Identifying regions of excess injury risks associated with distracted driving: A case study in Central Ohio, USA

Youngbin Lym [a,*], Seunghoon Kim [b,**], Ki-Jung Kim [c]

[a] Department of Public Administration, Kookmin University, 77 Jeongneung-ro, Seongbuk-gu, Seoul, 02707, Republic of Korea
[b] National Infrastructure Research Division Korea Research Institute for Human Settlements (KRIHS) 5 Gukchargyeonguwon-ro, Sejong-si, 30147, Republic of Korea
[c] Department of Smart Car Engineering, Doowon Technical University 159 Jurawu-gil, Paju-eup, Paju-si, Gyeonggi-do, 10838, Republic of Korea

ARTICLE INFO

ABSTRACT

This study examines the latent influence of spatial locations on the relative risks of crash injuries associated with distracted driving (DD) and identifies regions of excess risks for policy intervention. Using a sample of aggregated injury and fatal DD crash records for the period 2015–2019 across 1,024 census block groups in Central Ohio (i. e., the Columbus Metropolitan Area) in the United States, we investigate the role of latent effects along with several covariates such as land-use mix, sociodemographic features, and the built environment. To this end, we specifically leverage a full Bayesian hierarchical formulation with conditional autoregressive priors to account for uncertainty (i.e., spatially structured random effects) stemming from adjacent census block groups. Furthermore, we consider uncorrelated random effects from upper-level administrative units within which each block group is nested (i.e., census tracts and counties). Our analysis reveals that (1) addressing spatial correlation improves the model's performance, (2) block-group-level variability substantially explains the residual random fluctuation, and (3) intersection density appears negatively associated with the relative risks of crash injuries, while more diversified land use can increase injury risk. Based on these findings, we present spatial clusters with twice the relative risks compared to other block groups, suggesting that policies be devised to mitigate severe injuries due to DD and therefore enhance public health.

1. Introduction

The National Highway Traffic Safety Administration (NHTSA) (2020) defines distracted driving (DD) as driving that involves any activity (e.g., texting, cellphone usage, and drinking or eating) that takes the driver's attention away from the primary task, that is, safe driving. Due to its negative consequences, DD-related transportation safety seemed to gain wider attention among scholars, planners, government agencies, and policymakers as well as health officials (ODOT, 2019; CDC, 2020). Chen and Lym (2021) showed that vehicle crashes associated with DD are more likely to lead to severe outcomes (i.e., injuries rather than property damage only) compared to their non-distraction-affected counterparts. In addition, Fatmi and Habib (2019) found that certain built-environment features such as higher land-use mix, higher population density, and longer sidewalks could mitigate the severity of crashes due to DD. Chen and Lym (2021) showed the negative influence of roundabouts on minimizing the severity of DD

crashes in Ohio. From a different perspective, the effects of the built environment on DD crash risk reduction can be explained by drivers' behavioral responses, as they are inclined to adjust their behavior with higher driving complexity (e.g., driving under demanding roadway environments) (Oviedo-Trespalacios et al., 2019, 2020).

A number of studies have investigated various facets of DD, such as drivers' behavioral responses (e.g., by teenagers, age and execution, and young adults) and risk compensation (Gershon et al., 2019; Ortiz et al., 2018; Oviedo-Trespalacios et al., 2019), driving performance under experimental settings (Haque et al., 2016), the influence of the built environment on crash frequency and severity, and uncertainty from unobserved heterogeneity (e.g., Lym & Chen, 2020; Oviedo-Trespalacios et al., 2020). However, we believe that the literature has yet to fully consider unobserved heterogeneity in DD crashes. Hence, this research investigates the effects of unobserved heterogeneity on vehicle crashes associated with DD, focusing on latent influences from spatial alignments of administrative units (Dupont et al., 2013; Flask &

Schneider, 2013; Huang & Abdel-Aty, 2010). To this end, we aim to answer the following research questions:

1. Does the inclusion of spatially structured latent effects in the modeling process better explain crash injury risks due to DD?
2. How do we identify hot spots where policy intervention could mitigate the potential injury risks from DD in the Columbus Metropolitan Area (CMA), Ohio?

To address these questions, this study adopted a full Bayesian hierarchical approach with conditional autoregressive (CAR) priors (Besag et al., 1991), which allow us to account for latent spatial influences on DD injury risks along with other influential factors characterized by covariates.

The remainder of this paper is outlined as follows. Section 2 reviews the relevant literature, while Section 3 explains the adopted methodology. Section 4 presents and discusses our data. Section 5 evaluates our models and discusses our findings. Finally, Section 6 concludes our study and provides suggestions for future research.

2. Literature review

2.1. General description of distracted driving

DD refers to anything that diverts drivers' attention from safe driving (NHTSA, 2020). Recently, DD has become a major contributor to severe and fatal crashes, raising broad public awareness as well as research interests among scholars (Dingus et al., 2016, AAA Foundation for Traffic Safety [AAAFTS], 2018). For example, Kidd and Chaudhary's (2019) study on the prevalence of DD among Northern Virginia drivers showed an approximately 57% increase in the likelihood of cellphone manipulation in 2018 compared with 2014.

A considerable number of DD studies have attempted to clarify the association between drivers' behavioral response or performance and a series of distraction factors under simulated settings. Birrell and Young (2011) showed that smart driving information has no influence on driver distraction. Kircher and Ahlstrom's (2012) research on how tunnel design (tunnel wall color) and illumination affect drivers' exhibited behavior revealed the relative importance of light-colored walls over illumination. Haque et al. (2016) investigated young drivers' gap acceptance behavior at roundabouts under three mobile conditions (no phone, hands-free, and handheld conversation), identifying an acceptance of a smaller safety margin among distracted drivers. Kountouriotis and Merat (2016) found that driver distraction influenced speed, gaze patterns, and steering control, and that road geometry and the presence of other cars could interact with driver distraction. Using surveys and simulators, Oviedo-Trespalacios et al. (2016) and Papantoniou et al. (2017) provided systematic and comprehensive reviews of how distraction affects driving performance. Employing a different perspective, some scholars have considered whether age affects distraction. Gershon et al. (2017, 2019) suggested that teenagers are more likely to engage in secondary tasks associated with DD. Braitman and Braitman (2017), focusing on young adult drivers, conducted latent profile analysis to identify three profile classes of DD behaviors among them. Guo et al. (2017) and Ortiz et al. (2018) investigated the effects of age on crash risks related to DD, revealing that distraction more adversely affects senior drivers than middle-aged drivers.

In addition, many scholars have examined risk-compensating behavior, which refers to the avoidance of risky behavior to initiate secondary driving tasks (e.g., DD actions). For example, studies have used surveys (Parnell et al., 2020), simulations (Li et al., 2019), and empirical analysis (Oviedo-Trespalacios et al., 2017) to analyze factors leading to a driver's decision to engage in DD.

Meanwhile, other studies have examined various measures to reduce DD behavior and its consequences. Arnold et al. (2019, pp. 1–13) conducted a comprehensive review of the effectiveness of DD-related

legislation and regulations. They found that several measures are expected to reduce DD behavior and its associated crashes, but they rely on comparing pre- and post-intervention outcomes and controlling for potential confounders; in addition, finding appropriate control groups is often challenging. Most existing measures focus on cellphone use while behind the wheel, and studies based on these might not fully reflect the true nature, leading to biased outcomes and requiring further research. Meanwhile, Chen and Lym (2021) and Fatmi and Habib (2019) analyzed actual crash data and investigated statistical linkages between the role of the built environment and outcomes of DD crashes.

2.2. Understanding spatial correlation and unobserved heterogeneity

Much effort has been devoted to overcoming the challenges in understanding vehicle crashes from a statistical modeling standpoint (see comprehensive review by Lord & Mannering, 2010; Savolainen et al., 2011; Mannering et al., 2016). Since crashes tend to occur in certain locations (or coordinates), they inevitably entail spatial features. If a location is associated with many crashes, nearby locations may also share many of its characteristics (e.g., unobserved effects). For example, road segments or census units with higher frequencies of crashes may share certain geometric features or traffic flows or sociodemographic characteristics with adjacent ones. This leads to the potential influences of space between observations, and failure to account for these in the modeling process may result in incorrect inferences of covariate effects. Thus, it is reasonable to address the spatially structured random/latent effects in statistical models for vehicle crashes, which help visually identify regions of higher/excess crash risks to improve our understanding of relative risks of crashes.

Barua et al. (2014) found that crashes (collisions) are georeferenced (i.e., with known locations) and that the number of crashes observed in a certain period is likely to exhibit spatial dependency across different locations. Others have argued that properly incorporating spatial correlation into the crash modeling procedure substantially enhances the model's explanatory power (Aguero-Valverde, 2013; Aguero-Valverde & Jovanis, 2006; Boulieri et al., 2017; Liu & Sharma, 2018; Lym & Chen, 2020). In this regard, Cressie (2015) asserted that spatial random effects can successfully account for unknown and relevant covariates as well as unobserved heterogeneity.

Several traffic safety studies have adopted the Bayesian hierarchical approach as an analytic operational framework because of its flexibility in handling small samples (MacNab, 2004; Quddus, 2008). To address spatial correlation and unobserved heterogeneity, the Besag–York–Mollié (BYM or convolution) prior,[1] suggested by Besag et al. (1991), has been widely leveraged within transportation safety research (Aguero-Valverde & Jovanis, 2006; Barua et al., 2014; Boulieri et al., 2017; Lym & Chen, 2020; Ma et al., 2017; Zeng et al., 2019, 2020). Furthermore, Huang and Abdel-Aty (2010) argued that ignoring the multilevel structure of crash data can lead to unreliable parameter estimates and result in incorrect inferences about the crash generation process. This implies that properly addressing cross-group unobserved heterogeneity produces more trustworthy estimates and can improve the statistical model. In the same vein, Dupont et al. (2013) comprehensively discussed multilevel analysis in road safety research, and Flask and Schneider (2013) and Park et al. (2017) showed empirical evidence of its relative strength over its nonhierarchical counterpart.

2.3. Key features of this research

As noted, crash modeling studies have highlighted the importance of the proper inclusion of unobserved heterogeneity and random effects

---

[1] The Besag–York–Mollié (BYM) prior or model considers latent spatial effects as the sum of a spatially structured component for spatial smoothing and an additional unstructured random component for unobserved heterogeneity.

into statistical analysis; however, only a few have considered both spatial correlation and multilevel heterogeneity. To the best of our knowledge, research on crashes relative to DD that specifically emphasizes spatially structured random effects as well as latent influences from various levels of crash data aggregation is scant. Thus, we investigated the unique features of DD crashes along with the latent influences of space on the relative risks of DD. Through model-based approaches, we intend to provide different insights into vehicle crashes associated with DD, and our study can help fill the knowledge gap in traffic safety research.

3. Methodology

As a preliminary measure to formal statistical investigation, we conducted an exploratory data analysis to summarize data, find anomalies, and detect spatial patterns in the area of interest. To verify the existence of spatial correlation in DD-related injury crashes, we used a commonly adopted measure of spatial association, Moran's I, which tests for global spatial association among areal (lattice) units (in our study, census block groups) (Anselin, 1988). Transportation safety studies that have used this measure include Quddus (2008), Liu and Sharma (2018), and Lym and Chen (2020).

We also adopted a hierarchical Bayesian approach, which provides the flexibility to easily account for a multilevel cross-correlation alongside between-level spatial random effects. This means we considered the spatial correlation between census block groups (the lowest administrative unit in our study) and attempted to include latent effects from the alignment of higher-level administrative units such as census tracts and counties. To address between-level spatial random effects, we specifically followed a Bayesian disease mapping methodology, whose major focus is the relative risks of disease occurrence within small areas and their smoothing (Blangiardo & Cameletti, 2015; Lawson, 2018; Moraga, 2019). The relative risk of each unit (here, the number of crashes aggregated to each census block group) $\{\theta_i\}$, indexed by geographical unit i, was assumed to be random so that it (i.e., distribution of the relative risk θ_i) is generated from a set of hyperparameters.

The Poisson parameter, expected rate of occurrence $\{\lambda_i\}$, was decomposed into $\lambda_i = E_i \times \theta_i$. E_i is an offset to adjust for differences in traffic volume (i.e., vehicle miles traveled, VMT) across block groups, while θ_i denotes a relative risk of injury crashes. Conditional on the observed response Y_i, a natural framework that follows a full Bayesian hierarchical structure:

$$Data: \ Y_i|\lambda_i \ \sim \ Poisson(\lambda_i), \ \ \lambda_i = E_i \times \theta_i \quad (1)$$

$$Process: \ \theta_i|\ \psi \ \sim \ p(\bullet|\psi), \ \ \eta_i = \log(\theta_i)$$

$$\eta_i = \sum_{m=1}^{p} \beta_m X_{im} + s_i + \alpha_{k(i)} + \alpha_{l(k(i))}$$

$$\alpha_k \ \sim \ Normal \ (0, \tau_k)$$

$$\alpha_l \ \sim \ Normal \ (0, \tau_l)$$

$$s_i \ \sim \ BYM2 \ specification \ (\varphi, \tau_s)$$

$$Parameter: \ \psi \ \sim \ f(\bullet)$$

$$i = 1, 2, \cdots, n; \ k = 1, 2, \cdots, K; \ l = 1, 2, \cdots, L$$

where η_i corresponds to a linear predictor (i.e., logarithmic link function) that linearly relates a vector of predictors and random components to θ_i. Regarding the latent influences from coarser spatial units, we assumed two independent Gaussian distributions for the second (census tract, k) and third (county, l) levels. These can be characterized by two uncorrelated random effects (i.e., $\alpha_k \sim Normal \ (0, \tau_k)$ and $\alpha_l \sim Normal \ (0, \tau_l)$) after accounting for both fixed and spatially correlated random effects from the first-level observations (census block group, i). τ

denotes the precision parameter, which is an inverse of the variance of Gaussian distribution (i.e., $\tau^{-1} = \sigma^2$; here, τ_k and τ_l denote the precisions of census tracts and counties, respectively), for unobserved heterogeneity. $p(\bullet|\psi)$ and $f(\bullet)$ are probability distributions that generate η_i and (hyper)parameters ψ, respectively.

Meanwhile, we paid particular attention to the influences from the observation level (census block group). Fixed effects (coefficients) are estimated by a series of covariates at census block groups (i.e., $X_i = (X_{i1}, X_{i2}, X_{i3}, \cdots, X_{ip})^T$) as $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \cdots, \beta_p)^T$. For the block-group-level random effects, we employed the BYM2 model proposed by Riebler et al. (2016) and Simpson et al. (2017), which was modified (reparametrized) from the BYM (convolution) model (Besag et al., 1991). When considering priors for the hyperparameters of the BYM2 model (i.e., marginal precision (τ_s) and a mixing parameter (φ)), we used weakly informative penalized complexity (PC) priors (Simpson et al., 2017). See Simpson et al. (2017) for technical details.

Based on equation (1) above, we designed three models (models 1, 2, and 3) for the logarithm of relative risks of DD injuries ($\eta_i = \log(\theta_i)$) as follows:

Model 1: independent variables + block-group-level heterogeneity.
Model 2: Model 1 + census-tract-level heterogeneity.
Model 3: Model 2 + county-level heterogeneity.

Model 1 concerns only latent effects/unobserved heterogeneity from the arrangement of census block groups whereas models 2 and 3 further incorporate random influences from coarser levels of the hierarchy of the administrative units besides those in model 1.

4. Description of data

4.1. Study area

The study region included 1,024 census block groups within five counties in the CMA in Ohio. Although the CMA has ten counties, we focused on five counties following the 2016–2040 Metropolitan Transportation Plan (MTP)[2] by the Mid-Ohio Regional Planning Commission (MORPC) to support the goals and objectives of the plan (Mid-Ohio Regional Planning Commission (MORPC), 2020). Therefore, to adhere to this plan and potentially contribute to its implementation, we restricted our focus to DD crashes in these five counties between 2015 and 2019.

We followed the U.S. Census Bureau's standard hierarchy of census geographical entities, in which lower levels of statistical areas are completely nested within higher levels (U.S. Census Bureau, 2020). Hence, each block group is nested within a census tract, which then belongs to a CMA county. Fig. 1 illustrates the hierarchical relations among different entities in the CMA. The MORPC region, our point of interest, is highlighted in blue and is filled with census tracts and census block groups. This allows us to incorporate the unobserved multilevel cross-correlation stemming from higher levels of administrative units (i.e., census tracts and counties) and between-level spatial correlation across census block groups. When block groups share a border (boundary), we regarded them as neighbors.

4.2. Crash data

The Ohio Department of Public Safety (ODPS) publishes a database of motor vehicle crashes in the state (ODPS, 2020). We obtained five years (2015–2019) of crash data for the five CMA counties and considered and geocoded each record over the region as the raw dataset
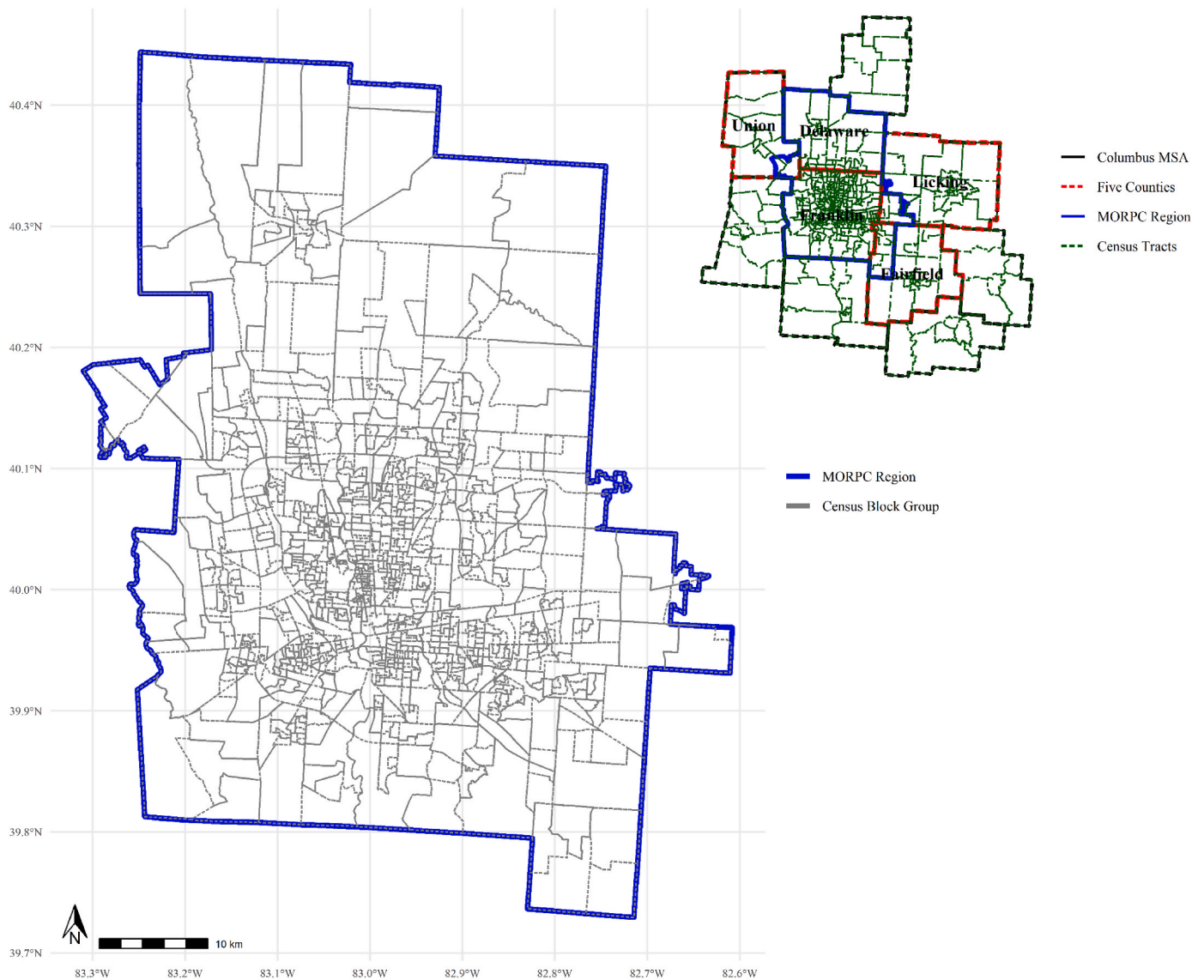

Fig. 1. Study region.

including the crash locations (latitude and longitude of each crash).[3] We assigned each record to its census block group, which is then nested within a census tract as well as a county, resulting in the multilevel (hierarchical) structure addressed in the formal analysis. One issue here is that distraction-affected crashes were identified based on law enforcement reports. People's reluctance to admit that crashes are due to DD may indicate an underreporting problem (Lym & Chen, 2021). Nevertheless, we assumed that our dataset provides an accurate reflection or at least a random sample of DD-related vehicle crashes in Central Ohio.

Our dataset consisted of 12,541 reported DD crashes for the period 2015–2019 in 1,024 block groups, which are completely nested within 338 census tracts across the 5 CMA counties. We focused on crashes leading to injury, aggregating the reported categories of evident injury, severe injury, and fatality (FA), and excluding crashes resulting in

Table 1
Systemic breakdown of distraction-affected crashes by severity and level of urbanization.

Severity	Level of urbanization			
	Urban	Suburban	Rural	Total
PDO (A)	3,499 (41.5%)	3,954 (46.9%)	972 (11.5%)	8,425 (100%)
Possible Injury (B)	782 (34.8%)	969 (48.2%)	283 (16.9%)	2,034 (100%)
Evident, Severe, and Fatal (C)	864 (41.5%)	928 (44.6%)	290 (13.9%)	2,082 (100%)
Total crashes (A + B + C)	5,145 (41.0%)	5,851 (46.7%)	1,545 (12.3%)	12,541 (100%)

Note: Numbers in parentheses are calculated based on the last column on the right (authors' calculation).
Source: Ohio Department of Public Safety (ODPS, 2020).

---

[3] Raw crash data are publicly available. The data can be requested via https://ohtrafficdata.dps.ohio.gov/crashstatistics/home (ODPS, 2020).

property damage only (PDO) or possible injury. Based on this definition, our injury measure represented the most severe category of crashes (C in Table 1).[4] Of the 12,541 reported crashes in our study area, 2,082 (16.6%) fell under this class. We also found that this proportion varied according to the degree of urbanization of crash locations (Table 1), suggesting that urbanization is an important variable. For example, about 67.18% (8,425) of the total DD crashes were classified as PDO, of which 88.4% (41.5% + 46.9%) occurred in either urban or suburban locations during that period. Regarding injuries and fatal crashes (2,082), 86.1% (41.5% + 44.6%) happened in urban and suburban areas.

From a different perspective, Fig. 2 presents the frequency distribution of injuries and fatal distraction-induced crashes (C in Table 1) according to urbanization levels. The number of crashes varied with urbanization; that is, 241 of 1,024 census block groups (23.5%) showed zero injury crashes in urban areas, while no injury crashes were observed in 76 suburban (7.42%) and 19 rural (1.9%) block groups. The average number of crashes was 2.03 with a variance of 7.49, indicating a significant overdispersion that requires a model to address such variability. In addition, 336 of 1,024 block groups (32.8%) observed neither injury nor fatal crashes, leading us to suspect a potential zero inflation.[5]

Fig. 3 illustrates the spatial distribution of DD-induced injuries and fatal crashes at the census block-group level. We considered raw crash counts (left panel) and crash rates adjusted by VMT[6] (right panel) to reveal patterns. Although some differences seemed to exist in the degree of each distribution, we could verify that census block groups with similar injuries and fatal crashes appear closely located, indicating a positive spatial autocorrelation. The measures of global spatial autocorrelation, Moran's I test statistic, were estimated as 0.230 for raw crash counts (p-value < 0.0001) and 0.207 for crash rates (p-value < 0.0001). The formal statistical assessment results clearly confirm a spatial dependency or correlation in DD injury crashes occurring across the MORPC region. Ignoring the latent influences of space may result in unreliable estimates of relative injury risks, justifying our adoption of spatial models.

4.3. Variables considered for this study

Following previous transportation safety studies, we considered explanatory variables such as the built environment (e.g., land-use mix, urbanization level, gross activity density, and roadway network), sociodemographic features (e.g., proportion of age cohorts, educational attainments, and employment status) to differentiate the latent spatial influences on DD-induced crash injuries.

To develop several sociodemographic, employment-related, and built-environment variables for each census block group in the CMA, we used the Smart Location Database (SLD)[7] provided by the

Environmental Protection Agency (EPA) (SLD, 2020). For example, total network density and intersection density reflected the existing roadway built environment. Gross activity density was defined by the sum of jobs and housing units per unprotected land in each census block group, while the employment mix (entropy)[8] of the five-tier employment categories was used to measure land-use diversity. To account for differences in the likelihood of exposure to vehicle crashes, we leveraged VMT as an exposure measure (an offset). Table 2 summarizes the statistics of the candidate variables for this study. Notably, there seem to be strong correlations among several covariates, leading us to be highly selective in terms of modeling.[9]

5. Results and discussion

5.1. Model selection

Bayesian models rely on a computationally expensive Markov chain Monte Carlo (MCMC) simulation. To circumvent the computational costs of an MCMC simulation for Bayesian spatial models, we leveraged R-INLA (Rue et al., 2017), an efficient alternative that provides approximate Bayesian results for latent Gaussian models (Rue et al., 2009; Rue & Held, 2005).

As previously discussed, we suggested three models that could account for unobserved heterogeneity from the multilevel structure of administrative units. Table 3 presents the results of the computed goodness-of-fit measures for several models considered in this study.

Regarding the performance of each model, we employed the deviance information criterion (DIC), the widely applicable information criterion (WAIC), and log pseudo marginal likelihood (LPML) for an optimal model selection. DIC, a Bayesian analog of the AIC, penalizes model complexity (e.g., models with more parameters or latent effects) by adding the effective number of parameters to the posterior mean of deviance (Spiegelhalter et al., 2002). Thus, models with a smaller DIC are preferable, and the same is true of the WAIC. Conversely, models with higher LPML values are preferred. In that regard, we found that the spatial models appeared to outperform their nonspatial counterparts because of significant drops in DIC and WAIC values in spatial models. Moreover, we have already observed excessive zeros in the crash data (Fig. 2); hence, we also employed zero-inflated Poisson (ZIP) models to address them. The performance metrics (i.e., DIC, WAIC, and LPML values) were in favor of nonspatial and spatial models, indicating that the spatial ZIP models are likely to overfit the data, and after accounting for fixed effects by independent variables, the potential zero-inflation problems were resolved.

5.2. Results

Table 3 allows us to select optimal structures for further inferences (in this case, spatial models), and Table 4 presents the detailed results, which include posterior means, standard deviations, and 95% Bayesian credible intervals for models 1, 2, and 3. We observed that models 2 and 3 outperformed model 1 and had minor differences in DIC values (the

---

[4] Although we focused on injury and fatal crashes relative to distracted driving, we also conducted a formal assessment of total DD crashes for comparison purposes, available in Appendix Table A1.

[5] The zero-inflation issue seems to emerge if we only consider the raw crash data, but this is resolved by the formal modeling process (by accounting for fixed and random effects). We also employed models that specifically address zero inflation (i.e., zero-inflated Poisson model), which results in poorer performance (i.e., zero-inflated models appear to overfit the data) compared to our optimal (suggested) model. These models are compared in Table 3.

[6] We calculated crash rates as the ratio of crash counts to vehicle miles traveled in thousands.

[7] https://www.epa.gov/sites/production/files/2014-03/documents/sld_user guide.pdf.

[8] Entropy $= -\sum_{i=1}^{5} p_i \ln p_i$ where p_i is the proportion of employment in type i in the census block group (U.S. Department of Transportation, 2021). We considered a five-tier employment classification scheme (retail, office, industrial, service, and entertainment) from the U.S. Census LEHD Origin-Destination Employment Statistics Work Area Characteristics tables (LEHD, 2021).

[9] We excluded covariates that showed a correlation greater than 0.4 in our final variable selection for advanced spatial models since the estimated effects of correlated variables appeared to swing widely and weaken the statistical power of the regression model.

[10] With regard to nonspatial models (baseline models), we employed generalized linear mixed models. It is worth noting that the difference between nonspatial and spatial models is the adoption of BYM2 priors to account for spatial correlation.
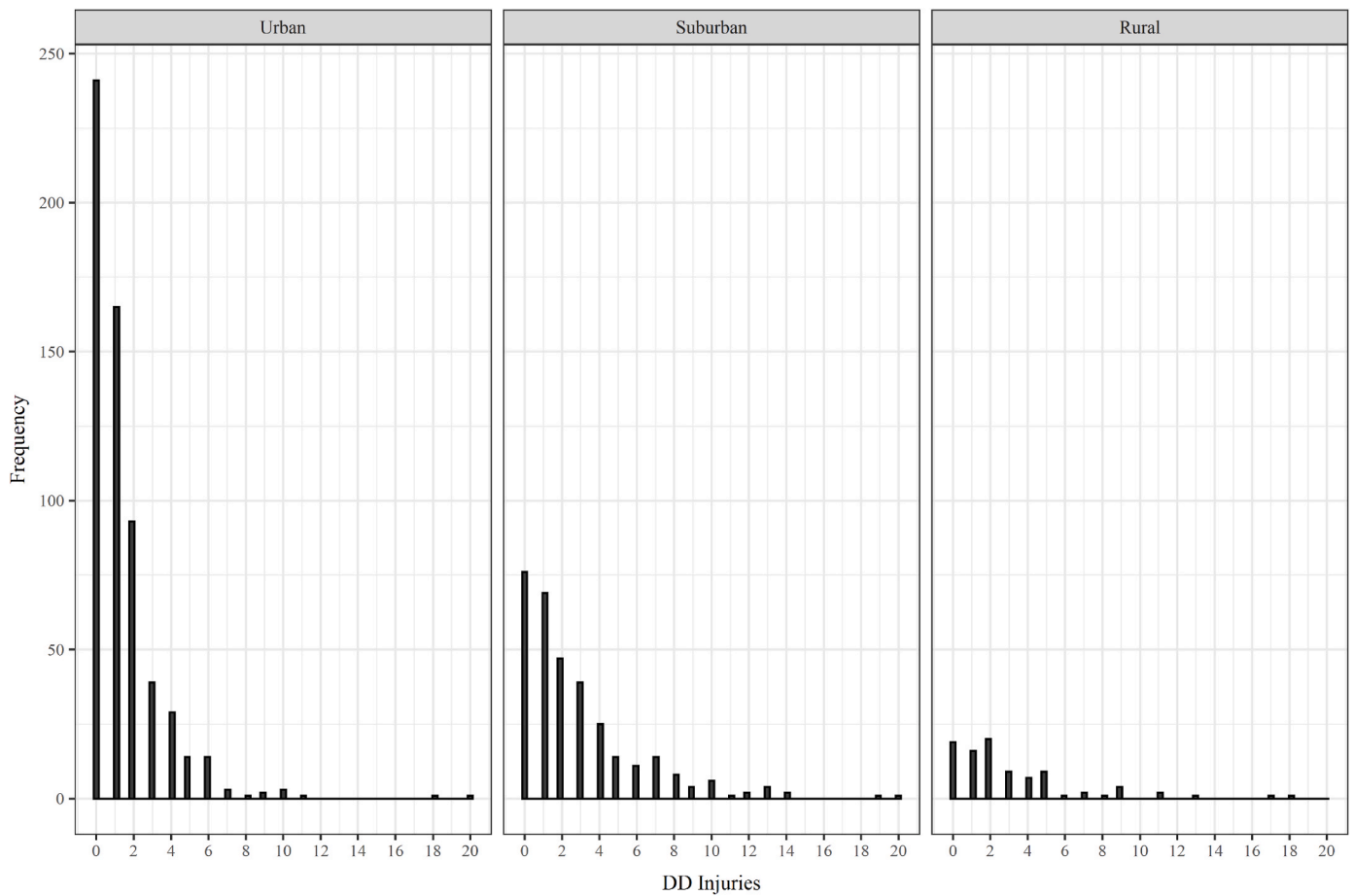
Fig. 2. Frequency distribution of injuries and fatal crashes according to urbanization.

LPML of model 3 is even higher than that of model 2). Among the spatial models, we selected model 3 as our optimal choice; thus, our discussion of outcomes will be based on model 3. When interpreting the results, one should note that the effects are linked to the *relative* risks of injury and fatal crashes because the VMT offset effectively makes the dependent variable λ/VMT.

Regarding fixed effects, we considered several contextual variables including sociodemographic factors, transportation-related variables, and built-environment features in each census block group.[11] Our results showed that the relative risks of crashes are positively associated with having a high school or college education, while population density was negatively correlated with the risks. Additionally, although the estimated parameters were not statistically meaningful, the influence of younger cohorts (aged 15–19 and 20–24) appeared to be positive, indicating an elevated risk in these age groups (consistent with the findings of other studies such as Huang et al. (2010), Aguero-Valverde (2013), and Lym and Chen (2020)). Gross activity density seemed to be negatively associated with DD-related injuries.

Regarding the effects of built environments, intersection density was positively associated with relative injury risks by DD. Due to the strong multicollinearity, however, we were unable to utilize the roadway network density variable, which may be an important contributing

factor in understanding injury risks. We also considered the level of urbanization, consisting of rural, suburban, and urban, and selected rural as the reference category for comparison with the influence of other categories. In reference to a rural block group, a census block group classified as urban is more likely to have elevated injury and fatality risks whereas based on a 95% Bayesian credible interval, we did not find any strong statistical evidence for suburban block groups. A higher proportion of commercial land use in a census block group increases risks of injury and fatal crashes. Likewise, we verified that land-use diversity was positively correlated with relative risks, suggesting possible higher risks due to more diversified land use.

Table 4 also presents a systemic breakdown of random effects, focusing on various types of unobserved heterogeneity originating from the spatial alignments of the study region. We assumed that the unobserved heterogeneity (latent influences) of level 1 (block group) is characterized by spatially structured random effects under the BYM2 specification. Those of level 2 (census tract) and level 3 (county) stem from two uncorrelated random effects. For level 1 (block group) heterogeneity, the posterior mean of the mixing parameter (φ) was estimated by 0.706 (Table 4 and Fig. 4, left panel), suggesting a significant marginal variance contribution of the spatial component. That is, the scaled spatially structured random component accounted for 84% ($\sqrt{0.706} = 0.84$) of the overall variability of residual relative risks across census block groups. This justifies our adoption of the BYM2 model instead of a nonspatial model.

Furthermore, a comparison of random effects among models 1, 2, and 3 revealed that the estimated posterior mean of lower-administrative-level precision parameters (τ_s, τ_k) increases with the introduction of upper-administrative-level heterogeneity. We observed a similar increase in the estimated posterior mean of τ_k (precision of

---

[11] Before employing the formal regression models, we performed correlation analysis. Variables showing correlations greater than 0.4 were excluded from the subsequent statistical analysis. These included total roadway network density; number of jobs; median household income; proportion of population aged 30–39, 40–49, 50–64, and 65 and above; proportion of nonwhite population; unemployment rate; and vacancy rate of each census block group.
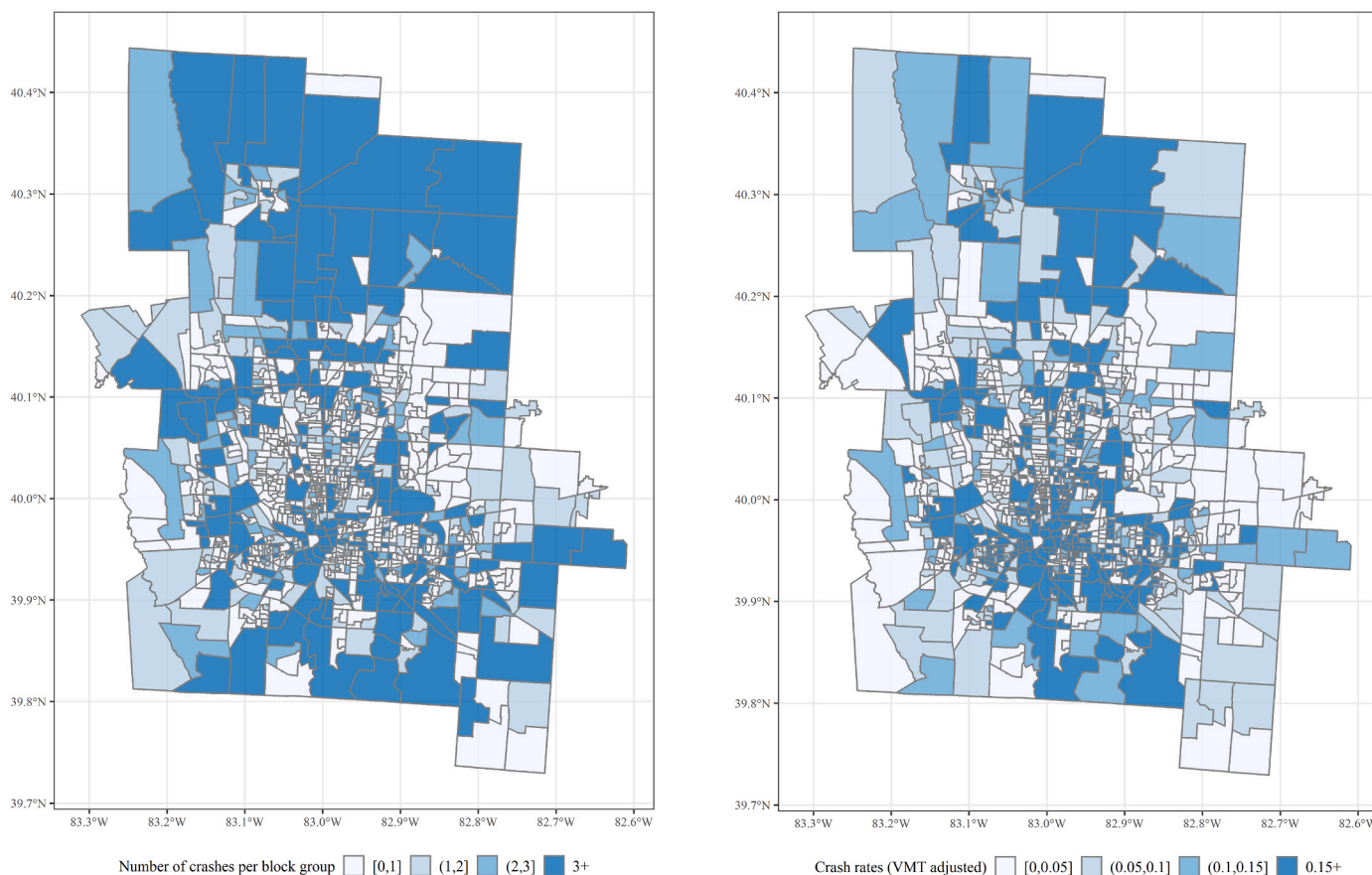
Fig. 3. Spatial distribution of injury crashes due to distracted driving.

census tracts) from 11.126 to 11.806 when the model complexity increased (i.e., by county-level random effects). One should note that precision is the inverse of variance; hence, higher precision indicates smaller variability. This supports our adoption of the multilevel structure in understanding DD-related crash injury risks.

Two precision parameters (τ_k, τ_l) addressed the multilevel unobserved heterogeneity from level 2 (census tract) and level 3 (county). The magnitude of the posterior mean of each precision parameter showed that certain amounts of random fluctuation were derived from the multilevel structure (levels 2 and 3) even after accounting for both fixed and random effects at level 1 (block group). The right panel of Fig. 4 presents the distribution of marginal standard deviation parameters[12] where we can compare each level's relative contribution to overall variability. With the same scale, we confirmed that level 1 dominated the other two. In light of this, we presented the intraclass correlation coefficient, a measure of relative contribution to the overall residual variability, for level 1. This was estimated at 80.4% (0.804) on average (posterior mean), suggesting a substantial amount of unobserved heterogeneity originating from level 1 (block group).

5.3. Places with higher relative risks of injury and fatal DD crashes

The left panel of Fig. 5 shows the posterior mean of the spatially structured random effects associated with DD in Central Ohio. The left panel suggests that after controlling for the effects of the fixed covariates, we identified the hidden risks of injury and fatal distraction-

affected crashes that may have been concealed in the raw observations. The spatially structured random effects imply that census block groups in the central city (Columbus) and more developed areas are likely to have increased relative risks, while those outside the central areas are exposed to fewer risks, revealing their dependency on neighboring locations. This indicates that DD crashes that cause injury or fatalities do not occur randomly but rather happen at specific locations.

One important advantage of our Bayesian framework was that we could examine the posterior probability of relative risk, that is, the probability of observing a certain number of crashes per VMT. To illustrate, suppose we consider two crashes per VMT as exceptionally high (i.e., twice the risk of injury and fatality, $Pr(\theta_i > 2|data)$[13]). The right-hand panel of Fig. 5 shows this high crash frequency level in our study area. Unlike Fig. 3, which presents injury and fatal DD crash distribution of raw counts and crash rates of census block groups, we adopted a model-based analytic approach to identify hot spots (places for further policy consideration) of DD crash injury risks. The darker the color, the higher the probability; these are areas where mitigation strategies may be most effective. We believe that identifying areas with excessive relative risks of distraction-affected injury and fatal crashes can support transportation safety and health practitioners' efforts to design new safety measures and enhance existing ones.

---

[12] The inverse of a precision parameter is a variance. The standard deviation is obtained when we apply the square root to the variance. We scaled each parameter to 1 so that their relative influences can be comparable.

[13] The exceedance probability of the relative risks for each census block group was based on the posterior distribution of the risks. When calculating the probability, we considered the fixed effects from covariates and multilevel unobserved heterogeneity.

Table 2

Summary of study variables (N = 1024 census block groups).

Variable	Description	Min	Max	Mean	Variance
Response Variables (Crash Counts)					
Total[1]	Total crashes	0	116	12.2	213.73
Injury[1]	Evident + severe + fatal crashes	0	20	2.03	7.49
Traffic-Related Variables					
VMT[2]	Vehicle miles traveled in thousands	0	212.75	22.64	390.54
Road network[3]	Total road network density	1.45	38.14	15.51	49.04
Intersection[3]	Auto-oriented intersection density	0	24.4	1.1	8.06
Areal Descriptors					
A. Measured in Numbers					
Population density	Population density (persons/square miles)	0	68798	5185	24505591
Jobs	Number of jobs	0	39463	804	5508830
Income	Median household income ($/year)	0	250001	65531	1372256939
Activity[4]	Gross activity density	0.028	241.35	6.35	116.45
Landmix[5]	Land-use diversity	0	0.985	0.429	0.1046
B. Measured in Percentages					
A1519	Population aged 15–19 years	0	79.80	6.05	38.77
A2024	Population aged 20–24 years	0	92.05	7.79	121.45
A2529	Population aged 25–29 years	0	47.14	8.94	58.1
A3039	Population aged 30–39 years	0	39.85	14.3	39.37
A4049	Population aged 40–49 years	0	53.33	12.91	36.19
A5064	Population aged 50–64 years	0	50.87	18.83	55.92
Nonwhite	Ethnicity other than white	0	100	21.34	643.28
High school	High school degree or less	0	100	10.47	124.06
College	College degree	0	91.3	52.05	362.9
Vehicle	Households with 1 or 2 vehicles available	0	100	73.7	186.06
Unemployment	Unemployment rate	0	52.01	7.01	56.7
Vacancy	Vacancy rate	0	88	10.21	135.21
Commercial	Commercial land use	0	84	10.41	283.13
Residential	Residential land use	0	100	71.72	776.54
Industrial	Industrial land use	0	44	1.21	23.75

Note: 1) The measurement is based on census block group (level 1), which is the unit of analysis for this study (N = 1024). There are 338 census tracts (level 2) and 5 counties (level 3) in our study region.2) As an offset (exposure), we employed vehicle miles traveled (VMT). 3) Adopted from the Smart Location Database (U.S. Environmental Protection Agency (EPA), 2020) 4) (Gross) Activity density consists of the sum of employment and housing units in each census block group. 5) An entropy measure for the five-tier employment categories was applied to calculate land-use diversity (U.S. Department of Transportation, 2021).

6. Concluding remarks

By examining the spatial alignments of administrative units in the Columbus Metropolitan Statistical Area, we explored the role of unobserved heterogeneity in DD-induced injury risks in addition to the influence of fixed effects by a series of covariates. Adopting a census block group as a unit of analysis, we were able to incorporate the spatial alignment of administrative units, which allowed us to identify the multilevel latent influences on the relative risks of DD-related injury and fatal crashes. Our analysis confirmed spatially structured latent effects shared by neighboring block groups as well as random influences from upper-level hierarchical units (census tracts and counties). The selected model further showed the unobserved heterogeneity whose variability or fluctuation can be largely contributed by block-group and tract-level random effects. We also found that the relative contribution of spatially structured random effects (intraclass correlation) in the block-group level accounted for 80.4% of the total residual variability (after controlling for fixed effects by covariates), justifying our use of a multilevel analytical framework.

We observed that a higher population density is associated with lower injury and fatal crashes caused by DD. The negative effect of population density on the risks may be explained by the stronger safety standards/regulations imposed by policymakers or safety officials to prevent DD-related crashes in a densely populated area. Meanwhile, we also found that the proportion of commercial land use, land-use diversity, and intersection density were positively associated with crash occurrence, contributing to an elevated risk of injury and fatality. Notably, the positive influence of land-use diversity on DD-induced injuries differs from the findings of Fatmi and Habib (2019). Fatmi and Habib (2019) investigated the severity of DD-related crashes on individuals and the effects of surrounding environments such as land-use mix and sidewalk length whereas our study examined relative risks (i.e., frequency of injuries) using aggregated data (i.e., census block group as a unit of analysis). The structure of the study design and driving culture between two regions (i.e., Nova Scotia in Canada and CMA in Ohio, USA) might have influenced the outcomes. Our study further suggests that a census block group classified as urban is likely to be credibly associated with relative risks compared with those categorized as rural.

These findings suggest several practical implications. As previously noted, our research focus area was grounded in the 2016–2040 MTP by the MORPC (Mid-Ohio Regional Planning Commission (MORPC), 2020). After the MTP, the MORPC developed the Central Ohio Transportation Safety Plan (Mid-Ohio Regional Planning Commission (MORPC), 2021) for the region, where DD is a primary driving safety concern. The safety plan, however, does not consider the multifaceted feature of DD in Central Ohio, which provided us the motivation to conduct this research. We argue that this paper can help support and improve transportation safety in Central Ohio.

For instance, although the evidence is not strong, younger age cohorts (e.g., 15–19, 20–24) are prone to DD, which might cause injury or fatal crashes. Thus, reinforcing or enhancing safety measures for these groups should be considered (e.g., mandatory driver's education designed to reduce DD behavior or the promotion of cellphone or texting bans for novice drivers). As previously noted, block groups with higher intersection density may result in more DD-induced injuries or fatal crashes. We also identified hot spots with twice the relative risks compared to the overall average across census block groups. These aspects may together suggest regions of policy intervention (e.g., higher intersection density as well as excessive relative risks) and the creation of safety management plans to prioritize scarce resource allocation.

Several limitations were not fully addressed in this study. Although our research confirmed the existence of multilevel unobserved heterogeneity from the spatial arrangement of administrative units and spatial correlation among adjacent census block groups, we did not consider covariates specific to the higher level of hierarchies that might be confounded with lower-level variables. We also assumed that the

Table 3
Comparison of goodness-of-fit measures of various models.

	Nonspatial[10]			Spatial			ZIP spatial[2]		
	Model 1[1]	Model 2[1]	Model 3[1]	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
DIC	4039.1	3353.2	3353	3320.9	3313.8	3314.9	5214.9	5209.4	5206.3
WAIC	4089.8	3343.5	3342.1	3297	3290.9	3295.5	5870	5860	5848.3
LPML	−2287.7	−1963.9	−1964.5	−2020.5	−1948	−1929.4	−5682.1	−5689.3	−5643.2

Note: 1) Model 1 only considers level 1 (1024 census block groups) random influences, while models 2 and 3 account for level 1 and 2 (census tracts) and levels 1, 2, and 3 (counties) multilevel heterogeneity, respectively. Hence, model 3 incorporates random effects from both 338 census tracts and 5 counties. 2) We employ a zero-inflated Poisson (ZIP) model to deal with potential zero-inflation issues.

Table 4
Comparison of the spatial models (posterior means and 95% credible intervals).

	Dependent variable: injury and fatal crash counts (Y_i)					
	Model 1	Model 2			Model 3	
	Mean (S.D.)	95% C.I	Mean (S.D.)	95% C.I	Mean (S.D.)	95% C.I
Fixed effects						
Intercept	−8.074 (0.366)	(−8.792, −7.356)	−7.986 (0.376)	(−8.721, −7.247)	−8.032 (0.377)	(−8.771, −7.289)
Sociodemographic						
Log (Population density)[1]	−0.35 (0.062)	(−0.472, −0.229)	−0.371 (0.063)	(−0.494, −0.248)	−0.37 (0.062)	(−0.493, −0.248)
Ages 15–19	0.013 (0.007)	(−0.002, 0.026)	0.012 (0.007)	(−0.002, 0.026)	0.012 (0.007)	(−0.002, 0.026)
Ages 20–24	0.002 (0.005)	(−0.008, 0.012)	0.003 (0.005)	(−0.007, 0.013)	0.003 (0.005)	(−0.007, 0.013)
Ages 25–29	−0.001 (0.006)	(−0.012, 0.011)	−0.002 (0.006)	(−0.014, 0.009)	−0.002 (0.006)	(−0.014, 0.009)
High school degree	0.018 (0.004)	(0.01, 0.027)	0.018 (0.004)	(0.009, 0.026)	0.018 (0.004)	(0.009, 0.026)
College degree	0.009 (0.003)	(0.004, 0.014)	0.009 (0.003)	(0.004, 0.014)	0.009 (0.003)	(0.004, 0.014)
Vehicle ownership	0.001 (0.003)	(−0.005, 0.007)	0.001 (0.003)	(−0.005, 0.006)	0.001 (0.003)	(−0.005, 0.006)
Gross activity density	−0.001 (0.003)	(−0.008, 0.002)	−0.001 (0.003)	(−0.007, 0.005)	−0.001 (0.003)	(−0.007, 0.005)
Land use						
Commercial land use	0.013 (0.003)	(0.008, 0.018)	0.014 (0.003)	(0.009, 0.019)	0.014 (0.003)	(0.009, 0.019)
Residential land use	−0.003 (0.002)	(−0.008, 0.002)	−0.002 (0.002)	(−0.007, 0.003)	−0.002 (0.002)	(−0.007, 0.003)
Industrial land use	0 (0.007)	(−0.013, 0.013)	0.001 (0.007)	(−0.013, 0.014)	0.001 (0.007)	(−0.013, 0.014)
Land−use diversity	0.315 (0.129)	(0.061, 0.569)	0.35 (0.129)	(0.097, 0.604)	0.345 (0.128)	(0.094, 0.598)
Roadway network						
Intersection density	0.066 (0.011)	(0.044, 0.088)	0.062 (0.012)	(0.039, 0.085)	0.062 (0.011)	(0.04, 0.085)
Level of urbanization[2]						
Urban	0.49 (0.195)	(0.107, 0.872)	0.56 (0.208)	(0.152, 0.968)	0.567 (0.21)	(0.154, 0.98)
Suburban	0.301 (0.164)	(−0.022, 0.623)	0.333 (0.175)	(−0.011, 0.678)	0.332 (0.177)	(−0.015, 0.679)
Random effects (Hyperparameters)						
φ (Mixing parameter)	0.666 (0.091)	(0.477, 0.831)	0.707 (0.099)	(0.492, 0.872)	0.706 (0.105)	(0.495, 0.894)
τ_s (Precision of level 1)[3]	1.600 (0.181)	(1.273, 1.982)	1.951 (0.307)	(1.449, 2.649)	2.021 (0.335)	(1.384, 2.688)
τ_k (Precision of level 2)[3]			11.126 (4.715)	(4.932, 23.021)	11.806 (5.918)	(4.951, 27.26)
τ_l (Precision of level 3)[3]					1760 (30792)	(6.168, 10781)
Intraclass correlation						
$\sigma_s^2 / (\sigma_s^2 + \sigma_k^2)$ [4]			0.834 (0.065)	(0.684, 0.936)		
$\sigma_s^2 / (\sigma_s^2 + \sigma_k^2 + \sigma_l^2)$ [4]					0.804 (0.088)	(0.610, 0.941)
Goodness of fit						
DIC	3320.85		3313.78		3314.94	
WAIC	3296.95		3290.92		3295.46	
Marginal log likelihood	−1287.22		−1285.40		−1285.44	

Note: 1) To stabilize the variance and skewness of the population density distribution, we applied logarithmic transformation. 2) Reference category: rural. 3) Precision is the inverse of variance (i.e., $\tau^{-1} = \sigma^2$); level 1 (block group), level 2 (census tract), level 3 (county). 4) Intraclass correlation: a relative contribution of spatially correlated random effects (level 1) on overall variability. 5) Exposure variable: vehicle miles traveled (VMT) used as an offset of the differences in traffic volume.

relation between the outcomes of DD crashes and some predictors was linear, ignoring the possibility of quadratic or nonparametric terms. In addition, our study did not consider other types of crashes besides those induced by DD and focused only on DD-related injuries and fatal crashes, excluding possible injuries and PDO crashes, which may reveal different outcomes. As suggested by Rahman Shaon et al. (2019), different injury severity levels are likely to be correlated, which would require a multivariate model to account for cross-severity correlations. Our study did not explore these aspects. Furthermore, various types of crashes other than DD may share similar features of spatial influences on DD, which can help us develop a comprehensive transportation safety plan. This research used a five-year aggregate crash data to focus on the spatial pattern of the relative risks; consequently, we ignored the temporal dimension of crash occurrences and its latent influences.

Nevertheless, by providing different insights, we believe that our efforts can help improve the understanding of DD-induced crashes.
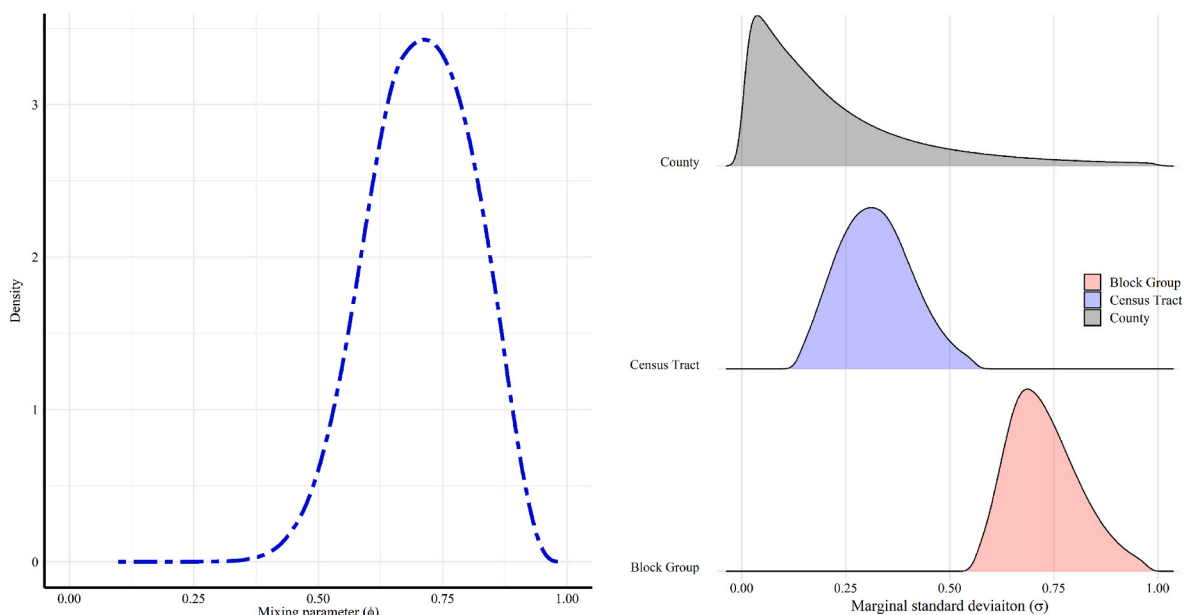
Ethics statement

Not applicable.

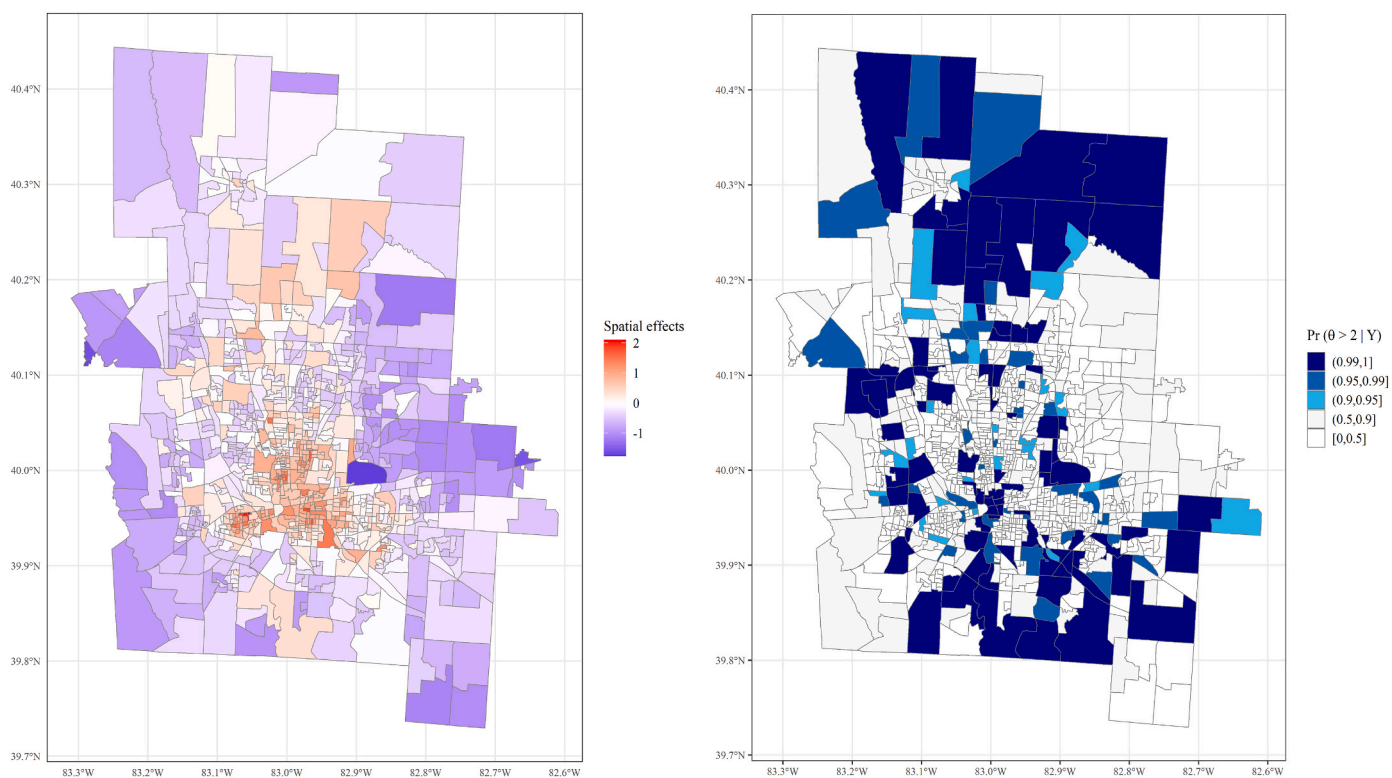**Fig. 4.** Posterior distribution of hyperparameters.

Fig. 5. Places with further policy intervention.
Note: Spatial effects (residual) by BYM2 (left); exceedance probability of twice the (overall) relative risks (right).

(KRIHS).

Credit author statement

Youngbin Lym: Conceptualization, Methodology, Software, Data curation, Formal analysis, Writing—Original draft, Visualization, Investigation, Writing—Review & editing, and Validation. Ki-Jung Kim: Methodology, Formal analysis, Investigation, Writing—Review & editing Seunghoon Kim: Conceptualization, Data curation, Investigation,

Validation, Writing—Review & editing, and Supervision.

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.

Appendix

Table A1
Comparison of outcomes of the optimal model (injury vs. total crashes)

	Injury and fatal crashes		Total crashes	
	Mean (S.D.) [5]	95% C.I. [5]	Mean (S.D.)	95% C.I.
Fixed effects				
Intercept	−8.032 (0.377)	(−8.771, −7.289)	−6.07 (0.292)	(−6.643, −5.496)
Sociodemographic				
Log (Population density) [1]	−0.37 (0.062)	(−0.493, −0.248)	−0.366 (0.045)	(−0.454, −0.277)
Ages 15–19	0.012 (0.007)	(−0.002, 0.026)	0.013 (0.005)	(0.003, 0.022)
Ages 20–24	0.003 (0.005)	(−0.007, 0.013)	0.003 (0.003)	(−0.004, 0.01)
Ages 25–29	−0.002 (0.006)	(−0.014, 0.009)	−0.005 (0.004)	(−0.013, 0.003)
High school degree	0.018 (0.004)	(0.009, 0.026)	0.012 (0.003)	(0.006, 0.018)
College degree	0.009 (0.003)	(0.004, 0.014)	0.005 (0.002)	(0.001, 0.009)
Vehicle ownership	0.001 (0.003)	(−0.005, 0.006)	0.001 (0.002)	(−0.003, 0.005)
Gross activity density	−0.001 (0.003)	(−0.007, 0.005)	−0.004 (0.002)	(−0.009, 0)
Land use				
Commercial land use	0.014 (0.003)	(0.009, 0.019)	0.012 (0.002)	(0.008, 0.016)
Residential land use	−0.002 (0.002)	(−0.007, 0.003)	−0.003 (0.002)	(−0.006, 0.001)
Industrial land use	0.001 (0.007)	(−0.013, 0.014)	−0.001 (0.005)	(−0.011, 0.008)
Land-use diversity	0.345 (0.128)	(0.094, 0.598)	0.523 (0.086)	(0.353, 0.693)
Roadway network				
Intersection density	0.062 (0.011)	(0.04, 0.085)	0.068 (0.009)	(0.051, 0.085)
Level of urbanization [2]				
Urban	0.567 (0.21)	(0.154, 0.98)	0.727 (0.161)	(0.413, 1.044)
Suburban	0.332 (0.177)	(−0.015, 0.679)	0.506 (0.139)	(0.235, 0.779)
Random effects (Hyperparameters)				
φ (Mixing parameter)	0.706 (0.105)	(0.495, 0.894)	0.725 (0.075)	(0.558, 0.850)
τ_s (Precision of level 1) [3]	2.021 (0.335)	(1.384, 2.688)	1.868 (0.173)	(1.549, 2.230)
τ_j (Precision of level 2) [3]	11.806 (5.918)	(4.951, 27.26)	21.888 (10.285)	(9.75, 48.66)
τ_k (Precision of level 3) [3]	1760 (30792)	(6.168, 10781)	1016.294 (7977.417)	(9.452, 6837.27)
Intraclass correlation				
$\sigma_s^2 / (\sigma_s^2 + \sigma_j^2 + \sigma_k^2)$ [4]	0.804 (0.088)	(0.610, 0.941)	0.887 (0.053)	(0.759, 0.959)
Goodness of fit				
DIC	3314.94		5639.85	
WAIC	3295.46		5510.03	
Marginal log likelihood	−1285.44		−2760.97	

Note: 1) To stabilize the variance and skewness of the population density distribution, we applied logarithmic transformation. In addition, when transforming the population density variable, we added 1 (a constant) to circumvent the issue of logarithmic transformation of zero values. 2) Reference category: rural. 3) The inverse of variance is precision (i.e., $\tau^{-1} = \sigma^2$). Level 2 (census tract); level 3 (county). 4) Intraclass correlation: a relative contribution of spatially structured random effects on overall variability. 5) Exposure variable: vehicle miles traveled (VMT); S.D.: standard deviation, C.I.: credible interval.

Table A2
Comparison of goodness-of-fit measures of various models

	Nonspatial			Spatial (BYM2 + PC prior)								
	M1	M2	M3	M1	M2	M2-1	M2-2	M3	M3-1	M3-2	M3-3	M3-4
DIC	4039.1	3353.2	3353	3320.9	3313.8	3313.8	3314.1	3314.9	3314.3	3313.2	3313.7	3313.6
WAIC	4089.8	3343.5	3342.1	3297	3290.9	3297.1	3295.4	3295.5	3299	3293.4	3294.7	3295.2
LPML	−2287.7	−1963.9	−1964.5	−2020.5	−1948	−1877.9	−1907.7	−1929.4	−1870.8	−1915.2	−1911.3	−1898.2

Note: 1) M1 (model 1) only considers level 1 (census block groups) random influences, while M2 and M3 account for levels 1 and 2 (census tracts) and levels 1, 2, and 3 (counties) cross-group heterogeneity, respectively. Hence, M3 incorporates random effects from both 338 census tracts and 5 counties. 2) In the case of spatial models, M1: BYM2 (level 1); M2: BYM2 (level 1) + IID (level 2); M2-1: BYM2 (level 1) + ICAR (level 2); M2-2: BYM2 (level 1) + BYM2 (level 2); M3: BYM2 (level 1) + IID (level 2) + IID (level 3); M3-1: BYM2 (level 1) + ICAR (level 2) + IID (level 3); M3-2: BYM2 (level 1) + BYM2 (level 2) + IID (level 3); M3-3: BYM2 (level 1) + BYM2 (level 2) + ICAR (level 3); M3-4: BYM2 (level 1) + BYM2 (level 2) + BYM2 (level 3).

Table A2 presents the results of the performance measures of several spatial models investigated in this study. M1 (model 1) considers the block-group-level latent influences characterized by the BYM2 specification, while M2 (model 2) and M3 (model 3) incorporate latent effects from census tracts and tracts + counties, respectively. The tested spatial models include.

- M1: BYM2 (block group only)
- M2: BYM2 (block group) + IID (tract)
- M2-1: BYM2 (block group) + ICAR (tract)
- M2-2: BYM2 (block group) + BYM2 (tract)
- M3: BYM2 (block group) + IID (tract) + IID (county)
- M3-1: BYM2 (block group) + ICAR (tract) + IID (county)
- M3-2: BYM2 (block group) + BYM2 (tract) + IID (county)
- M3-3: BYM2 (block group) + BYM2 (tract) + ICAR (county)
- M3-4: BYM2 (block group) + BYM2 (tract) + BYM2 (county)

Model complexity increases as we incorporate latent influences from upper-level administrative units. We observed that changes in DIC and WAIC values from M1 to M2 of the spatial models appeared significant. After that, however, we observed extremely small (negligible) differences in DIC and WAIC values among sophisticated spatial models (M2, M2-1, … M3,.. M3-4). This indicates that addressing spatially structured random effects at the tract level or county level does not improve the performance of the models. Hence, we selected simpler models (M2 and M3) as our optimal choices since they perform as efficiently as the more sophisticated ones.

References

AAA. (2018). Foundation for traffic safety, 2017 *Traffic Safety Culture Index*. Retrieved from https://aaafoundation.org/wp-content/uploads/2018/03/TSCI-2017-Report. pdf. (Accessed 16 January 2021).

Aguero-Valverde, J. (2013). Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. *Accident Analysis & Prevention, 59*, 365–373.

Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention, 38*(3), 618–625.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Netherlands: Springer.

Arnold, L. S., Benson, A. J., Tefft, B. C., Barragan, D., Jin, L., Kolek, S., & Horrey, W. J. (2019). *Effectiveness of distracted driving countermeasures: A review of the literature (research brief)* (pp. 1–13). Washington, D.C.: AAA Foundation for Traffic Safety.

Barua, S., El-Basyouny, K., & Islam, M. T. (2014). A full Bayesian multivariate count data model of collision severity with spatial correlation. *Analytic Methods in Accident Research, 3*(4), 28–43. https://doi.org/10.1016/j.amar.2014.09.001

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics, 43*(1), 1–20. https://doi.org/10.1007/bf00116466

Birrell, S. A., & Young, M. S. (2011). The impact of smart driving aids on driving performance and driver distraction. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(6), 484–493. https://doi.org/10.1016/j.trf.2011.08.004

Blangiardo, M., & Cameletti, M. (2015). *Spatial and spatio-temporal bayesian models with R-INLA*. Wiley.

Boulieri, A., Liverani, S., de Hoogh, K., & Blangiardo, M. (2017). A space-time multivariate Bayesian model to analyse road traffic accidents by severity. *Journal of the Royal Statistical Society: Series A, 180*(1), 119–139. https://doi.org/10.1111/rssa.12178

Braitman, K. A., & Braitman, A. L. (2017). Patterns of distracted driving behaviors among young adult drivers: Exploring relationships with personality variables. *Transportation Research Part F: Traffic Psychology and Behaviour, 46*, 169–176. https://doi.org/10.1016/j.trf.2017.01.015

Center for Disease Control and Prevention. (2020). Distracted driving. Retrieved https://www.cdc.gov/transportationsafety/distracted_driving/index.html. (Accessed 20 September 2020).

Chen, Z., & Lym, Y. (2021). The influence of built environment on distracted driving related crashes in Ohio. *Transport Policy, 101*, 34–45. https://doi.org/10.1016/j.tranpol.2020.11.011

Cressie, N. (2015). In Rev (Ed.), *Statistics for spatial data*. Wiley.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences of the United States of America, 113*(10), 2636–2641. https://doi.org/10.1073/pnas.1513271113

Dupont, E., Papadimitriou, E., Martensen, H., & Yannis, G. (2013). Multilevel analysis in road safety research. *Accident Analysis & Prevention, 60*, 402–411. https://doi.org/10.1016/j.aap.2013.04.035

Fatmi, M. R., & Habib, M. A. (2019). Modeling vehicle collision injury severity involving distracted driving: Assessing the effects of land use and built environment. *Transportation Research Record: Journal of the Transportation Research Board, 263*(7), 181–191. https://doi.org/10.1177/0361198119849060

Flask, T., & Schneider, W. (2013). A Bayesian analysis of multi-level spatial correlation in single vehicle motorcycle crashes in Ohio. *Safety Science, 53*, 1–10. https://doi.org/10.1016/j.ssci.2012.08.008

Gershon, P., Sita, K. R., Zhu, C., Ehsani, J. P., Klauer, S. G., Dingus, T. A., & Simons-Morton, B. G. (2019). Distracted driving, visual inattention, and crash risk among teenage drivers. *American Journal of Preventive Medicine, 56*(4), 494–500. https://doi.org/10.1016/j.amepre.2018.11.024

Gershon, P., Zhu, C., Klauer, S. G., Dingus, T., & Simons-Morton, B. (2017). Teens' distracted driving behavior: Prevalence and predictors. *Journal of Safety Research, 63*, 157–161. https://doi.org/10.1016/j.jsr.2017.10.002

Guo, F., Klauer, S. G., Fang, Y., Hankey, J. M., Antin, J. F., Perez, M. A., Lee, S. E., & Dingus, T. A. (2017). The effects of age on crash risk associated with driver distraction. *International Journal of Epidemiology, 46*(1), 258–265. https://doi.org/10.1093/ije/dyw234

Haque, M. M., Oviedo-Trespalacios, O., Debnath, A. K., & Washington, S. (2016). Gap acceptance behavior of mobile phone–distracted drivers at roundabouts. *Transportation Research Record: Journal of the Transportation Research Board, 2602*(1), 43–51. https://doi.org/10.3141/2602-06

Huang, H., & Abdel-Aty, M. (2010). Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis & Prevention, 42*(6), 1556–1565. https://doi.org/10.1016/j.aap.2010.03.013

Huang, H., Abdel-Aty, M. A., & Darwiche, A. L. (2010). County-level crash risk analysis in Florida: Bayesian spatial modeling. *Transportation Research Record: Journal of the Transportation Research Board, 2148*(1), 27–37. https://doi.org/10.3141/2148-04

Kidd, D. G., & Chaudhary, N. K. (2019). Changes in the sources of distracted driving among northern Virginia drivers in 2014 and 2018: A comparison of results from two roadside observation surveys. *Journal of Safety Research, 68*, 131–138. https://doi.org/10.1016/j.jsr.2018.12.004

Kircher, K., & Ahlstrom, C. (2012). The impact of tunnel design and lighting on the performance of attentive and visually distracted drivers. *Accident Analysis & Prevention, 47*, 153–161. https://doi.org/10.1016/j.aap.2012.01.019

Kountouriotis, G. K., & Merat, N. (2016). Leading to distraction: Driver distraction, lead car, and road environment. *Accident Analysis & Prevention, 89*, 22–30. https://doi.org/10.1016/j.aap.2015.12.027

Lawson, A. (2018). *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology*. Boca Raton, FL: CRC Press.

Li, X., Oviedo-Trespalacios, O., Rakotonirainy, A., & Yan, X. (2019). Collision risk management of cognitively distracted drivers in a car-following situation. *Transportation Research Part F: Traffic Psychology and Behaviour, 60*, 288–298. https://doi.org/10.1016/j.trf.2018.10.011

Liu, C., & Sharma, A. (2018). Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. *Analytic Methods in Accident Research, 17*, 14–31. https://doi.org/10.1016/j.amar.2018.02.001

Longitudinal employer-household dynamics (LEHD). (2021). US census Bureau. *Center for Economic Studies*. Retrieved on https://lehd.ces.census.gov/data/. (Accessed 26 April 2021).

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.

Lym, Y., & Chen, Z. (2020). Does space influence on the frequency and severity of the distraction-affected vehicle crashes? An empirical evidence from the central Ohio. *Accident Analysis & Prevention, 144*, Article 105606. https://doi.org/10.1016/j.aap.2020.105606

Lym, Y., & Chen, Z. (2021). Influence of built environment on the severity of vehicle crashes caused by distracted driving: A multi-state comparison. *Accident Analysis & Prevention, 150*, Article 105920. https://doi.org/10.1016/j.aap.2020.105920

Ma, X., Chen, S., & Chen, F. (2017). Multivariate space-time modeling of crash frequencies by injury severity levels. *Analytic Methods in Accident Research, 15*, 29–40.

MacNab, Y. C. (2004). Bayesian spatial and ecological models for small-area accident and injury analysis. *Accident Analysis & Prevention, 36*(6), 1019–1028.

Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research, 11*, 1–16.

Mid-Ohio Regional Planning Commission (MORPC). (2020). The 2016-2040 metropolitan transportation plan (MTP). Retrieved from https://www.morpc.org/mtp2040/. (Accessed 25 December 2020).

Mid-Ohio Regional Planning Commission (MORPC). (2021). The central Ohio transportation safety plan, 2019, Retrieved from https://www.morpc.org/program-service/transportation-safety/. (Accessed 20 May 2021).

Moraga, P. (2019). Geospatial health data: Modeling and visualization with R-INLA and shiny. In *Chapman & Hall/CRC biostatistics series)* (1st ed.) Chapman & Hall/CRC.

National Highway Traffic Safety Administration (NHTSA). (2020). *Distracted driving*. Retrieved on https://www.nhtsa.gov/risky-driving/distracted-driving. (Accessed 1 August 2020).

Ortiz, C., Ortiz-Peregrina, S., Castro, J. J., Casares-López, M., & Salas, C. (2018). Driver distraction by smartphone use (WhatsApp) in different age groups. *Accident Analysis & Prevention, 117*, 239–249. https://doi.org/10.1016/j.aap.2018.04.018

Oviedo-Trespalacios, O., Afghari, A. P., & Haque, M. M. (2020). A hierarchical Bayesian multivariate ordered model of distracted drivers' decision to initiate risk-compensating behaviour. *Analytic Methods in Accident Research, 26*, Article 100121. https://doi.org/10.1016/j.amar.2020.100121

Oviedo-Trespalacios, O., Haque, M. M., King, M., & Washington, S. (2016). Understanding the impacts of mobile phone distraction on driving performance: A systematic review. *Transportation Research Part C: Emerging Technologies, 72*, 360–380.

Oviedo-Trespalacios, O., Haque, M. M., King, M., & Washington, S. (2017). Effects of road infrastructure and traffic complexity in speed adaptation behaviour of distracted drivers. *Accident Analysis & Prevention, 101*, 67–77. https://doi.org/10.1016/j.aap.2017.01.018

Oviedo-Trespalacios, O., Haque, M. M., King, M., & Washington, S. (2019). Mate! I'm running 10 min late": An investigation into the self-regulation of mobile phone tasks while driving. *Accident Analysis & Prevention, 122*, 134–142. https://doi.org/10.1016/j.aap.2018.09.020

Papantoniou, P., Papadimitriou, E., & Yannis, G. (2017). Review of driving performance parameters critical for distracted driving research. *Transportation Research Procedia, 25*, 1796–1805. https://doi.org/10.1016/j.trpro.2017.05.148

Park, H.-C., Kim, D.-K., Kho, S.-Y., & Park, P. Y. (2017). Cross-classified multilevel models for severity of commercial motor vehicle crashes considering heterogeneity among companies and regions. *Accident Analysis & Prevention, 106*, 305–314. https://doi.org/10.1016/j.aap.2017.06.009

Parnell, K. J., Rand, J., & Plant, K. L. (2020). A diary study of distracted driving behaviours. *Transportation Research Part F: Traffic Psychology and Behaviour, 74*, 1–14. https://doi.org/10.1016/j.trf.2020.08.003

Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accident Analysis & Prevention, 40* (4), 1486–1497.

Rahman Shaon, M. R., Qin, X., Afghari, A. P., Washington, S., & Haque, M. M. (2019). Incorporating behavioral variables into crash count prediction by severity: A multivariate multiple risk source approach. *Accident Analysis & Prevention, 129*, 277–288. https://doi.org/10.1016/j.aap.2019.05.010

Riebler, A., Sørbye, S. H., Simpson, D., & Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research, 25*(4), 1145–1165. https://doi.org/10.1177/0962280216660421

Rue, H., & Held, L. (2005). Gaussian Markov random fields: Theory and applications. In *Chapman & Hall/CRC monographs on statistics and applied probability)* (1st ed.) Chapman & Hall/CRC.

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, 71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with inla: A review. *Annual Review of Statistics and Its Application, 4*(1), 395–421, 2017.

Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis & Prevention, 43*(5), 1666–1676. https://doi.org/10.1016/j.aap.2011.03.025

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science, 32*(1), 1–28. https://doi.org/10.1214/16-sts576

Smart location database. *US EPA*, (2020). Retrieved from https://www.epa.gov/smartgrowth/smart-location-mapping#SLD. (Accessed 1 November 2020).

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B, 64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353

United States census Bureau. (2020). *Hierarchy diagrams*. Retrieved on https://www.census.gov/programs-surveys/geography/guidance/hierarchy.html. (Accessed 19 December 2020).

Zeng, Q., Gu, W., Zhang, X., Wen, H., Lee, J., & Hao, W. (2019). Analyzing freeway crash severity using a Bayesian spatial generalized ordered logit model with conditional autoregressive priors. *Accident Analysis & Prevention, 127*, 87–95. https://doi.org/10.1016/j.aap.2019.02.029

Zeng, Q., Hao, W., Lee, J., & Chen, F. (2020). Investigating the impacts of real-time weather conditions on freeway crash severity: A bayesian spatial analysis. *International Journal of Environmental Research and Public Health, 17*(8), 2768. https://doi.org/10.3390/ijerph17082768