash safety analyses (Brazil and Kirk 2016, Brazil and Kirk 2020, Kirk *et al.* 2020). However, few considered mode-specific exposures when modeling ride-hailing or taxi crashes at the city level. In addition, like mode-specific crash frequency modeling at the city level, only Ma *et al.* (2019) considered exposures and mode-specific exposures for taxi crash modeling at the census tract level in the existing literature. More specifically, the mode-specific exposures for taxi crashes, taxi VMT, were estimated by trip assignments at the census tract level using the Euclidean distance between each taxi OD pair as the expected route. However, the actual route distance tended to be longer than the Euclidean route distance for the same OD pair because of the geometry characteristics of the actual road networks. Further, the relationships between mode-specific crashes and mode-specific exposures were also investigated at the individual level (Mao *et al.* 2021). More specifically, the vehicle-kilometer-traveled (VKT) being ride-hailing was involved in ride-hailing crash modeling. However, crash records and mode-specific exposures for ride-hailing services were unavailable to the public due to privacy and commercial reasons in Mao *et al.* (2021).

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Besides the crash exposures above, other transportation factors are summarized more extensively from previous crash frequency modeling studies (Xie *et al.* 2019, Cui and Xie 2021, Kabir *et al.* 2021, Xu *et al.* 2022) because of the limited studies on ride-hailing crashes and taxi crashes. For example, crash frequencies were found to be positively associated with the bus stop number (Wei and Lovegrove 2013, Xie *et al.* 2019), stop sign density (Ding *et al.* 2018a), traffic signal characteristics (Kabir *et al.* 2021), intersection number (Marshall and Garrick 2011), road length (Kamel *et al.* 2019, Cui and Xie 2021), and truck ratio (Hou *et al.* 2018).

In terms of land use factors, the number of points of interest (Ma *et al.* 2019), the number of education sites (Ukkusuri *et al.* 2012), and the number of bars (Mitra and Washington 2012b) were found to be positively correlated to crash frequencies. However, ratios of residential areas, open space areas, and institutional areas were found to be insensitive to taxi crashes (Ma *et al.* 2019). Moreover, demographic factors are also summarized in a more extensive range to help understand mode-specific crash frequencies. For instance, crash frequencies are found to be positively associated with more population (Cui and Xie 2021), higher employment density (Cai *et al.* 2016), and higher median household income (Xie *et al.* 2019). It should be noted that the ratio of commuters by public transit or walking was also crucial in understanding the potential traffic conflicts between ride-hailing services/taxis and public transits/pedestrians (Ding *et al.* 2018a, Xie *et al.* 2019).

It is worth mentioning that some studies have investigated ride-hailing crashes (Mao *et al.* 2021) and taxi crashes (Ma *et al.* 2019), considering mode-specific exposures. However,

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

the corresponding mode-specific exposures are either inaccessible to the public (Mao *et al.* 2021) or unreliably enough estimated by Ma *et al.* (2019). More reliable estimations of mode-specific exposures should be incorporated in modeling ride-hailing crashes and taxi crashes, except for other commonly used transportation, land use, and demographic factors.

2.2. Modeling approaches for crash frequencies

Traditional crash frequency modeling approaches, such as Poisson (Ma *et al.* 2019) and Poisson-Gamma (Brazil and Kirk 2020, Mao *et al.* 2021), have been used to understand ride-hailing or taxi crash frequencies. Such crash frequency modeling approaches assume that crash observations are independent. However, the assumption is often violated by spatial correlations of crash observation sites and inherent correlations across various types of crashes (Xie *et al.* 2019, Ziakopoulos and Yannis 2020).

In terms of the spatial correlations of crash frequencies, observed or unobserved safety factors of one site might affect crash occurrences at neighboring sites (Xie *et al.* 2019). The correlation matrices among different sites were primarily designed to accommodate the potential spatial correlations in generalized estimating equations (GEEs) (Abdel-Aty and Wang 2006, Mohammadi *et al.* 2014), spatial autoregressive (SAR) models (Xie *et al.* 2015a, Gaweesh *et al.* 2019), and conditional autoregressive (CAR) models (Ma *et al.* 2019, Xie *et al.* 2019, Xu *et al.* 2022). The CAR models are then reviewed to help further understand the crash frequency modeling approach used in this study (please refer to Mohammadi *et al.* (2014) and Xie *et al.* (2015a) for more details on GEEs and SAR models, respectively). In addition, SAR models are often considered special types of CAR models

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

(Cressie 2015). In particular, the CAR models enabled complicated model settings and faster computation speeds than SAR models, especially with larger datasets (Wang and Kockelman 2013). On the other hand, the CAR models have also been widely used in crash safety analyses at intersections (Xie *et al.* 2014), road segments (Yang *et al.* 2021), census blocks (Saha *et al.* 2018), census tracts (Ma *et al.* 2019), and traffic analysis zones (Kamel and Sayed 2020). Therefore, the CAR model specification is developed in this study to accommodate the spatial correlations across different crash observation sites.

The inherent correlations are usually caused by unobserved safety factors across various types of crashes. Multivariate models (Xie *et al.* 2015c, Bhowmik *et al.* 2019, Xie *et al.* 2019) and the multinomial generalized Poisson models (Chiou and Fu 2013, Chiou *et al.* 2014) were developed to accommodate the inherent correlations by shared error terms. For instance, Xie *et al.* (2015c) developed multivariate spatial count models to accommodate inherent correlations across different types of truck crashes. To jointly account for the potential spatial and inherent correlations above, multivariate conditional autoregressive (MVCAR) models have been proposed in previous crash safety studies (Wang and Kockelman 2013, Cheng *et al.* 2018, Xie *et al.* 2019, Yang *et al.* 2019). Besides the spatial and inherent correlations, the measurement errors of mode-specific exposures have also been widely considered to get more reliable crash safety estimations and inferences (Xie *et al.* 2015c, Kamel and Sayed 2020, Xu *et al.* 2022).

To sum up, limited studies have jointly accommodated spatial correlations of different sites, inherent correlations across different types of crashes, and measurement errors in mode-

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

specific exposures. Therefore, the MVCARME model is developed to investigate mode-specific crashes (i.e., ride-hailing and taxi crashes) by injury severity in this study.

3. Data Preparation

This study uses the operations of ride-hailing services and taxis in Chicago as a case study. We collected crash data, transportation, land use, and demographic factors in Chicago in 2019. The census tracts ($N=801$) in Chicago were used as spatial analysis units for ride-hailing and taxi crash frequency modeling.

3.1. Crash data

Only crash data in 2019 were used to model ride-hailing and taxi crash frequencies (Chicago Police Department 2019a) to avoid the potential impacts of COVID-19 and the congestion surcharge policy initiated in 2020. In addition, trip data of ride-hailing services were only available for the public from November 2018 (Chicago Department of Business Affairs & Consumer Protection 2019b), which were the critical components to estimating mode-specific exposures for ride-hailing services.

One big challenge in mode-specific crash safety inference is the lack of identifiers for ride-hailing crashes and taxi crashes. The crash vehicle data in Chicago provided opportunities to identify ride-hailing and taxi crashes by the term named vehicle-use (Chicago Police Department 2019b). More specially, one crash was regarded as a ride-hailing crash only if one of the vehicle-use in the same crash was the rideshare service; one crash was classified as a taxi crash only if one of the vehicle-use in the same crash was the taxi/for hire. By

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

severity, severe and minor crashes were identified by the most severe injury in crash data. Severe crashes involved fatal injury, incapacitating injury, and non-incapacitating injury crashes. Similarly, minor crashes included no indication of injury and no evident crashes.

3.2. Exposure estimation

Regarding exposures, VMT could be estimated by multiplying traffic volumes in 2019 and the corresponding road length in each census tract (Illinois Department of Transportation 2019). Because vehicle trajectory data are unavailable for ride-hailing services and taxis in this study, we cannot obtain the actual mode-specific exposures in each census tract using the ride-hailing OD data (Chicago Department of Business Affairs & Consumer Protection 2019b) and taxi OD data (Chicago Department of Business Affairs & Consumer Protection 2019a). To get reliable estimations of the mode-specific exposures, we assume that the origins and destinations of all ride-hailing or taxi trips occurred at centroids of the corresponding census tracts. It should be noted that only 68.48% of ride-hailing trips and 64.78% of taxi trips have census tract identifiers in both trip origins and destinations, mainly due to privacy concerns. We assume that trip distances of OD pairs with both origins and destinations for ride-hailing services and taxis are proportional to the actual trip distances. For instance, the trip distances of ride-hailing services with identifiers in both trip origins and destinations account for 68.48% of the total trip distances of ride-hailing services in Chicago in 2019.

As indicated in Figure 1, the procedure to estimate mode-specific exposure, $ModeVMT_i^k$, is described as follows:

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Step 1. Obtain the number of trips N_{ij}^{k} and trip distances D_{ij}^{k} for OD_{ij} where i and j are census tract identifiers, k denotes ride-hailing services or taxis.

Step 2. Classify OD_{ij} into $IntraOD_{ii}$ and $InterOD_{ij}\left(i \neq j\right)$ where only d_{ii}^{k} for $IntraOD_{ii}$ is the actual trip distances for census tract i.

Step 3. Estimate the expected shortest $Route_{ij}$ and distance $d_{ij}^{'}$ for $InterOD_{ij}$ on the actual road network with OpenStreetMap (OSM) Routing Services (Giraud 2022).

Step 4. Assign d_{ij}^{k} for $InterOD_{ij}$ into census tract vector $\mathbf{I} = (1, 2, 3, ..., N)$ by the proportion of $d_{ij}^{'}$ in each census tract.

Step 5. Calculate d_{i}^{k} by summing d_{ij}^{k} for all $InterOD_{ij}$ in census tract i.

Step 6. Calculate the mode-specific exposure $ModeVMT_{i}^{k}$ for ride-hailing services or taxis by summing d_{ii}^{k} from $IntraOD_{ii}$ and d_{i}^{k} from $InterOD_{ij}$.

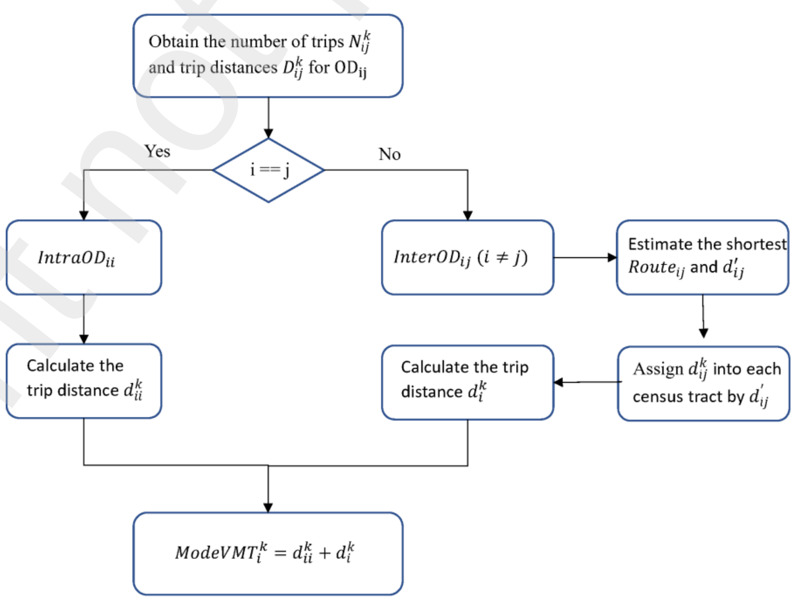


Figure 1. Estimation procedure of mode-specific exposures

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

3.3. Safety factors

Besides VMT and mode-specific VMT, other transportation factors are aggregated into census tracts by the *sf* package in the programming language R (Pebesma 2018), such as bus stop numbers, metro station numbers, crosswalk numbers, stop sign numbers, and traffic signal numbers (OpenStreetMap 2019b). We also collected some land use factors (OpenStreetMap 2019a) for each census tract, such as commercial ratio, residential ratio, recreational ratio, green space ratio, number of education sites, and number of alcohol-related sites by the *sf* package in the programming language R (Pebesma 2018). Demographic factors at the census tract level were also collected from the American Community Survey (ACS) released in 2019 by the *tidycensus* package of the programming language R (Walker *et al.* 2021). For instance, commonly used demographic factors included the number of populations, the number of populations younger than 18, median household income in USD, and the number of house units. Moreover, demographic factors for commuters were also involved, such as transit ratio, cycling ratio, and walk ratio.

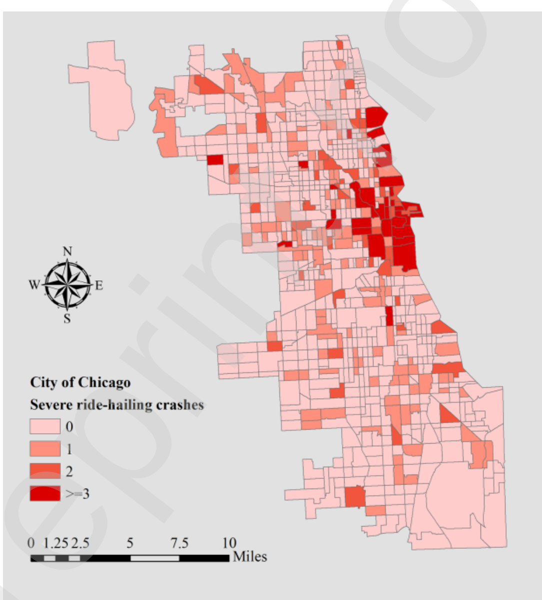
3.4. Descriptive analysis

Table 1 summarizes crash data, transportation, land use, and demographic factors. For example, the average severe crash frequency involving ride-hailing services at the census tract level was 0.47, much higher than that of taxis (0.33). In contrast, the average number of minor ride-hailing crashes (3.40) was much lower than that of minor taxi crashes (3.55). In addition, Figure 2 suggests significant spatial correlations of crash occurrence at the census tract level. For instance, most ride-hailing crashes and taxi crashes are observed in the central areas of Chicago. More severe and minor ride-hailing crashes are distributed in

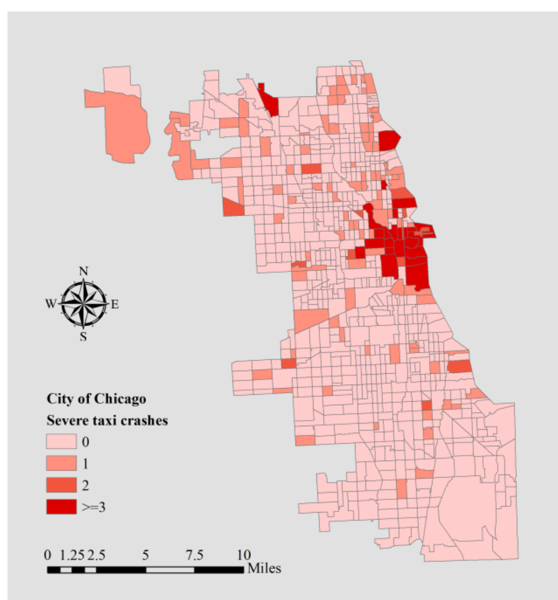
Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

suburban and rural areas of Chicago. Ride-hailing services are more likely to operate in such areas than taxis, partially due to the high-efficiency matching algorithms between potential ride-hailing drivers and passengers (Acheampong *et al.* 2020). Further, there are inherent correlations for different types of crashes (ride-hailing and taxi crashes by severity) in Figure 2. For example, minor ride-hailing crashes are distributed similarly to minor taxi crashes, especially in the central and northwestern areas of Chicago.

The average number of the natural logarithm of VMT for all motor vehicles was approximately 15.54, much higher than that for ride-hailing services (10.77) and taxis (7.03) in Table 1. In particular, the mode-specific exposures for ride-hailing services and taxis also have different spatial distributions in Figure 3. More specially, ride-hailing services would operate in more extensive service areas than taxis, especially in suburban and rural areas. Besides the mode-specific exposures, other transportation, land use, and demographic factors were assumed to be the same for ride-hailing and taxi crashes in 2019.



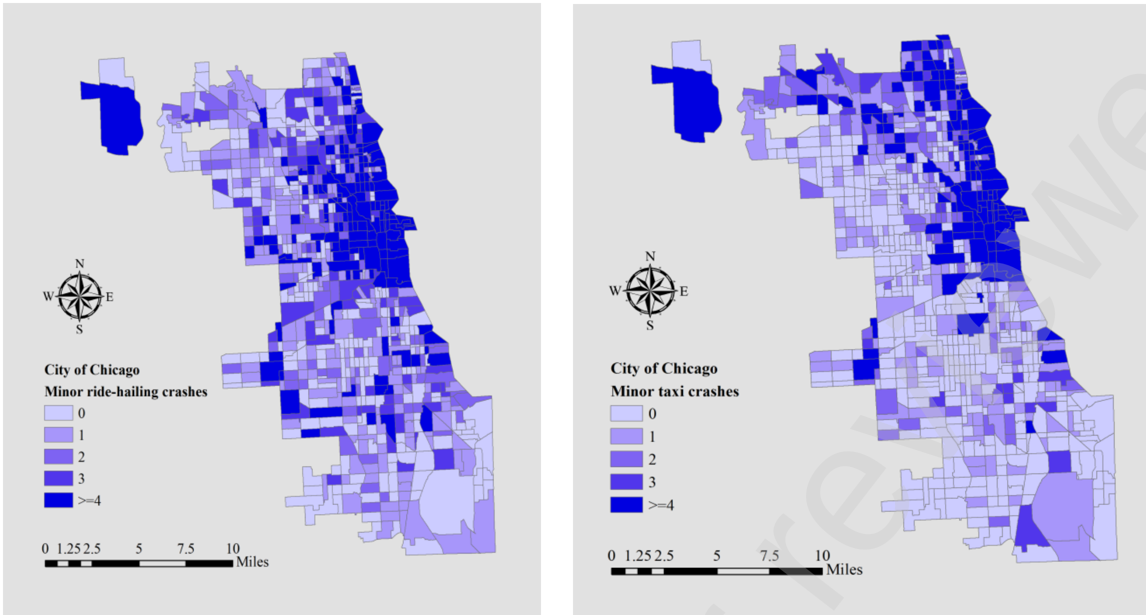
(a) Severe ride-hailing crashes



(b) Severe taxi crashes

Preprint not peer reviewed

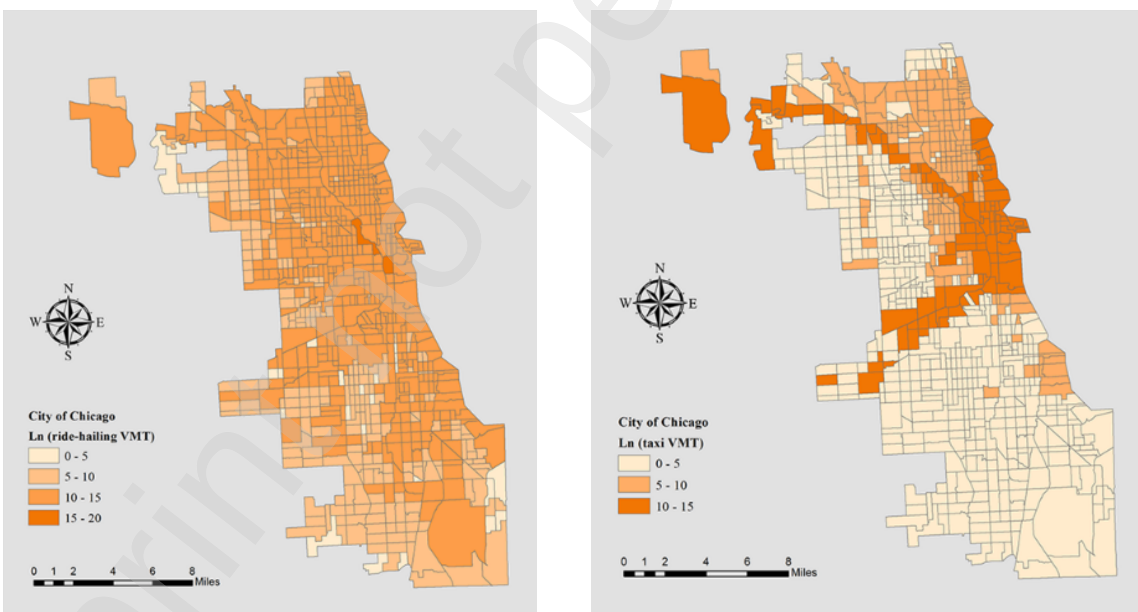
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669



(c) Minor ride-hailing crashes

(d) Minor taxi crashes

Figure 2. Spatial distributions of mode-specific crashes



(a) Ln(ride-hailing VMT)

(b) Ln(taxi VMT)

Figure 3. Spatial distributions of mode-specific exposures

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

1 Table 1 Descriptive analysis for the prepared data (N = 801)

Variables	Definition	Mean	SD
Crashes			
Severe ride-hailing crashes	Severe crashes involving ride-hailing services	0.47	1.24
Minor ride-hailing crashes	Minor crashes involving ride-hailing services	3.40	10.36
Severe taxi crashes	Severe crashes involving ride-hailing services	0.33	1.56
Minor taxi crashes	Minor crashes involving ride-hailing services	3.55	17.45
Transportation factors			
Ln (VMT)	Natural logarithm of exposures (vehicle. mile) for all motor vehicles	15.54	1.30
Ln (ride-hailing VMT)	Natural logarithm of exposures (vehicle. mile) for ride-hailing services	10.77	1.74
Ln (taxi VMT)	Natural logarithm of exposures (vehicle. mile) for taxis	7.03	3.29
Bus stop number	Number of bus stops	11.82	8.59
Metro station number	Number of metro stations	0.20	0.63
Crosswalk number	Number of crosswalks	2.84	8.37
Stop sign number	Number of stop signs	2.95	6.19
Traffic signal number	Number of traffic signals for vehicles	4.70	6.64
Land use factors			
Commercial ratio	Ratio of commercial areas to the whole area	0.09	0.21
Residential ratio	Ratio of residential areas to the whole area	0.29	0.39
Green space ratio	Ratio of green space areas to the whole area	0.35	0.43
Other ratio	Ratio of other areas to the whole area	0.27	0.42
Number of education sites	Number of kindergartens, schools, colleges, and universities	0.59	0.95

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Number of alcohol-related sites	Number of bars, beverages, nightclubs, and pubs	0.61	1.68
Demographic factors			
Population	Number of populations in thousands	3.43	1.87
Population younger than 18	Number of population younger than 18 in thousands	0.72	0.47
Median household income	Median household income (in 10^4 USD)	3.39	1.81
Number of house units	Number of house units	3.32	1.82
Transit ratio	Ratio of commuters by transit	0.29	0.13
Bike ratio	Ratio of commuters by bike	0.02	0.02
Walk ratio	Ratio of commuters by walking	0.05	0.08

Note: SD denotes the standard deviation.

4. Methodology

4.1. Model specification

4.1.1. Multivariate conditional autoregressive (MVCAR) model

The observed crash frequency y_i^k at the site i for the crash type k ($i = 1, 2, ..., n$, n denotes the total number of census tracts in Chicago) is commonly assumed to follow a Poisson distribution with the mean value λ_i^k. It should be noted that there are four types of crashes, including severe ride-hailing crashes, minor ride-hailing crashes, severe taxi crashes, and minor taxi crashes. To be more specific, the probability of observed crash frequency y_i^k at site i for crash type k can be given by Equation (1):

$$P(y_i^k | \lambda_i^k) = \frac{e^{-\lambda_i^k} \lambda_i^{k\, y_i^k}}{y_i^k\,!} \quad (1)$$

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Let $ModeVMT_i^k$ at site i denotes the mode-specific exposures for ride-hailing crashes and taxi crashes, including $RHVMT_i^k$ for ride-hailing crashes and $TXVMT_i^k$ for taxi crashes. The Poisson parameter λ_i^k can be specified by the estimated mode-specific exposures $ModeVMT_i^k$ and a series of other explanatory variables X_{pi}^k (i.e., other transportation factors, land use factors, and demographic factors) in Equation (2):

$$\ln\left(\lambda_i^k\right) = \beta_0^k + \beta_1^k \ln(ModeVMT_i^k) + \sum_{p=2}^{P} \beta_p^k X_{pi}^k + \varepsilon_i \tag{2}$$

where, $p = 2,3,...,P$, P is the total number of explanatory variables except for the estimated mode-specific exposures $ModeVMT_i^k$. β_0^k, β_1^k, and β_p^k are the regression coefficients to be estimated. In addition, $\exp\left(\varepsilon_i\right)$ is assumed to be gamma-distributed with mean one and variance α^2 across different sites to address the over-dispersion issues, formulating commonly used negative binomial (NB) models in crash safety studies (Lord and Mannering 2010, Zhai *et al.* 2022, Zhang *et al.* 2022).

The proposed MVCAR is developed to capture the potential spatial correlations of neighboring sites and inherent correlations across various types of crashes simultaneously by adding a multivariate spatial latent effect term, S_{ki} into Equation (3) (Palmí-Perales *et al.* 2019, Xie *et al.* 2019):

$$\ln\left(\lambda_i^k\right) = \beta_0^k + \beta_1^k \ln(ModeVMT_i^k) + \sum_{p=2}^{P} \beta_p^k X_{pi}^k + S_{ki} + \varepsilon_i \tag{3}$$

The full conditional distribution of $\mathbf{S}_i = \left(S_{1i}, S_{2i}, ..., S_{Ki}\right)'$ follows a K-dimensional multivariate normal distribution (Thomas *et al.* 2004) in Equation (4):

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

$$\mathbf{S}_i \mid \mathbf{S}_{-i} \sim MVN_K\left(\sum_{j\neq i}\frac{w_{ij}}{w_{i+}}\mathbf{S}_j, \frac{\Omega}{w_{i+}}\right) \tag{4}$$

where, $\mathbf{S}_{-i}$ is the set of $\mathbf{S}_j$ for any $j \neq i$. w_{ij} denotes the spatial correlations (weights) between site i and site j. Specifically, $w_{ij} = 1$ if site i and site j are adjacent and $w_{ij} = 0$ otherwise. w_{i+} is the aggregation of spatial weights for site i, with $w_{i+} = \sum_{j=1}^{n} w_{ij}$. Ω is the variance-covariance matrix for the spatial and inherent correlations in Equation (5):

$$\Omega = \begin{pmatrix} \sigma_{S11}^2 & \cdots & \sigma_{S1K}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{SK1}^2 & \cdots & \sigma_{SKK}^2 \end{pmatrix} \tag{5}$$

Diagonal elements of Ω denote the conditional variance of the spatial correlations of neighboring sites, and the off-diagonal elements represent the conditional variance of the inherent correlations across various types of crashes for the same site (Thomas *et al.* 2004, Palmí-Perales *et al.* 2019).

4.1.2. Measurement errors in mode-specific exposures

Without detailed vehicle trajectory data for ride-hailing services and taxis, potential differences between actual and expected mode-specific exposures should be appropriately considered in the modeling process. The classical measurement error structure in log-scale mode-specific exposures is described (Muff *et al.* 2015) in Equation (6):

$$\ln\left(ModeVMT_i^k\right) = \ln\left(ModeVMT_i^{k*}\right) + \tau_i \tag{6}$$

where $ModeVMT_i^{k*}$ is the unknown actual mode-specific exposures of ride-hailing services and taxis. Because mode-specific exposures (i.e., ride-hailing VMT and taxi VMT)

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Preprint not peer reviewed

are estimated in the same logic, the measurement error term, τ_i, is assumed to be normally distributed with zero mean and Gaussian noise σ_{τ_i} across different sites.

After integrating the measurement error structure into the MVCAR model, the proposed MVCARME model can be developed as Equation (7):

$$\ln\left(\lambda_i^k\right) = \beta_0^k + \beta_1^k \ln\left(ModeVMT_i^{k*}\right) + \sum_{p=2}^{P} \beta_p^k X_{pi}^k + S_{ki} + \varepsilon_i \tag{7}$$

4.2. Model assessment

The deviance information criterion (DIC) has been widely used to measure model fitting and complexity in Bayesian modeling assessments (Spiegelhalter *et al.* 2002). Specifically, the DIC can be estimated as Equation (8):

$$DIC = \overline{D(\theta)} + p_D \tag{8}$$

where, $\overline{D(\theta)}$ is the posterior mean of the deviance of the estimated parameters θ, which can be considered as a Bayesian measure about the goodness-of-fit. p_D denotes the effective number of parameters and can be taken as a measure of model complexity. As indicated by Lunn *et al.* (2013), the models with DIC values smaller than five are considered to have the same fitness and complexity. A smaller DIC is associated with better statistical performance.

Besides DIC, the widely applicable information criterion (WAIC) is also used to assess Bayesian model fitness, with simpler estimates of predictive errors but requiring additional computational steps (Vehtari *et al.* 2017). The WAIC could be estimated by Equation (9)

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

where $lppd$ denotes the log point-wise predictive density to measure the prediction accuracy and p_{WAIC_2} sum the variance of individual terms in the log predictive density to adjust for overfitting (Gelman *et al.* 2021). Similarly, Models with smaller WAIC values tend to be preferred.

$$WAIC = -2(lppd - p_{WAIC_2}) \tag{9}$$

In addition, the mean absolute error (MAE) and root-mean-square error (RMSE) are also considered to measure goodness-of-fit for the Bayesian inference in this study.

4.3. Bayesian estimation with INLA

All the crash frequency models above (i.e., NB, MVCAR, and MVCARME) are estimated in the full Bayesian framework. The Bayesian method combines prior distributions with a likelihood function obtained to create posterior distributions as estimates (Cui and Xie 2021, Gelman *et al.* 2021). The theoretical framework for Bayesian inference can be described as Equation (10):

$$p(\theta \mid y) \propto L\left(y|\theta\right) p\left(\theta\right) \tag{10}$$

where, y is the vector of observed crash frequencies for ride-hailing crashes and taxi crashes by severity; θ has been defined previously in Equation 8; $p(\theta \mid y)$ denotes the posterior distribution of θ given y; $L\left(y|\theta\right)$ is likelihood function; and $p\left(\theta\right)$ is the prior distribution of θ.

In practice, the Markov Chain Monte Carlo (MCMC) algorithm is commonly used to estimate parameters of Bayesian models (Thomas *et al.* 2004, Lunn *et al.* 2013). Despite

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

the flexibility of Bayesian inference, the computational burden of the MCMC algorithm is tremendous, especially when some variables are with no-Gaussian distributions (Cui and Xie 2021). Such an inefficient or time-consuming algorithm will not be applied to Bayesian estimation in this study.

Alternatively, the INLA method proposed by Rue *et al.* (2009) is more efficient in Bayesian inference than the MCMC algorithm, especially for complicated Bayesian models. The Laplace approximation technique is the most essential component of the INLA method, which can approximate any distributions with Gaussian distributions and thus improve estimation efficiency. Under the Bayesian framework, the posterior marginal distributions of interest can be expressed as Equation (11),

$$p(\theta|y) = \int p(\theta,\pi|y)\,d\pi = \int \exp\left(\log\left(p(\theta,\pi|y)\right)\right)d\pi \tag{11}$$

where, π is the vector of hyperparameters; $\log\left(p(\theta,\pi|y)\right)$ can be represented by a Taylor series expansion; thus $p(\theta|y)$ can be turned into Equation (12),

$$p(\theta|y) \approx \exp\left(\log\left(p\left(\theta,\pi^{*}|y\right)\right)\right)\int \exp\left(-\frac{\left(p(\theta,\pi|y)-p\left(\theta,\pi^{*}|y\right)\right)^{2}}{2\sigma^{2*}}\right)d\pi \tag{12}$$

where, $\pi^{*} = argmax_{\pi}\log\left(p(\theta,\pi|y)\right)$; $\sigma^{2*} = -1/\frac{\partial}{\partial\pi^{2}}\Big|\pi=\pi^{*}$. At last, inside the integration is normally distributed with mean π^{*} and variance σ^{2*}, which significantly improves estimation efficiency for Bayesian inference (Cui and Xie 2021). To be more specific, the MVCAR components are developed by the INLAMSM package (Palmí-Perales *et al.*

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

2019), and the classical measurement error effects are integrated into the MVCAR model by the latent effect named 'mec' in the INLA package (Muff *et al.* 2015).

5. Results and discussions

5.1. Modeling results

We developed three types of crash frequency models specified above to understand ride-hailing crashes and taxi crashes, involving the NB models, the MVCAR model, and the MVCARME model. Specifically, the first (NB) models were developed by the INLA package (Rue *et al.* 2009) while the latter two models, MVCAR and MCVARME models, were estimated by the INLMSM package and the INLA package simultaneously (Rue *et al.* 2009, Muff *et al.* 2015, Palmí-Perales *et al.* 2019). For variable selections, insignificant variables were removed if the variables were insignificant at the 0.05 significance level for all types of crashes (i.e., severe ride-hailing crashes, minor ride-hailing crashes, severe taxi crashes, and minor taxi crashes). Variance inflation factors (VIFs) were used to examine the multicollinearity problems in modeling all types of crashes, respectively. Generally, a VIF value lower than five is acceptable in statistical models (O'brien 2007). Table 2 shows no multicollinearity issues are detected due to lower VIF values.

For comparison, Table 3 summarizes the goodness-of-fit values for the crash frequency models above (i.e., NB, MVCAR, and MVCARME). The MVCARME outperforms the other models with the lowest DIC, WAIC, MAE, and RMSE values. In addition, the crash frequency modeling is improved by accounting for the spatial and inherent correlations because of the significantly reduced DIC, WAIC, MAE, and RMSE values from NB to

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

MVCAR, which validate the existence of the spatial and inherent correlations. Then, the crash frequency modeling is further improved after considering the measurement errors in mode-specific exposures due to slightly lower DIC, WAIC, MAE, and RMSE values of the MVCARME model compared to the MVCAR model. Such improvements in the goodness-of-fit should be attributed to the incorporation of measurement errors in mode-specific exposures.

Table 2 Results of the multicollinear test for all types of crashes

	VIF			
Variables	Severe ride-hailing crashes	Minor ride-hailing crashes	Severe taxi crashes	Minor taxi crashes
Transportation factors				
Ln (VMT)	2.15	1.89	2.35	1.59
Ln (mode-specific VMT)	3.59	2.80	3.96	1.73
Traffic signal number	2.59	1.95	4.20	1.85
Land use factors				
Number of education sites	1.06	1.03	1.51	1.04
Number of alcohol-related sites	2.36	1.72	3.08	1.64
Demographic factors				
Median household income	2.25	1.94	2.39	1.55
Transit ratio	1.46	1.36	1.51	1.13
Walk ratio	2.34	1.65	3.44	1.45

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Table 3 Summary of model performance

Goodness-of-fit	NB	MVCAR	MVCARME
DIC	7,199.90	6,812.06	6,739.61
WAIC	7,204.71	6,800.57	6,751.56
MAE	2.35	1.26	1.18
RMSE	51.11	8.95	7.99

Table 4 presents the modeling results of the MVCARME model. For transportation factors, a one-percent increase in VMT was found to be positively associated with a 0.30% increase in severe ride-hailing crashes, a 0.22% increase in minor ride-hailing crashes, a 0.31% increase in severe taxi crashes, and a 0.29% increase in minor taxi crashes. The positive correlations between crash frequencies and VMT can be commonly observed in previous studies (National Research Council 2010, Zhai *et al.* 2022). Mode-specific VMT was also found to impact severe ride-hailing and taxi crashes in similar intensity positively. A one-percent increase in mode-specific VMT would increase severe ride-hailing crashes by 0.11% and severe taxi crashes by 0.10%. In contrast, a one-percent increase in mode-specific VMT would increase minor ride-hailing crashes by 0.17% while minor taxi crashes by 0.07%. Such positive relationships are consistent with Ma *et al.* (2019), where taxi VMT was also positively associated with taxi crashes. In particular, a large variance (0.23) for measurement errors of mode-specific VMT emphasizes the importance (i.e., mitigating the impacts of the uncertainty of mode-specific VMT) of incorporating the measurement error structure in modeling ride-hailing and taxi crashes; otherwise, the modeling results may lead to biased inferences. Traffic signal numbers were also positively associated with ride-hailing and taxi crashes. One possible reason is the complicated vehicle movements at

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

intersections with traffic signal controls, especially with higher numbers of signal phases (Chin and Quddus 2003). Another possible reason is the potential interactions between ride-hailing or taxi vehicles and vulnerable road users (such as pedestrians and cyclists) at intersections even controlled by traffic signals.

Regarding land use factors, educational sites were also positively correlated with ride-hailing crashes and taxi crashes. A possible reason is that vulnerable road users, especially teenagers, are likely to interact more with motor vehicles at educational sites (Warsh *et al.* 2009). High proportions of inexperienced drivers around the educational sites might induce more dangerous interactions with ride-hailing and taxi drivers (Mitra and Washington 2012a). Similarly, more alcohol-related sites were positively associated with ride-hailing and taxi crashes. Intoxicated vulnerable road users (i.e., pedestrians and cyclists) and drivers are more likely to be observed in alcohol-related sites. A higher likelihood of crash occurrence is positively associated with the reduced ability to detect potential collision risks and decreased reaction time to unexpected events (Mitra and Washington 2012a).

In terms of demographic factors, median household income was found to be positively correlated with ride-hailing crashes and taxi crashes. A higher median household income would positively affect private vehicle use and ownership, thus inducing more interactions between private vehicles and ride-hailing or taxi vehicles. The findings above were consistent with previous studies (Xie *et al.* 2019, Xu *et al.* 2022). Additionally, a higher transit or walk ratio for commuters would increase the likelihood of minor ride-hailing

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

crashes and taxi crashes because of the increased exposure to transit vehicles and pedestrians (Chen and Zhou 2016, Ding *et al.* 2018b, Mohammadi *et al.* 2018).

Table 4 Modeling results of the MVCARME model

Variables	Severe ride-hailing crashes		Minor ride-hailing crashes		Severe taxi crashes		Minor taxi crashes	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Intercept	-8.49*	0.80	-6.61*	0.81	-9.06*	0.84	-6.53*	0.72
Transportation factors								
Ln (VMT)	0.30*	0.05	0.22*	0.05	0.31*	0.05	0.29*	0.05
Ln (mode-specific VMT)	0.11*	0.03	0.17*	0.03	0.10*	0.03	0.07*	0.03
Traffic signal number	0.02*	0.01	0.02*	0.01	0.02*	0.01	0.03*	0.01
Land use factors								
Number of education sites	0.12*	0.04	0.08*	0.04	0.08*	0.04	0.09*	0.04
Number of alcohol-related sites	0.05*	0.02	0.07*	0.02	0.05*	0.02	0.06*	0.02
Demographic factors								
Median household income	0.08*	0.03	0.05*	0.03	0.10*	0.03	0.10*	0.03
Transit ratio	1.30*	0.52	1.08*	0.45	1.09*	0.52	1.06*	0.44
Walk ratio	0.84	0.82	1.38*	0.65	3.38*	0.73	1.90*	0.66

Notes:

1. Measure errors $\sigma_{\tau_i}^2 \sim N(0.23, 0.11)$
2. Dispersion $\alpha \sim N(0.14, 0.04)$
3. SD denotes the standard deviation.
4. * denotes 95% Bayesian credible interval

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

5.2. Safety comparison of ride-hailing crashes and taxi crashes

The statistical t-tests were conducted to assess the safety performance of ride-hailing services and taxis by comparing the coefficients of ln (mode-specific VMT). As presented in Table 5, no significant difference was found for the risks of severe injury crashes between ride-hailing services and taxis because of higher *p*-values (> 0.05): NB (*p*-value = 0.30), MVCAR (*p*-value = 0.86), and MVCARME (*p*-value = 0.81). In particular, by accounting for the spatial correlations, the inherent correlations, and the measurement errors, we found that a one percent increase in mode-specific VMT would increase severe ride-hailing crashes by 0.11% and severe taxi crashes by 0.10%. In addition, Table 5 also indicates that taxis are exposed to lower risks of minor injury crashes than ride-hailing services due to lower *p*-values (<0.05) for the three models above (i.e., NB, MVCAR, MVCARME). For the modeling results of MVCARME, a one percent increase in mode-specific exposures will induce a 0.17% increase in minor ride-hailing crashes and a 0.07% increase in taxi crashes.

In summary, taxis are associated with lower risks of minor injury crashes than ride-hailing services, even though there is no significant difference in safety performance for severe injury crashes. There are three possible reasons. First, ride-hailing drivers must interact with passengers via ride-hailing apps while driving to receive ride-hailing orders, pick up the targeted passengers, and follow the pre-planned routes. Such behaviors might distract the driver's visual attention and reduce the situation awareness, even if the presence of compensatory behaviors, such as speed reductions (Chen *et al.* 2022). Second, taxi drivers are professionally trained and more experienced, while ride-hailing drivers are semi-

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

professional or unprofessional drivers with less driving experience. For example, taxi drivers tend to have safer speed control because of their familiarity with road environments (Mi *et al.* 2021). Thirdly, taxi companies have implemented well-established regulations to improve the safety performance of taxi drivers. For instance, taxi drivers are guaranteed a minimum wage to reduce work-related stress, thus reducing the workload in driving and improving safety performance (Park *et al.* 2021). Taxi drivers are not allowed to work overtime to minimize the risks of fatigue driving and improve crash safety (Park *et al.* 2021).

Table 5 Safety comparison of ride-hailing crashes and taxi crashes

Models	Crash types	Ln (mode-specific VMT)		*T test*
		Mean	SD	*p*-value
NB	Severe ride-hailing crashes	0.14	0.06	0.30
	Severe taxi crashes	0.07	0.03	
	Minor ride-hailing crashes	0.26	0.03	< 0.01
	Minor taxi crashes	0.05	0.02	
MVCAR	Severe ride-hailing crashes	0.12	0.03	0.86
	Severe taxi crashes	0.11	0.03	
	Minor ride-hailing crashes	0.16	0.03	0.03
	Minor taxi crashes	0.08	0.02	
MVCARME	Severe ride-hailing crashes	0.11	0.03	0.81
	Severe taxi crashes	0.10	0.03	
	Minor ride-hailing crashes	0.17	0.03	0.02
	Minor taxi crashes	0.07	0.03	

Note: SD denotes the standard deviation.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

6. Conclusions

This study proposed a novel multivariate conditional autoregressive model with accommodation of exposure uncertainty to compare the safety performance of ride-hailing services and taxis. The proposed model can jointly account for the spatial correlations among different sites, the inherent correlations across different types of crashes, and measurement errors in mode-specific exposures. Mode-specific exposures at each census tract were estimated by trip assignments, where the shortest route paths for different origin-destination (OD) pairs are calculated based on the actual route networks in Chicago. The measurement error structure of the proposed model can mitigate the impact of the uncertainty in mode-specific exposures that stems from two sources: 1) ride-hailing or taxi drivers might have different preferences on routing choices spatially and temporally for the same OD pairs (Liu and Jiang 2022), and 2) the centroids of OD pairs were still not the actual trip origin or destination locations. Additionally, the proposed model has a multivariate spatial specification and can account for the spatial and inherent correlations among crash observations. The proposed model outperforms the other two alternatives (i.e., negative binomial models and the multivariate conditional autoregressive model) with the lowest deviance information criterion (DIC), Watanabe-Akaike information criterion (WAIC) value, mean absolute error (MAE), and root-mean-square error (RMSE).

Our proposed approach suggests that ride-hailing services and taxis have no significant difference in safety performance in terms of severe crashes. A one percent increase in mode-specific VMT was associated with a 0.11% increase in severe ride-hailing crashes and a 0.10% increase in severe taxi crashes. However, ride-hailing services are found to be

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

prone to a higher risk of minor injury crashes compared with taxis. More specially, a one percent increase in mode-specific VMT would increase minor ride-hailing crashes by 0.17% and minor taxi crashes by 0.07%.

Three potential reasons are summarized to explain the better safety performance of taxis compared to ride-hailing services, such as professionally trained and experienced taxi drivers and well-established regulations to reduce work stress and driver fatigue. Correspondingly, transportation network companies should consider how to minimize the visual interactions between drivers and ride-hailing apps while picking up or dropping off passengers. Voice interactions might be a possible alternative to visual interactions with more intelligent speech identification and response algorithms. Transportation network companies could also assign orders to drivers more familiar with the road environment and restrict the work time of ride-hailing drivers to reduce fatigue driving. On the other hand, government agencies could set a minimum requirement of driving experience for potential ride-hailing drivers and develop regulations to ensure that ride-hailing drivers are professionally trained.

There are also some limitations of this study. First, it was infeasible to obtain the actual mode-specific exposures because the high-resolution vehicle trajectory data are unavailable for privacy concerns. Instead, we estimated the mode-specific exposures by trip assignments and accounted for uncertain exposure in the proposed model. In addition, we only used one-year data to exclude the impacts of COVID-19 on driving behaviors (Dong *et al.* 2022) without looking into the safety performance of ride-hailing and taxi

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

drivers that might change over time. Thus, it would be valuable to revisit the problem when new data become available. Further, validating the findings in other study areas also deserves attention.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

Guocong Zhai: Methodology, Software, Validation, Data curation. **Kun Xie**: Conceptualization, Methodology, Software, Writing – review & editing. **Hong Yang**: Methodology, Writing – review & editing. **Di Yang**: Methodology, Writing – review & editing.

Acknowledgments

The work is partially funded by the Transportation Informatics Lab, Department of Civil and Environmental Engineering at Old Dominion University (ODU). The contents of this paper present the views of the authors, who are responsible for the facts and accuracy of the data presented herein. The contents of the paper do not reflect the official views or policies of the agencies.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

References

Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: Analyzing spatial effect and identifying significant factors. Transportation Research Record 1953 (1), 98-111.

Acheampong, R.A., Siiba, A., Okyere, D.K., Tuffour, J.P., 2020. Mobility-on-demand: An empirical study of internet-based ride-hailing adoption factors, travel characteristics and mode substitution effects. Transportation Research Part C: Emerging Technologies 115, 102638.

Barrios, J.M., Hochberg, Y.V., Yi, H., 2020. The cost of convenience: Ridehailing and traffic fatalities. Journal of Operations Management.

Bhowmik, T., Yasmin, S., Eluru, N., 2019. Do we need multivariate modeling approaches to model crash frequency by crash types? A panel mixed approach to modeling crash frequency by crash types. Analytic methods in accident research 24, 100107.

Brazil, N., Kirk, D., 2020. Ridehailing and alcohol-involved traffic fatalities in the united states: The average and heterogeneous association of uber. PLoS One 15 (9), e0238744.

Brazil, N., Kirk, D.S., 2016. Uber and metropolitan traffic fatalities in the united states. Am J Epidemiol 184 (3), 192-8.

Brown, A., 2022. Not all fees are created equal: Equity implications of ride-hail fee structures and revenues submitted to transport policy. Transport Policy.

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. Accid Anal Prev 93, 14-22.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Chen, P., Zhou, J., 2016. Effects of the built environment on automobile-involved pedestrian crash frequency and risk. Journal of Transport & Health 3 (4), 448-456.

Chen, T., Oviedo-Trespalacios, O., Sze, N., Chen, S., 2022. Distractions by work-related activities: The impact of ride-hailing app and radio system on male taxi drivers. Accident Analysis & Prevention 178, 106849.

Cheng, W., Gill, G.S., Dasu, M., Jia, X., 2018. An empirical evaluation of multivariate spatial crash frequency models. Accid Anal Prev 119, 290-306.

Chicago Department of Business Affairs & Consumer Protection, 2019a. Taxi trips - 2019. In: Chicago, C.O. ed.

Chicago Department of Business Affairs & Consumer Protection, 2019b. Transportation network providers - trips - 2019. In: Chicago, C.O. ed.

Chicago Police Department, 2019a. Traffic crashes - crashes. In: Chicago, C.O. ed.

Chicago Police Department, 2019b. Traffic crashes - vehicles. In: Chicago, C.O. ed.

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. accident analysis & prevention 35 (2), 253-259.

Chiou, Y.-C., Fu, C., 2013. Modeling crash frequency and severity using multinomial-generalized poisson model with error components. Accident Analysis & Prevention 50, 73-82.

Chiou, Y.-C., Fu, C., Chih-Wei, H., 2014. Incorporating spatial dependence in simultaneously modeling crash frequency and severity. Analytic methods in accident research 2, 1-11.

Cressie, N., 2015. Statistics for spatial data John Wiley & Sons.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Cui, H., Xie, K., 2021. An accelerated hierarchical bayesian crash frequency model with accommodation of spatiotemporal interactions. Accid Anal Prev 153, 106018.

Dills, A.K., Mulholland, S.E., 2018. Ride-sharing, fatal crashes, and crime. Southern Economic Journal 84 (4), 965-991.

Ding, C., Chen, P., Jiao, J., 2018a. Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: A machine learning approach. Accid Anal Prev 112, 116-126.

Ding, C., Chen, P., Jiao, J., 2018b. Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: A machine learning approach. Accident Analysis & Prevention 112, 116-126.

Dong, X., Xie, K., Yang, H., 2022. How did covid-19 impact driving behaviors and crash severity? A multigroup structural equation modeling. Accident Analysis & Prevention 172, 106687.

Gaweesh, S.M., Ahmed, M.M., Piccorelli, A.V., 2019. Developing crash prediction models using parametric and nonparametric approaches for rural mountainous freeways: A case study on wyoming interstate 80. Accident Analysis & Prevention 123, 176-189.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2021. Bayesian data analysis, the third edition Chapman and Hall/CRC.

Giraud, T., 2022. Osrm: Interface between r and the openstreetmap-based routing service osrm. The Journal of Open Source Software.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Greenwood, B.N., Wattal, S., 2017. Show me the way to go home: An empirical investigation of ride-sharing and alcohol related motor vehicle fatalities. MIS Q. 41 (1), 163-187.

Hou, Q., Tarko, A.P., Meng, X., 2018. Analyzing crash frequency in freeway tunnels: A correlated random parameters approach. Accid Anal Prev 111, 94-100.

Illinois Department of Transportation, 2019. Traffic volumes. In: Idot ed.

Jiao, J., Bischak, C., Hyden, S., 2020. The impact of shared mobility on trip generation behavior in the us: Findings from the 2017 national household travel survey. Travel Behaviour and Society 19, 1-7.

Kabir, R., Remias, S.M., Lavrenz, S.M., Waddell, J., 2021. Assessing the impact of traffic signal performance on crash frequency for signalized intersections along urban arterials: A random parameter modeling approach. Accident Analysis & Prevention 149, 105868.

Kamel, M.B., Sayed, T., 2020. Cyclist-vehicle crash modeling with measurement error in traffic exposure. Accid Anal Prev 144, 105612.

Kamel, M.B., Sayed, T., Osama, A., 2019. Accounting for mediation in cyclist-vehicle crash models: A bayesian mediation analysis approach. Accid Anal Prev 131, 122-130.

Kirk, D.S., Cavalli, N., Brazil, N., 2020. The implications of ridehailing for risky driving and road accident injuries and fatalities. Soc Sci Med 250, 112793.

Liu, S., Jiang, H., 2022. Personalized route recommendation for ride-hailing with deep inverse reinforcement learning and real-time traffic conditions. Transportation Research Part E: Logistics and Transportation Review 164, 102780.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice 44 (5), 291-305.

Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2013. The bugs book. A Practical Introduction to Bayesian Analysis, Chapman Hall, London.

Ma, Q., Yang, H., Xie, K., Wang, Z., Hu, X., 2019. Taxicab crashes modeling with informative spatial autocorrelation. Accid Anal Prev 131, 297-307.

Mao, H., Deng, X., Jiang, H., Shi, L., Li, H., Tuo, L., Shi, D., Guo, F., 2021. Driving safety assessment for ride-hailing drivers. Accid Anal Prev 149, 105574.

Marshall, W.E., Garrick, N.W., 2011. Does street network design affect traffic safety? Accident Analysis & Prevention 43 (3), 769-781.

Mi, X., Dong, C., Li, N., Lin, Y., Shao, C., Fan, B., 2021. Operating safety evaluation of battery-electric taxi based on spatio-temporal speed parameters. Sustainability 13 (23), 13446.

Mitra, S., Washington, S., 2012a. On the significance of omitted variables in intersection crash modeling. Accident Analysis & Prevention 49, 439-448.

Mitra, S., Washington, S., 2012b. On the significance of omitted variables in intersection crash modeling. Accid Anal Prev 49, 439-48.

Mohammadi, M., Shafabakhsh, G., Naderan, A., 2018. Effects of modal shares on crash frequencies at aggregate level. Accident Analysis & Prevention 120, 295-303.

Mohammadi, M.A., Samaranayake, V., Bham, G.H., 2014. Crash frequency modeling using negative binomial models: An application of generalized estimating Equation to longitudinal data. Analytic Methods in Accident Research 2, 52-69.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Muff, S., Riebler, A., Held, L., Rue, H., Saner, P., 2015. Bayesian analysis of measurement error models using integrated nested laplace approximations. Journal of the Royal Statistical Society: Series C (Applied Statistics) 64 (2), 231-252.

National Research Council, 2010. Highway safety manual AASHTO.

O'brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Quality & quantity 41 (5), 673-690.

Openstreetmap, 2019a. Land use characteristics. In: Openstreetmap ed.

Openstreetmap, 2019b. Points of Interest (POI).

Palmí-Perales, F., Gómez-Rubio, V., Martinez-Beneito, M.A., 2019. Bayesian multivariate spatial models for lattice data with inla. arXiv preprint arXiv:1909.10804.

Park, J., Lee, S., Oh, C., Choe, B., 2021. A data mining approach to deriving safety policy implications for taxi drivers. Journal of safety research 76, 238-247.

Pebesma, E.J., 2018. Simple features for r: Standardized support for spatial vector data. R J. 10 (1), 439.

Rayle, L., Dai, D., Chan, N., Cervero, R., Shaheen, S., 2016. Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in san francisco. Transport Policy 45, 168-178.

Rue, H., Martino, S., Chopin, N., 2009. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. Journal of the royal statistical society: Series b (statistical methodology) 71 (2), 319-392.

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 & Prevention 118, 166-177.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Spiegelhalter, D.J., Best, N.G., Carlin, B.R., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B-Statistical Methodology 64, 583-616.

Thomas, A., Best, N., Lunn, D., Arnold, R., Spiegelhalter, D., 2004. Geobugs user manual. Cambridge: Medical Research Council Biostatistics Unit.

Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Safety science 50 (4), 1141-1151.

Vehtari, A., Gelman, A., Gabry, J., 2017. Practical bayesian model evaluation using leave-one-out cross-validation and waic. Statistics and computing 27 (5), 1413-1432.

Walker, K., Herman, M., Eberwein, K., Walker, M.K., 2021. Package 'tidycensus'. MIT.

Wang, Y., Kockelman, K.M., 2013. A poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accid Anal Prev 60, 71-84.

Warsh, J., Rothman, L., Slater, M., Steverango, C., Howard, A., 2009. Are school zones effective? An examination of motor vehicle versus child pedestrian crashes near schools. Injury prevention 15 (4), 226-229.

Wei, F., Lovegrove, G., 2013. An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. Accident Analysis & Prevention 61, 129-137.

Xie, K., Ozbay, K., Yang, H., 2015a. Spatial analysis of highway incident durations in the context of hurricane sandy. Accident Analysis & Prevention 74, 77-86.

Xie, K., Ozbay, K., Yang, H., 2019. A multivariate spatial approach to model crash counts by injury severity. Accid Anal Prev 122, 189-198.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Xie, K., Ozbay, K., Yang, H., Holguín-Veras, J., Morgul, E.F., 2015b. Modeling safety impacts of off-hour delivery programs in urban areas. Transportation research record 2478 (1), 19-27.

Xie, K., Ozbay, K., Yang, H., Holguín-Veras, J., Morgul, E.F., 2015c. Modeling safety impacts of off-hour delivery programs in urban areas. Transportation Research Record: Journal of the Transportation Research Board 2478 (1), 19-27.

Xie, K., Wang, X., Ozbay, K., Yang, H., 2014. Crash frequency modeling for signalized intersections in a high-density urban road network. Analytic methods in accident research 2, 39-51.

Xu, P., Bai, L., Pei, X., Wong, S.C., Zhou, H., 2022. Uncertainty matters: Bayesian modeling of bicycle crashes with incomplete exposure data. Accid Anal Prev 165, 106518.

Yang, D., Xie, K., Ozbay, K., Yang, H., 2021. Fusing crash data and surrogate safety measures for safety assessment: Development of a structural equation model with conditional autoregressive spatial effect and random parameters. Accid Anal Prev 152, 105971.

Yang, D., Xie, K., Ozbay, K., Yang, H., Budnick, N., 2019. Modeling of time-dependent safety performance using anonymized and aggregated smartphone-based dangerous driving event data. Accident Analysis & Prevention 132, 105286.

Yang, H., Wang, Z., Xie, K., Ozbay, K., Imprialou, M., 2018. Methodological evolution and frontiers of identifying, modeling and preventing secondary crashes on highways. Accident Analysis & Prevention 117, 40-54.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669

Yang, H., Zhai, G., Yang, L., Xie, K., 2022. How does the suspension of ride-sourcing affect the transportation system and environment? Transportation Research Part D: Transport and Environment 102, 103131.

Zhai, G., Xie, K., Yang, D., Yang, H., 2022. Assessing the safety effectiveness of citywide speed limit reduction: A causal inference approach integrating propensity score matching and spatial difference-in-differences. Transportation Research Part A: Policy and Practice 157, 94-106.

Zhang, Z., Zhai, G., Xie, K., Xiao, F., 2022. Exploring the nonlinear effects of ridesharing on public transit usage: A case study of San Diego. Journal of Transport Geography 104, 103449.

Ziakopoulos, A., Yannis, G., 2020. A review of spatial approaches in road safety. Accid Anal Prev 135, 105323.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4334669