

A Unified Framework for Modeling Traffic Crashes from Hierarchical Spatial Resolutions

Shahrrior Pervaz*

Post Doctoral Scholar

Department of Civil, Environmental and Construction Engineering

University of Central Florida

Tel: 1-407-561-0298; Fax: 1-407-823-3315

Email: shahrrior.pervaz@ucf.edu

ORCID number: 0000-0001-7966-7083

Tanmoy Bhowmik

Assistant Professor

Department of Civil and Environmental Engineering

Portland State University

Tel: 1-407-927-6574; Fax: 1-407-823-3315

Email: tbhowmik@pdx.edu

ORCID number: 0000-0002-0258-1692

Naveen Eluru

Professor

Department of Civil, Environmental and Construction Engineering

University of Central Florida

Tel: 407-823-4815, Fax: 407-823-3315

Email: naveen.eluru@ucf.edu

ORCID number: 0000-0003-1221-4113

*Corresponding author

ABSTRACT

Independent traffic crash modeling approaches do not account for the embedded relationships related to the multi-resolution data structure, leading to mis-specified estimations. The recently developed integrated frameworks demonstrate the capability of addressing this drawback. The current study proposes an integrated framework that accommodates information from multiple spatial units and observation resolutions. Specifically, the study develops an integrated model system that allows for the influence of independent variables from disaggregate crash record, micro-facility (segment and intersection) and macro (traffic analysis zone) level simultaneously within the macro level propensity estimation. The empirical analysis considers disaggregate crash records of 1,818 segments and 4,184 intersections from 300 traffic analysis zones in the City of Orlando, Florida. These crash records contain crash-specific factors, driver and vehicle factors, roadway, road environmental and weather information of each crash record. For micro-facility and macro levels, an exhaustive set of independent variables including roadway and traffic factors, land-use and built environment attributes, and sociodemographic characteristics are considered. The proposed model system can also accommodate for hierarchical correlations among the data across observation resolutions and parameter variability across the system. The empirical analysis is augmented by employing several goodness of fit and predictive measures. The results clearly demonstrate the improved performance offered by the proposed integrated model system relative to the non-integrated model. A validation exercise also highlights the superiority of the proposed framework. The application of the proposed integrated framework can allow transportation professionals to adopt policy-based, site-specific, and outcome-specific solutions simultaneously.

Keywords: Integrated framework, Multi-level data, Crash frequency, Crash severity, Unobserved effects.

1. BACKGROUND

Transportation safety professionals apply statistical and econometric models to analyze road traffic crashes, identify the causes of crash occurrence and the consequences, forecast future situations, and devise appropriate crash countermeasures to alleviate the situations. These models are applied at different spatial units and for different observation resolutions. For instance, researchers apply macro level (such as zonal or county level) models to formulate long-term policy and planning level strategic solutions while micro-facility level (such as segment facility or intersection facility level) models are employed to design facility-specific engineering solutions (Cai et al., 2019; Pervaz et al., 2022). Both the macro level and micro level models use aggregated crash data and examine crash occurrence via crash frequency models. In these crash frequency models, a host of variables including roadway and traffic factors, land-use and built environment attributes, and sociodemographic characteristics are aggregated at different spatial resolutions such as zones, corridors, segments and intersections (Cai et al., 2019; Ding et al., 2023; Pervaz et al., 2022)¹. Alternatively, disaggregate level (such as crash record or driver record) models are employed to obtain outcome-specific solutions using individual record level details including crash specific, driver, vehicle, roadway, weather and road environmental variables (Abdel-Aty, 2003).

Transportation safety literature has traditionally developed independent model systems for each of these observation resolutions. However, crash analysis should account for the following fundamental considerations. *First*, macro level crashes are obtained by summing up the micro level crashes (i.e., zonal level crashes are obtained by summing up the crashes from road segment and intersection levels). *Second*, macro level independent variables are often aggregated from micro level information (i.e., in macro models zonal speed limit is obtained by considering weighted average of the speed limits of the roads of that zone). *Third*, both the macro level and micro level crashes are aggregated from disaggregate crash records (i.e., total crashes of a zone/road segment are aggregated from the individual crashes that occurred within the zone/road segment). Since both the dependent and independent variables for aggregate resolutions are obtained from the smaller resolutions, there could be potential relationships in the information transfer across these levels. In model systems where these dependent variables are modeled separately the embedded relationships within the data relating to the analysis resolutions are often neglected or ignored leading to a biased model output. Further, the recovered impact of independent variables might not reflect the true impact of these variables due to mis-specification.

In recent safety literature, several econometric model frameworks have addressed the limitations of the traditional independent model systems by introducing models that consider data from different analysis resolutions simultaneously. These multi-resolution data structure modeling efforts can be categorized in two directions: a) hierarchical models and b) integrated models. Within multi-level *hierarchical modeling* approaches, research efforts considered upper level (such as zone/corridor) information while modeling crashes at lower level (such as segment, intersection or crash record) to accommodate for unobserved heterogeneity and correlation in crash analysis. For instance, Huang and Abdel-Aty (2010) presented a potential hierarchical structure of geographic region level – traffic site level – traffic crash level – driver-vehicle unit level – occupant level and suggested to properly accommodate the potential cross-group heterogeneity

¹ For the macro level models, the choice of a certain unit (traffic analysis zones, census tracts or blocks) over an area of interest is closely related to the well-known modifiable areal unit problem (MAUP) and boundary crash issues. Several studies used various areal units and identified the level of data aggregation that minimizes the impact of the modifiable areal unit problem (Briz-Redón et al., 2019; Xu et al., 2018, 2014; Zhai et al., 2025) and the allocation of crashes that occur in the boundary of spatial analysis units while estimating crash frequency data (Ding et al., 2023).

and spatiotemporal correlation due to the multilevel data structure in crash analysis. Xie et al. (2014) modeled intersection crashes in high-density road networks and revealed strong evidence for the presence of heterogeneity across corridors and spatial correlation among intersections. Lee et al. (2017) developed intersection crash models for total, severe, pedestrian, and bicycle crashes with macro level sociodemographic and land-use data for spatial units and concluded that the intersection crash prediction models with macro level observed and unobserved variables outperform the models with intersection level variables only. Han et al. (2018) considered the hierarchical structure where road entity is nested within the geographic region and examined the variations in effect of road-level factors on crash frequency across different regions. Pew et al. (2020) considered a hierarchical structure of the intersections under urban area classification to allow the effects of a given intersection attribute to vary across urban classifications and found that allowing such hierarchy significantly improved the fit and prediction accuracy. Recent studies have introduced multiple membership multilevel models to estimate intersection crashes and pedestrian crashes (Park et al., 2022, 2020; Zhu et al., 2024). These studies employed a distance based weighting approach to account for the spatial dependency in safety modeling. The overall goal of these studies is to reduce heterogeneity issues between zones in the crash prediction model while avoiding model misspecification. Research efforts also employed hierarchical data structure models for crash outcome analysis (Islam et al., 2023; Kim et al., 2017; Park et al., 2017). All these research efforts suggest the consideration of zone/corridor level observed and unobserved variables in analyzing crashes at road entity or disaggregate level. Another group of studies within multilevel hierarchical models jointly explore traffic safety at the segment and intersection level, with the consideration of zonal/corridor-level and sub corridor-level variables (Alarifi et al., 2018a, 2018b, 2017; Wang and Huang, 2016).² These modeling efforts also captured spatial effects, the effect of different neighboring structures as well as the correlations among individual road entity and between adjacent entities. All these joint modeling efforts highlight that micro level models are improved while considering macro level observed and unobserved variables in the analysis. However, none of these studies considered variables from lower level to study crashes at upper level. In other words, these studies did not consider observed and unobserved variables from both levels simultaneously within the model structure to enhance the aggregate level predictions.

The *integrated model* systems – an emerging safety analysis area – improve on the other approaches by considering observed and unobserved variables from different analysis resolutions simultaneously within a unified framework. Within these approaches, studies explicitly recognize that macro level (zonal) crashes are obtained from the micro level (segments and intersections) crashes, and hence crash counts at the two levels are correlated (Cai et al., 2019; Pervaz et al., 2022). Thus, by considering both micro and macro level variables simultaneously, these integrated models found significantly improved macro level predictions. These integrated modeling efforts recognize that crash frequency data are generated by aggregating individual crash records and capture rich information flow from disaggregate level crash type and/or severity models to aggregate level crash frequency by crash type and/or severity estimation (Haddad et al., 2024; Pervaz et al., 2024, 2023). For instance, Pervaz et al. (2023) proposed a unified model system that

² The reader would note that joint modeling approaches were proposed to simultaneously model the crash counts by attributes including crash counts by crash severity levels (Afghari et al., 2020; Ahmad et al., 2023; Wang et al., 2021; Xie et al., 2019; Yasmin and Eluru, 2018), collision types (Bhowmik et al., 2021, 2019, 2018; Hosseinpour et al., 2018; Jahan et al., 2024), number of vehicles (Ahmad et al., 2023), and transport modes (Cai et al., 2017; Cheng et al., 2018; Huang et al., 2017; Yasmin et al., 2018). However, these studies considered a single analysis level such as zones, census tracts or census blocks for their analysis.

enhances the accuracy of the aggregate level crash frequency by severity estimation by incorporating the influence of independent variables from the crash record level severity model. The authors extended the framework by incorporating disaggregate level information from unordered crash type and ordered severity models within the aggregate level propensity to jointly estimate crash frequency by crash type and severity (Pervaz et al., 2024). Another study by Haddad et al. (2024) proposed a novel integrated parametric framework that links the information at the disaggregate crash level from an unordered model structure and the aggregate level crash count and found a significant positive association between the dimensions. These integrated frameworks used disaggregate level information such as crash-specific factors, driver and vehicle factors, roadway characteristics, road environmental and weather information, and facility level or aggregate level roadway and traffic factors, land-use and built environment attributes, and sociodemographic characteristics for the analysis. All these integrated frameworks illustrated that consideration of additional observed and unobserved information from different observation resolutions simultaneously within the unified framework enhances the performance of macro or aggregate level crash frequency estimation. In addition, these integrated approaches can capture systemwide effects and individual level information simultaneously that can contribute to minimizing MAUP and boundary-crash issues.

2. STUDY CONTEXT

The current study builds on the aforementioned integrated modeling approaches by accommodating information from multiple spatial units and analysis resolutions (as shown in **Figure 1**). The proposed approach seamlessly integrates data from multiple disaggregate levels within the appropriate aggregate levels thus generalizing the overall crash analysis processes. In Pervaz et al. (2022), micro (segment and intersection facility) level information was incorporated in the macro (zone) level crash frequency models (Pervaz et al., 2022). On the other hand, Pervaz et al. (2023, 2024) incorporated disaggregate level (crash record) information in the modeling of zonal level crash frequency by crash type and/or severity (Pervaz et al., 2024, 2023). Haddad et al. (2024) developed an integrated parametric framework linking the information at the disaggregate crash level and the aggregate level (census block group) crash count (Haddad et al., 2024). These frameworks consider either macro-micro or aggregate-disaggregate combination to estimate zonal crashes and highlight that the estimated models offer superior performances than the independent model system. Therefore, it might be beneficial to consider macro–micro–disaggregate combination to accommodate for the information flow from crash record level–segment/intersection level–zonal level while also considering hierarchical relationships across the levels. The approach would involve summing up the crash propensity of each disaggregate level (crash record) within the micro resolutions (segment and intersection facility) and adding the generated values as new variables in the micro level propensities while also considering summed up micro level crash propensities as new variables for the macro level (zone) propensity equation. To summarize, in our current paper, a unified framework that explicitly allows for the information flow of observed and unobserved variables from the crash record level models into the micro and from the micro level models to macro level crash frequency by severity model is proposed and estimated.

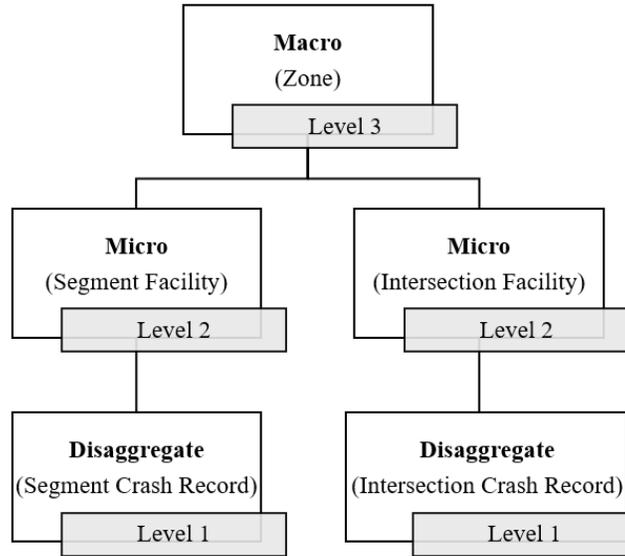


Figure 1: Hierarchical Data Structure of the Study

In our study, econometric building blocks including negative binomial-ordered probit fractional split (NB-OPFS) and ordered probit (OP) are employed to develop the integrated framework. For independent model system, the NB-OPFS framework can be employed separately at micro (for segment and intersection facility) and macro level (zone) to jointly estimate crash frequency by severity where the NB component models the total crashes and the OPFS component determines the proportion of crashes by severity class for each spatial analysis level. At disaggregate level, the crash severity variable can be examined using an OP model for crash records from each facility type (segment and intersection). The integrated model system can employ these econometric building blocks to consider information flow from the disaggregate and micro levels through propensity equations within a unified framework to estimate crashes at macro level. With these models, the integrated approach can take four potential forms as shown in **Table 1**. The different forms of integrated models arise in how the propensity sum variables from the lower-level models are accommodated as independent variables in the upper-level models. The propensity sum variables can be treated as fixed (as obtained from the micro/disaggregate model predictions) or allowed to vary and re-estimated on the integrated model. The different levels and the different decisions to fix or allow the parameters to vary will result in different integrated approaches.

Table 1: Integrated Modeling Approaches

Integrated Model Approaches	Count Component (NB Model Propensity)			Fraction Component (OPFS Model Propensity)		
	Macro Level Parameters	Micro Level Parameters	Disaggregate Level Parameters	Macro Level Parameters	Micro Level Parameters	Disaggregate Level Parameters
Integrated Approach 1	Allowed to vary	Fixed	Fixed	Allowed to vary	Fixed	Fixed
Integrated Approach 2	Allowed to vary	Allowed to vary	Fixed	Allowed to vary	Allowed to vary	Fixed
Integrated Approach 3	Allowed to vary	Fixed	Allowed to vary	Allowed to vary	Fixed	Allowed to vary

Integrated Model Approaches	Count Component (NB Model Propensity)			Fraction Component (OPFS Model Propensity)		
	Macro Level Parameters	Micro Level Parameters	Disaggregate Level Parameters	Macro Level Parameters	Micro Level Parameters	Disaggregate Level Parameters
Integrated Approach 4	Allowed to vary	Allowed to vary	Allowed to vary	Allowed to vary	Allowed to vary	Allowed to vary
Number of Additional Coefficients Estimated						
Integrated Approaches 1-4	Count Component (NB Model Propensity)			Fraction Component (OPFS Model Propensity)		
	Micro level propensity sums: 2 [Segment facility: 1, Intersection facility: 1] Disaggregate level propensity sums: 4 [Segment crash record within micro: 1, Segment crash record within macro: 1, Intersection crash record within micro: 1, Intersection crash record within macro: 1]			Micro level propensity sums: 2 [Segment facility: 1, Intersection facility: 1] Disaggregate level propensity sums: 4 [Segment crash record within micro: 1, Segment crash record within macro: 1, Intersection crash record within micro: 1, Intersection crash record within macro: 1]		

Table 1 outlines the different variants based on how the propensity sum variables can be treated as fixed as obtained from the micro/disaggregate model predictions (Approach 1) and the parameters within the propensity sum variables can be treated as endogenous and be allowed to vary (Approaches 2, 3, and 4). Approach 4 is computationally more involved as it allows for feedback from all the observational resolutions simultaneously. **Table 1** also presents the number of additional parameters estimated for composite variables (propensity sum scores) in the integrated model system. The different approaches can be estimated and the data fit of these approaches can be compared to identify the most appropriate model for the data considered (more details are provided in the methodology section). The model selection process can be accomplished using model fit measures such as Bayesian Information Criterion (BIC). Finally, it is important to highlight that the integrated approach accommodates for a host of unobserved factors within each econometric building block and across all econometric blocks (exact formulation details are included in the methodology section).

The proposed model system is estimated using data drawn from the City of Orlando, Florida for the year 2019. The study obtained 21,189 crash records (5,669 segment and 15,520 intersection crashes) for the disaggregate level model analysis. The records contain crash-specific factors, driver and vehicle factors, roadway attributes, road environmental and weather information of each crash record. For micro and macro level analysis, the study aggregated the crash records over 1,818 segments, 4,184 intersections and 300 traffic analysis zones (TAZs). An exhaustive set of independent variables including roadway and traffic factors, land-use and built environment attributes, and sociodemographic characteristics are considered in these levels.

3. METHODOLOGY

In this study, a negative binomial-ordered probit fractional split (NB-OPFS) model structure is employed to analyze crash frequency by severity at the micro level and macro level model analysis. Alternatively, for disaggregate level analysis an ordered probit (OP) model structure is employed. The NB-OPFS and OP model structures are employed to develop the integrated framework. Thus, the overall econometric framework of the study can be described along three components: the disaggregate level model structure (ordered probit), micro level and macro level model structure (NB-OPFS), and the integrated model structure.

3.1 Disaggregate Level Model Structure (Ordered Probit Model)

In the traditional ordered response model, the discrete injury severity levels ($v_{j(s/i)}$) for segment s or intersection i are assumed to be associated with an underlying continuous latent variable ($v_{j(s/i)}^*$). This latent variable is typically specified as the following linear function:

$$v_{j(s/i)}^* = X_{j(s/i)} \Theta_{(s/i)} + \theta_{(s/i)} + \varepsilon_{j(s/i)} \quad (1)$$

where $j_{(s/i)} = 1, 2, \dots, J_{(s/i)}$ represents the crash record for segment s or intersection i . $X_{j(s/i)}$ is a vector of exogenous variables for segment s or intersection i (excluding a constant). $\Theta_{(s/i)}$ is a vector of unknown parameters to be estimated for segment s or intersection i . $\theta_{(s/i)}$ is a vector of unobserved effects specific to the facility type (segment s or intersection i) for the crash records, highlighting the spatial arrangement within the same facility. This $\theta_{(s/i)}$ will be same across the crash records if they correspond to the same segment or intersection and thus the spatial dependency will be captured. The reader would note that the spatial unobserved heterogeneity can vary across the crash records. Therefore, in the current study, we parameterize the correlation parameter $\theta_{(s/i)}$ as a function of observed attributes as follows:

$$\theta_{(s/i)} = G_{(s/i)} R_{(s/i)} \quad (2)$$

where $R_{(s/i)}$ is a vector of exogenous variables at the facility level (including a constant) employed for crash records, $G_{(s/i)}$ is a vector of parameters to be estimated. $\varepsilon_{j(s/i)}$ is the random disturbance term assumed to be standard normal distribution. Let us assume k ($k = 1, 2, 3, \dots, k$) be the index to represent injury severity categories. In this study, k take the values of ‘no-injury’ ($k = 1$), ‘possible injury’ ($k = 2$), ‘non-incapacitating injury’ ($k = 3$) and ‘fatal and incapacitating injury’ ($k = 4$). $t_{k(s/i)}$ represents the thresholds associated with these severity levels. These unknown t_k s are assumed to partition the propensity into $k - 1$ intervals. The unobservable latent variable $v_{j(s/i)}^*$ is related to the observable ordinal variable $v_{j(s/i)}$ by the $t_{k(s/i)}$ with a response mechanism of the following form:

$$v_{j(s/i)} = k, \text{ if } t_{(s/i)(k-1)} < v_{j(s/i)}^* < t_{(s/i)k}, \text{ for } k = 1, 2, \dots, k \quad (3)$$

In order to ensure the well-defined intervals and natural ordering of observed severity, the thresholds are assumed to be ascending in order, such that $t_{(s/i)0} < t_{(s/i)1} < \dots < t_{(s/i)k}$ where $t_{(s/i)0} = -\infty$ and $t_{(s/i)k} = +\infty$. Given these relationships across the different parameters, the resulting probability expressions for record $j_{(s/i)}$ and alternative k for the ordered probit take the following form:

$$\begin{aligned} \pi_{j(s/i)k} &= Pr \left(v_{j(s/i)} = k \mid X_{j(s/i)} \right) \\ &= Y \left(t_{(s/i)k} - (X_{j(s/i)} \Theta_{(s/i)} + \theta_{(s/i)}) \right) - Y \left(t_{(s/i)(k-1)} \right. \\ &\quad \left. - (X_{j(s/i)} \Theta_{(s/i)} + \theta_{(s/i)}) \right) \end{aligned} \quad (4)$$

where $Y(\cdot)$ represents the standard normal cumulative distribution function.

3.2 Micro and Macro Level Model Structure (NB-OPFS Model)

3.2.1 Count framework

In our research, crash counts per TAZ (for macro), per segment and per intersection (for micro) are considered as the dependent variables in the model estimations. Since crash counts are non-negative integers, count-data modeling techniques are appropriate for crash frequency data analysis. Among the count-modeling techniques, a Poisson model may not always be appropriate because the Poisson distribution restricts the mean and variance to be equal whereas crash frequency data are generally over-dispersed (Lord and Mannering, 2010). Therefore, in our research, we consider the negative binomial regression framework to account for over-dispersed data. The approach can be extended readily to any possible mathematical model such as Poisson lognormal and other variants of negative binomial such as the zero inflated negative binomial model.

For a spatial unit l (segment s , intersection i , or zone z), negative binomial (NB) model is employed to estimate total crash count. The probability density function of NB model can be written as,

$$P(c_l) = \frac{\Gamma\left(c_l + \frac{1}{\alpha_l}\right)}{\Gamma(c_l + 1)\Gamma\left(\frac{1}{\alpha_l}\right)} \left(\frac{1}{1 + \alpha_l\mu_l}\right)^{\frac{1}{\alpha_l}} \left(1 - \frac{1}{1 + \alpha_l\mu_l}\right)^{c_l} \quad (5)$$

where c_l be the index for crashes occurring over a period of time in a spatial unit l (segment s , intersection i , or zone z). $P(c_l)$ is the probability that unit l has c_l number of crashes. $\Gamma(\cdot)$ is the gamma function, α_l is negative binomial overdispersion parameter and μ_l is the expected number of crashes occurring in the unit l over a given time period. The equation for μ_l can be written as follows,

$$\mu_l = E(c_l|Y_l) = \exp((\delta_l + \zeta_l)Y_l + \theta_{z(s/i)} + \varepsilon_l + \eta_l) \quad (6)$$

where Y_l is a vector of explanatory variables associated with the analysis unit l . δ_l is a vector of coefficients to be estimated for the unit l . ζ_l a vector of unobserved factors on crash count propensity for unit l . $\theta_{z(s/i)}$ is a vector of unobserved effects specific to the zone exclusively for segments or intersections, highlighting the spatial arrangement of segments or intersections within the same zone. This θ_z will be same across all the segments or all the intersections if they correspond to the same zone, highlighting the spatial dependency. This spatial unobserved heterogeneity can vary across the segments or intersections. Therefore, in the current study, we parameterize the correlation parameter θ_z as a function of observed attributes as follows:

$$\theta_{z(s/i)} = G_{z(s/i)}R_{z(s/i)} \quad (7)$$

where $R_{z(s/i)}$ is a vector of exogenous variables at zonal level (including a constant) employed for facility type (segment or intersection), $G_{z(s/i)}$ is a vector of parameters to be estimated. ε_l is a gamma distributed error term with mean 1 and variance α_l . η_l captures the influence of common unobserved factors that impact the total number of crashes and proportion of crashes by severity for unit l .

3.2.2 Fractional split framework

The modeling of crash proportions by severity levels is undertaken using the ordered probit fractional split model (OPFS). In the ordered outcome framework, the actual injury severity proportions (y_{lk}) are assumed to be associated with an underlying continuous latent variable (y_l^*) as follows:

$$y_l^* = ((\boldsymbol{\beta}_l + \boldsymbol{\Sigma}_l)\mathbf{F}_l + \boldsymbol{\theta}_{z(s/i)} + \xi_l \pm \eta_l), y_{lk} = k \text{ if } \tau_{l(k-1)} < y_l^* < \tau_{lk} \quad (8)$$

The latent propensity y_l^* is mapped to the actual severity proportion categories y_{lk} by τ_l thresholds ($\tau_{l0} = -\infty$ and $\tau_{lK} = +\infty$) as presented in equation 8. \mathbf{F}_l is a vector of attributes (not including a constant) that influences the propensity associated with severity proportion categories for unit l . $\boldsymbol{\beta}_l$ is the corresponding vector of mean effects. $\boldsymbol{\Sigma}_l$ a vector of unobserved factors on severity proportion propensity for unit l . ξ_l is an idiosyncratic error term assumed to be identically and independently standard normally distributed across unit l . η_l term generates the correlation between equations for total number of crashes and crash proportions by severity levels and also allows for considering the influence of various unobserved factors affecting the frequency and proportion variables. The \pm sign in front of η_l indicates that the correlation in unobserved individual factors between total crashes and crash proportions by severity levels may be positive or negative. A positive sign implies that units with higher number of crashes are intrinsically more likely to incur higher proportions for severe crashes. On the other hand, negative sign implies that units with higher number of crashes intrinsically incur lower proportions for severe crashes. To determine the appropriate sign one can empirically test the models with both '+' and '-' signs independently. The model structure that offers superior data fit is considered as the final model.

It is important to note here that the unobserved heterogeneity between total crashes and crash proportions by severity levels can vary across units. Therefore, in the current study, the correlation parameter η_l is parameterized as a function of observed attributes as follows:

$$\eta_l = \mathbf{M}_l \mathbf{Q}_l \quad (9)$$

where \mathbf{Q}_l is a vector of exogenous variables, \mathbf{M}_l is a vector of unknown parameters to be estimated (including a constant).

To estimate the model presented in equation 8, we assume that:

$$E(y_{lk} | \mathbf{F}_l) = H_{lk}(\boldsymbol{\beta}_l, \tau_l), 0 \leq H_{lk} \leq 1, \sum_{k=1}^K H_{lk} = 1 \quad (10)$$

H_{lk} in our model takes the ordered probit probability (Λ) form for the severity category k .

Given these relationships across different parameters, the resulting probability (Λ) for the ordered probit fractional split model takes the following form:

$$\Lambda(y_{lk} = k) = \varphi\{\tau_{lk} - y_l^*\} - \varphi\{\tau_{l(k-1)} - y_l^*\} \quad (11)$$

where $\varphi(\cdot)$ is the standard normal cumulative distribution function.

3.3 Integrated Model Structure

To develop integrated model framework, we use propensity equations from disaggregate level OP model structure, and micro and macro level NB-OPFS model structure. To be specific, we incorporate composite sums of the disaggregate level OP model propensities (from segment crash OP model and intersection crash OP model) and sums for the micro level NB-OPFS model propensities (from segment facility NB-OPFS model and intersection facility NB-OPFS model) within the macro level (zonal) propensity equation to estimate crash frequency by severity. For example, if 5 crashes occur in a segment, we sum the OP model propensity for these 5 crashes and the composite score (summation of propensities) is incorporated as an independent variable within the NB-OPFS model equation for segment facility. Similarly, if a zone has 5 segments, we sum the propensities from NB and OPFS model components and incorporate the composite scores (one from NB and one from OPFS) within the zonal level crash propensities to estimate crash frequency by severity. We can also incorporate disaggregate level composite sum directly into the zonal propensity equation. For each composite score, we estimate a scalar parameter (coefficient) in the model system. These scalar parameters can be estimated by considering the composite scores as fixed value obtained from the respective OP or NB-OPFS model components or be allowed to vary based on the model fit. With these considerations, four model structures can be estimated. For NB model component, the propensity structures are:

$$\begin{aligned}
\mu'_z = E(c_z|Y_z) = & \exp \left((\delta_z + \zeta_z)Y_z + \rho_{c_s} * \ln \sum_{p=1}^{sz} \left(\exp \left(\ln \mu_s + \rho_{c_{ds}} \ln \sum_{p=1}^{sj_s} \left(\exp(v_{j_s}^*) \right) \right) \right) \right) \\
& + \rho_{c_{dsz}} \ln \sum_{p=1}^{zj_s} \left(\exp(v_{j_s}^*) \right) + \rho_{c_i} \\
& * \ln \sum_{p=1}^{iz} \left(\exp \left(\ln \mu_i + \rho_{c_{di}} \ln \sum_{p=1}^{ij_i} \left(\exp(v_{j_i}^*) \right) \right) \right) \\
& + \rho_{c_{diz}} \ln \sum_{p=1}^{zj_i} \left(\exp(v_{j_i}^*) \right) + \varepsilon_z + \eta_z
\end{aligned} \tag{12}$$

$$\begin{aligned}
\mu'_z = E(c_z|Y_z) = & \exp \left((\delta_z + \zeta_z)Y_z + \rho_{c_s} \right. \\
& * \ln \sum_{p=1}^{sz} \left(\exp \left((\delta_s + \zeta_s)Y_s + \theta_{z(s)} + \rho_{c_{ds}} \ln \sum_{p=1}^{sj_s} (\exp(v_{j_s}^*)) \right) \right) \\
& + \rho_{c_{dsz}} \ln \sum_{p=1}^{zj_s} (\exp(v_{j_s}^*)) + \rho_{c_i} \\
& * \ln \sum_{p=1}^{iz} \left(\exp \left((\delta_i + \zeta_i)Y_i + \theta_{z(i)} + \rho_{c_{di}} \ln \sum_{p=1}^{ij_i} (\exp(v_{j_i}^*)) \right) \right) \\
& \left. + \rho_{c_{diz}} \ln \sum_{p=1}^{zj_i} (\exp(v_{j_i}^*)) + \varepsilon_z + \eta_z \right)
\end{aligned} \tag{13}$$

$$\begin{aligned}
\mu'_z = E(c_z|Y_z) = & \exp \left((\delta_z + \zeta_z)Y_z + \rho_{c_s} \right. \\
& * \ln \sum_{p=1}^{sz} \left(\exp \left(\ln \mu_s + \rho_{c_{ds}} \ln \sum_{p=1}^{sj_s} (\exp(X_{j_s}\theta_s + \theta_s + \varepsilon_{j_s})) \right) \right) \\
& + \rho_{c_{dsz}} \ln \sum_{p=1}^{zj_s} (\exp(X_{j_s}\theta_s + \theta_s + \varepsilon_{j_s})) + \rho_{c_i} \\
& * \ln \sum_{p=1}^{iz} \left(\exp \left(\ln \mu_i + \rho_{c_{di}} \ln \sum_{p=1}^{ij_i} (\exp(X_{j_i}\theta_i + \theta_i + \varepsilon_{j_i})) \right) \right) \\
& \left. + \rho_{c_{diz}} \ln \sum_{p=1}^{zj_i} (\exp(X_{j_i}\theta_i + \theta_i + \varepsilon_{j_i})) + \varepsilon_z + \eta_z \right)
\end{aligned} \tag{14}$$

$$\begin{aligned}
\mu'_z = E(c_z|Y_z) = & \exp \left((\delta_z + \zeta_z)Y_z + \rho_{c_s} \right. \\
& * \ln \sum_{p=1}^{sz} \left(\exp \left((\delta_s + \zeta_s)Y_s + \theta_{z(s)} \right. \right. \\
& \left. \left. + \rho_{c_{ds}} \ln \sum_{p=1}^{sj_s} \left(\exp(X_{j_s}\theta_s + \theta_s + \varepsilon_{j_s}) \right) \right) \right) \\
& \left. + \rho_{c_{dsz}} \ln \sum_{p=1}^{zj_s} \left(\exp(X_{j_s}\theta_s + \theta_s + \varepsilon_{j_s}) \right) + \rho_{c_i} \right. \\
& * \ln \sum_{p=1}^{iz} \left(\exp \left((\delta_i + \zeta_i)Y_i + \theta_{z(i)} \right. \right. \\
& \left. \left. + \rho_{c_{di}} \ln \sum_{p=1}^{ij_i} \left(\exp(X_{j_i}\theta_i + \theta_i + \varepsilon_{j_i}) \right) \right) \right) \\
& \left. + \rho_{c_{diz}} \ln \sum_{p=1}^{zj_i} \left(\exp(X_{j_i}\theta_i + \theta_i + \varepsilon_{j_i}) \right) + \varepsilon_z + \eta_z \right)
\end{aligned} \tag{15}$$

where ρ_{c_s} is the scalar (coefficient) for segment facility propensity sum for count component, $\rho_{c_{ds}}$ is the scalar for disaggregate segment crash propensity sum within segment facility propensity for count component, $\rho_{c_{dsz}}$ is the scalar for disaggregate segment crash propensity sum within zone for count component, ρ_{c_i} is the scalar for intersection facility propensity sum for count component, $\rho_{c_{di}}$ is the scalar for disaggregate intersection crash propensity sum within intersection facility propensity for count component, $\rho_{c_{diz}}$ is the scalar for disaggregate intersection crash propensity sum within zone for count component.

In the first structure, the micro level and disaggregate model parameters are fixed (as obtained from facility-specific NB models and OP models) and only scalar parameters will be estimated along with the macro level variables. In the second approach, micro level parameters within the propensity will be allowed to vary based on model fit while estimating the scalar parameters and the macro level variables. In the third approach, disaggregate level parameters within the propensity will be allowed to vary based on model fit and in the fourth approach, both micro level and disaggregate level parameters will be allowed to vary while estimating the scalar parameters and the macro level parameters. Similar to the count component, the equations for OPFS component are:

$$\begin{aligned}
y_z^{*'} &= \left((\boldsymbol{\beta}_z + \boldsymbol{\Sigma}_z) \mathbf{F}_z + \rho_{f_s} * \ln \sum_{p=1}^{sz} \left(\exp \left(y_s^* + \rho_{f_{ds}} \ln \sum_{p=1}^{s_{j_s}} \left(\exp(v_{j_s}^*) \right) \right) \right) \right) + \\
&\rho_{f_{dsz}} \ln \sum_{p=1}^{z_{j_s}} \left(\exp(v_{j_s}^*) \right) + \rho_{f_i} * \ln \sum_{p=1}^{iz} \left(\exp \left(y_i^* + \rho_{f_{di}} \ln \sum_{p=1}^{i_{j_i}} \left(\exp(v_{j_i}^*) \right) \right) \right) + \\
&\rho_{f_{diz}} \ln \sum_{p=1}^{z_{j_i}} \left(\exp(v_{j_i}^*) \right) + \xi_z \pm \eta_z \Big), y'_{zk} = k \text{ if } \tau_{z(k-1)} < y_z^{*'} < \tau_{zk}
\end{aligned} \tag{16}$$

$$\begin{aligned}
y_z^{*'} &= \left((\boldsymbol{\beta}_z + \boldsymbol{\Sigma}_z) \mathbf{F}_z + \rho_{f_s} \right. \\
&* \ln \sum_{p=1}^{sz} \left(\exp \left((\boldsymbol{\beta}_s + \boldsymbol{\Sigma}_s) \mathbf{F}_s + \boldsymbol{\theta}_{z(s)} + \rho_{f_{ds}} \ln \sum_{p=1}^{s_{j_s}} \left(\exp(v_{j_s}^*) \right) \right) \right) \\
&+ \rho_{f_{dsz}} \ln \sum_{p=1}^{z_{j_s}} \left(\exp(v_{j_s}^*) \right) + \rho_{f_i} \\
&* \ln \sum_{p=1}^{iz} \left(\exp \left((\boldsymbol{\beta}_i + \boldsymbol{\Sigma}_i) \mathbf{F}_i + \boldsymbol{\theta}_{z(i)} + \rho_{f_{di}} \ln \sum_{p=1}^{i_{j_i}} \left(\exp(v_{j_i}^*) \right) \right) \right) \\
&\left. + \rho_{f_{diz}} \ln \sum_{p=1}^{z_{j_i}} \left(\exp(v_{j_i}^*) \right) + \xi_z \pm \eta_z \right), y'_{zk} = k \text{ if } \tau_{z(k-1)} < y_z^{*'} < \tau_{zk}
\end{aligned} \tag{17}$$

$$\begin{aligned}
y_z^{*'} &= \left((\boldsymbol{\beta}_z + \boldsymbol{\Sigma}_z) \mathbf{F}_z + \rho_{f_s} * \ln \sum_{p=1}^{sz} \left(\exp \left(y_s^* + \rho_{f_{ds}} \ln \sum_{p=1}^{s_{j_s}} \left(\exp(X_{j_s} \boldsymbol{\theta}_s + \boldsymbol{\theta}_s + \varepsilon_{j_s}) \right) \right) \right) \right) \\
&+ \rho_{f_{dsz}} \ln \sum_{p=1}^{z_{j_s}} \left(\exp(X_{j_s} \boldsymbol{\theta}_s + \boldsymbol{\theta}_s + \varepsilon_{j_s}) \right) + \rho_{f_i} \\
&* \ln \sum_{p=1}^{iz} \left(\exp \left(y_i^* + \rho_{f_{di}} \ln \sum_{p=1}^{i_{j_i}} \left(\exp(X_{j_i} \boldsymbol{\theta}_i + \boldsymbol{\theta}_i + \varepsilon_{j_i}) \right) \right) \right) \\
&+ \rho_{f_{diz}} \ln \sum_{p=1}^{z_{j_i}} \left(\exp(X_{j_i} \boldsymbol{\theta}_i + \boldsymbol{\theta}_i + \varepsilon_{j_i}) \right) + \xi_z \pm \eta_z \Big), y'_{zk} = k \text{ if } \tau_{z(k-1)} < y_z^{*'} \\
&< \tau_{zk}
\end{aligned} \tag{18}$$

$$\begin{aligned}
y_z^{*'} = & \left((\boldsymbol{\beta}_z + \boldsymbol{\Sigma}_z) \mathbf{F}_z + \rho_{f_s} \right. \\
& * \ln \sum_{p=1}^{sz} \left(\exp \left((\boldsymbol{\beta}_s + \boldsymbol{\Sigma}_s) \mathbf{F}_s + \boldsymbol{\theta}_{z(s)} \right. \right. \\
& \left. \left. + \rho_{f_{ds}} \ln \sum_{p=1}^{sj_s} \left(\exp(X_{j_s} \boldsymbol{\theta}_s + \boldsymbol{\theta}_s + \varepsilon_{j_s}) \right) \right) \right) \\
& + \rho_{f_{dsz}} \ln \sum_{p=1}^{zj_s} \left(\exp(X_{j_s} \boldsymbol{\theta}_s + \boldsymbol{\theta}_s + \varepsilon_{j_s}) \right) + \rho_{f_i} \\
& * \ln \sum_{p=1}^{iz} \left(\exp \left((\boldsymbol{\beta}_i + \boldsymbol{\Sigma}_i) \mathbf{F}_i + \boldsymbol{\theta}_{z(i)} \right. \right. \\
& \left. \left. + \rho_{f_{di}} \ln \sum_{p=1}^{ij_i} \left(\exp(X_{j_i} \boldsymbol{\theta}_i + \boldsymbol{\theta}_i + \varepsilon_{j_i}) \right) \right) \right) \\
& \left. + \rho_{f_{diz}} \ln \sum_{p=1}^{zj_i} \left(\exp(X_{j_i} \boldsymbol{\theta}_i + \boldsymbol{\theta}_i + \varepsilon_{j_i}) \right) + \xi_z \pm \eta_z \right), y'_{zk} = k \text{ if } \tau_{z(k-1)} < y_z^{*'} \\
& < \tau_{zk}
\end{aligned} \tag{19}$$

where ρ_{f_s} is the scalar (coefficient) for segment facility propensity sum for severity proportion, $\rho_{f_{ds}}$ is the scalar for disaggregate segment crash propensity sum within segment facility propensity for severity proportion, $\rho_{f_{dsz}}$ is the scalar for disaggregate segment crash propensity sum within zone for severity proportion, ρ_{f_i} is the scalar for intersection facility propensity sum for severity proportion, $\rho_{f_{di}}$ is the scalar for disaggregate intersection crash propensity sum within intersection facility propensity for severity proportion, and $\rho_{f_{diz}}$ is the scalar for disaggregate intersection crash propensity sum within zone for severity proportion.

The correlation parameter η_z is parameterized as a function of observed zonal attributes as follows (with the notation from equation 9):

$$\eta_z = M_z \mathbf{Q}_z \tag{20}$$

With the integrated propensity, the updated probability equation for macro level NB is:

$$P'(c_z) = \frac{\Gamma(c_z + \frac{1}{\alpha_z})}{\Gamma(c_z + 1) \Gamma(\frac{1}{\alpha_z})} \left(\frac{1}{1 + \alpha_z \mu'_z} \right)^{\frac{1}{\alpha_z}} \left(1 - \frac{1}{1 + \alpha_z \mu'_z} \right)^{c_z} \tag{21}$$

And fractional split component probability for macro level is,

$$\Lambda(y'_{zk} = k) = \varphi\{\tau_{zk} - y_z^{*'}\} - \varphi\{\tau_{z(k-1)} - y_z^{*'}\} \quad (22)$$

3.4 Model Estimation

In examining the model structure of total crash count and proportions of crashes by severity level, it is necessary to specify the structure for the unobserved vectors \mathbf{G} , \mathbf{M} , $\boldsymbol{\zeta}$, $\boldsymbol{\varepsilon}$ represented by Ω . In this study, it is assumed that the elements are drawn from independent realization from normal population: $\Omega \sim N(0, (\boldsymbol{\sigma}_1^2, \boldsymbol{\sigma}_2^2, \boldsymbol{\sigma}_3^2, \boldsymbol{\sigma}_4^2))$. Thus, conditional on Ω , the likelihood function for the integrated probability can be expressed as:

$$\begin{aligned} L_z = \int_{\Omega} & P'(c_z) \times \prod_{p=1}^{Sz} P(c_s)^{w_s} \times \prod_{p=1}^{Iz} P(c_i)^{w_i} \times \prod_{k=1}^K (\Lambda(y'_{zk} = k))^{\varpi_z d_{zk}} \\ & \times \prod_{p=1}^{Sz} \prod_{k=1}^K (\Lambda(y_{sk} = k))^{\varpi_s d_{sk}} \times \prod_{p=1}^{Iz} \prod_{k=1}^K (\Lambda(y_{ik} = k))^{\varpi_i d_{ik}} \\ & \times \prod_{p=1}^{j_{sz}} \prod_{k=1}^K \pi_{j_{sk}} \times \prod_{p=1}^{j_{iz}} \prod_{k=1}^K \pi_{j_{ik}} d\Omega \end{aligned} \quad (23)$$

where w_s (w_i) is a dummy variable taking a value of 1 if the corresponding zone z has segments (intersections) in it and 0 otherwise. ϖ_z is a dummy with $\varpi_z = 1$ if zone z has at least one crash over the study period and 0 otherwise. d_{zk} is the proportion of crashes in severity category k in zone z . ϖ_s (ϖ_i) is a dummy with ϖ_s (ϖ_i) = 1 if segment s (intersection i) has at least one crash over the study period and 0 otherwise. d_{sk} is the proportion of crashes in severity category k at segment s and d_{ik} is the proportion of crashes in severity category k at intersection i . Finally, the log-likelihood function is:

$$LL = \sum_z \text{Ln}(L_z) \quad (24)$$

All the parameters in the model are estimated by maximizing the logarithmic function LL presented in equation 24. To estimate the proposed model, we apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate the integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function across individuals (please see Bhat, 2001; Yasmin and Eluru, 2013 for details). In our research, we tested the model specification with several Halton realization levels (such as 50, 100, 300, ..., 500, 1000). We found that model parameters were stable at around 500 and 1000 levels. We use the GAUSS matrix programming software to run the models (Aptech, 2015).

4. DATA PREPARATION

The current study conducts empirical analysis using data from the Orlando city region of Florida. The study area is composed of 300 traffic analysis zones (TAZs), 1,818 segments, and 4,184 intersections. The study extracted crash record level data for the year 2019 from Signal Four Analytics database. Each record contains details of crash level information. After processing and

cleaning the data, we obtained 21,189 crash records (5,669 segment and 15,520 intersection crashes) for the disaggregate level analysis. For assigning intersection crashes, a 250 feet buffer around the center of each intersection was considered and the crashes were spatially assigned by using ArcGIS tools (see Cai et al., 2019; Ingle and Gates, 2021; Sharafeldin et al., 2022; Yue, 2024 that adopted this approach). Based on the severity class, these crashes could be classified into a five-point severity scale: fatal injury (FI), incapacitating injury (II), non-incapacitating injury (NII), possible injury (PI), and no-injury crashes (NI). The distributions of FI, II, NII, PI and NI are 0.32%, 1.89%, 8.20%, 17.27%, and 72.32%, respectively for segment facility and 0.28%, 1.66%, 8.94%, 18.88%, and 70.24%, respectively for intersection facility. This study combines FI and II as FII for the disaggregate level models estimation.

For micro and macro levels model estimation, the study aggregated crash records by facility level and TAZ level, respectively. For micro and macro level crash count components, total crash counts at the observation resolutions are considered as the dependent variable while for severity proportion components, crash proportions by severity level (number of crashes of specific severity level/total crashes) are considered as the dependent variable. The four severity proportions are: (1) proportion of no-injury crashes (PNI), (2) proportion of possible injury crashes (PPI), (3) proportion of non-incapacitating injury crashes (PNII), and (4) proportion of fatal and incapacitating injury crashes (PFII). A comprehensive set of independent variables including roadway and traffic factors, land-use and built environment attributes, and sociodemographic characteristics are considered in these levels. This study randomly selects 270 TAZs that cover 1,613 segments with 4,635 crash records, and 3,746 intersections with 13,784 crash records for model estimation and set aside the remaining 30 TAZs that include 205 segments with 1,034 crash records, and 438 intersections with 1,736 crash records for the validation of the models.

4.1 Variables Considered

The variables for disaggregate, micro, and macro level analysis were sourced from different databases including Signal Four Analytics (S4A), Florida Department of Transportation (FDOT) Transportation Statistics Division, US Census Bureau and American Community Survey, Florida Department of Revenue, and Florida Geographic Data Library. Disaggregate level segment and intersection OP models consider the information present in the crash records including crash-specific variables (such as first harmful events), driver and vehicle factors (such as driving under influence related, distraction related, presence of passengers), roadway factors (such as location of the crashes, speed limit, shoulder type), road environmental factors (such as time of the day, lighting condition) and weather information (such as clear, rain, fog).

For the segment facility level, the variables were spatially assigned to the segments while for intersection facility level, variables were aggregated by taking 0.5-mile buffer zone around each intersection by using ArcGIS tools. The segment level variables include roadway and traffic factors (such as AADT, Truck AADT, functional class, number of lanes, speed limit, median width, and shoulder width). The intersection level variables include roadway and traffic factors (such as AADT, proportion of roads by functional class, number of lanes, and average shoulder width), land-use attributes (such as proportion of residential, commercial, institutional, industrial, recreational and mixed area), built environment attributes (such as number of restaurants, business centers, commercial centers, educational centers, and shopping centers), and sociodemographic characteristics (such as population density, proportion of males and females, household density, median household income, proportion of car, drive alone, non-motorized means of transport, different population group by age level, household with vehicle availability, and population with

different races). For macro level analysis, the explanatory variables were aggregated at the TAZ level. Macro level analysis uses roadway and traffic factors (such as AADT, truck AADT, proportion of roads by functional class, number of lanes, average speed limit, average shoulder width, average sidewalk width and median width, intersection density, and traffic signal density), and land-use, built environment, and sociodemographic characteristics similar to the intersection facility.

In estimating the model, several functional forms, and combination of variables are considered and those that provide the best fit are retained in the final specification. The final specification of the model was based on removing the statistically insignificant variables in a systematic process based on 90% confidence level.

Figure 2 shows the sample share of the variables at disaggregate levels considered for the final model estimation while the micro level and macro level variables are presented in **Table 2** with the appropriate definition and summary statistics.

5. EMPIRICAL ANALYSIS

5.1 Model Specification and Overall Measure of Fit

A series of models are estimated to conduct empirical analysis of the proposed framework and the independent model systems. First, we estimate OP models for segment crash records and intersection crash records, NB-OPFS models for segment facility, intersection facility and zonal level to estimate crash counts by severity class. These model frameworks together provide the non-integrated model system. Second, we develop our proposed integrated model system following four approaches as discussed in the methodology section.

Integrated approach 1: Focuses on optimizing the joint log-likelihood of the macro, micro, and disaggregate level models by only estimating the scalar parameters for the micro level and disaggregate model propensity sums (we tested statistical significance of six additional coefficients for count component: two for facility-specific propensity from NB, two for facility specific crash record level propensity from OP incorporated within the facility level NB, two for facility specific crash record level propensity from OP incorporated directly within the zonal NB, and six additional coefficients for severity proportion component: two for facility-specific propensity from OPFS, two for facility specific crash record level propensity from OP incorporated within the facility level OPFS, two for facility specific crash record level propensity from OP incorporated directly within the zonal OPFS) along with macro level factors as shown in equations 12 and 16.

Integrated approach 2: The micro level parameters in the segment propensity and intersection propensity are allowed to vary along with macro level factors and scalars for propensity sums within the integrated model equations as shown in equations 13 and 17.

Integrated approach 3: The disaggregate level parameters in the segment OP propensity and intersection OP propensity are allowed to vary along with macro level factors and scalars for propensity sums within the integrated model equations as shown in equations 14 and 18.

Integrated approach 4: The micro level and disaggregate level parameters are allowed to vary along with macro level factors and scalars for propensity sums within the integrated model equations as shown in equations 15 and 19.

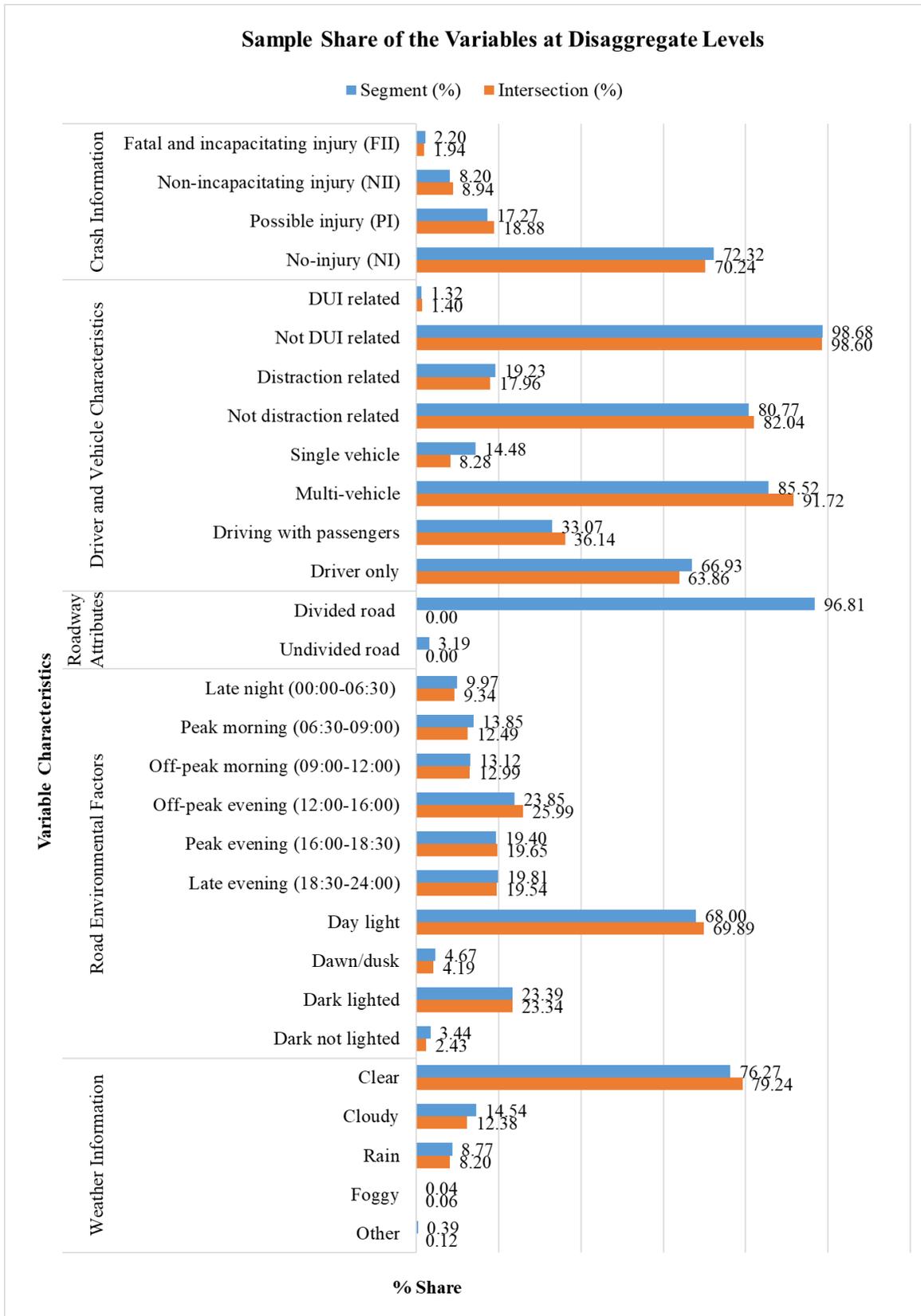


Figure 2: Sample Share of the Variables at Disaggregate Levels

Table 2: Summary Statistics of the Variables at Micro and Macro Levels

Variables	Min.	Max.	Mean	Std. Dev.
<i>Segment Facility Level (Micro)</i>				
Total segment crash count	0.000	242.000	3.118	9.530
Proportion of fatal and incapacitating injury crashes	0.000	1.000	0.014	0.084
Proportion of non-incapacitating injury crashes	0.000	1.000	0.042	0.141
Proportion of possible injury crashes	0.000	1.000	0.072	0.177
Proportion of no-injury crashes	0.000	1.000	0.364	0.427
Ln (Segment length, miles)	-12.852	1.144	-2.758	1.770
Ln (AADT)	6.686	12.095	9.987	0.891
Speed limit <=40 mph	0.000	1.000	0.457	0.498
Number of lanes	1.000	5.000	2.128	0.720
Ln (Sidewalk width +1, feet)	0.000	2.708	1.789	0.491
Traffic signal density	0.000	74.048	0.862	13.620
<i>Intersection Facility Level (Micro)</i>				
Total intersection crash count	0.000	99.000	3.709	8.422
Proportion of fatal and incapacitating injury crashes	0.000	1.000	0.011	0.069
Proportion of non-incapacitating injury crashes	0.000	1.000	0.051	0.155
Proportion of possible injury crashes	0.000	1.000	0.100	0.215
Proportion of no-injury crashes	0.000	1.000	0.359	0.412
Proportion of arterial roads	0.000	1.000	0.416	0.275
Proportion of minor roads	0.000	1.000	0.312	0.257
Proportion of >=3-lane roads	0.000	1.000	0.222	0.207
Average inside shoulder width, feet	0.000	21.000	4.005	3.878
Proportion of institutional area	0.000	1.000	0.069	0.091
Proportion of recreational area	0.000	1.000	0.068	0.138
Number of finance centers	0.000	5.000	4.604	9.302
<i>TAZ Level (Macro)</i>				
Total TAZ crash count	0.000	342.000	70.630	65.789
Proportion of fatal and incapacitating injury crashes	0.000	0.500	0.022	0.040
Proportion of non-incapacitating injury crashes	0.000	0.333	0.084	0.059
Proportion of possible injury crashes	0.000	1.000	0.173	0.104
Proportion of no-injury crashes	0.000	1.000	0.671	0.190
Ln (Truck AADT +1)	0.000	11.302	8.326	1.613
Proportion of >=3-lane roads	0.000	1.000	0.231	0.274
Average inside shoulder width, feet	0.000	18.000	3.008	3.742
Proportion of divided roads	0.000	1.000	0.610	0.357
Intersection density	0.000	0.770	0.085	0.115
Traffic signal density	0.000	1.000	0.058	0.106
Number of restaurants	0.000	4.000	3.357	5.626
Proportion of residential area	0.000	0.998	0.490	0.350
Proportion of commercial area	0.000	1.000	0.242	0.274
Proportion of African American population	0.000	0.978	0.222	0.246

In the third step, we identify the best model by comparing model performance based on Bayesian Information Criterion (BIC) and corrected Akaike Information Criterion (AICc). The BIC and AICc for a given empirical model are equal to:

$$BIC = -2LL + N_p \ln(Obs.) \quad (25)$$

$$AICc = -2LL + 2N_p + \frac{2N_p(N_p+1)}{Obs.-N_p-1} \quad (26)$$

where LL is the log-likelihood value at convergence, N_p is the number of parameters and $Obs.$ is the number of observations. The model with the lower BIC and AICc is the preferred model. The corresponding LL, NP, BIC and AICc values are presented in **Table 3**.

Table 3: Comparison of the Models

Model	LL	N _P	BIC	AICc
Non-integrated (Separate OP and NB-OPFS models for observation units)	-29,969.460	64	60,297.219	60,107.505
<i>Integrated models</i>				
Approach 1	-29,464.830	62	59,276.762	59,091.399
Approach 2	-29,945.160	59	60,220.627	60,042.034
Approach 3	-29,461.320	61	59,264.144	59,081.005
Approach 4	-29,939.490	57	60,198.090	60,024.169
<i>Integrated models with unobserved heterogeneity</i>				
Approach 3 with unobserved heterogeneity	-29,401.650	66	59,172.796	58,978.867

Based on these BIC and AICc values, several observations could be drawn. First, all the integrated models provide improved data fit as evidenced by the lower BIC and AICc values in comparison to the non-integrated model system. Second, within the integrated systems, our proposed integrated approach 3 provides the lowest BIC and AICc indicating the best data fit in comparison to the proposed other integrated approaches. For this selected model, we capture unobserved heterogeneity in the model and find that integrated model accounting for unobserved heterogeneity further improves model performance.

5.2 Model Estimation Results

The results of the proposed integrated model approach 3 with unobserved heterogeneity are discussed in this section. **Table 4** presents the model estimation results for the proposed model. The reader would note that a positive (negative) sign for a variable in **Table 4** indicates that an increase in the variable is likely to result in more (less) crashes as well as exhibit a higher (lower) impact on severity. The results of the non-integrated model system are presented in **Table A1** in the Appendix.

5.3 Disaggregate Level Attributes

The threshold parameters demarcate the various severity categories and do not have any substantive interpretation.

For the disaggregate level crash severity model, with respect to the driver and vehicle attributes, driving under influence of drugs and alcohol, distracted driving, and presence of passengers in the vehicle contribute to the likelihood of higher severity across the facility types while single vehicle crashes such as roll over and run-off-road crashes are likely to result in severe crashes for segment facility (see Das et al., 2009; Marcoux et al., 2024, 2018; Paleti et al., 2010; Pervaz et al., 2023; Yasmin and Eluru, 2013 for similar results).

With regards to roadway attributes, the results show that divided roadways show lower impact on severity of segment crashes.

Table 4: Estimation Results of the Proposed Integrated Model Approach 3 with Unobserved Heterogeneity

Disaggregate Level				
Variables	Segment OP		Intersection OP	
	Est.	t-stat.	Est.	t-stat.
Threshold between NI-PI	0.507	4.661	0.722	37.327
Threshold between PI-NII	1.206	11.287	1.453	59.456
Threshold between NII-FII	2.001	17.142	2.300	66.553
DUI related	0.361	2.241	0.659	7.609
Distraction related	0.249	4.822	0.240	8.528
Single vehicle	0.435	7.149	--	--
Driving with passengers	0.361	9.006	0.265	13.205
Divided road	-0.395	-3.879	--	--
Late night	--	--	0.091	1.873
Dark lighted	0.083	2.047	0.098	3.123
Dark not lighted	0.294	2.666	0.128	1.655
Micro Level				
Variables	Segment NB		Segment OPFS	
	Est.	t-stat.	Est.	t-stat.
Constant	-0.464	-0.790	--	--
Threshold between PNI-PPI	--	--	0.493	10.476
Threshold between PPI-PNII	--	--	1.062	20.394
Threshold between PNII-PFII	--	--	1.746	25.455
Ln (AADT)	0.224	3.813	--	--
Ln (Segment length)	0.645	7.747	0.043	1.927
Speed limit <=40mph	--	--	-0.217	-2.981
Number of lanes	0.303	4.029	--	--
Ln (Sidewalk width+1)	-0.158	-2.088	--	--
Traffic signal density	0.031	5.191	--	--
<i>Standard deviation</i>	0.009	2.589	--	--
Over dispersion parameter	1.797	10.541	--	--
Variables	Intersection NB		Intersection OPFS	
	Est.	t-stat.	Est.	t-stat.
Constant	1.161	13.024	--	--
Threshold between PNI-PPI	--	--	0.405	8.303
Threshold between PPI-PNII	--	--	1.094	21.267
Threshold between PNII-PFII	--	--	1.933	32.844
Proportion of arterial roads	0.143	1.650	--	--
Proportion of minor roads	--	--	-0.107	-1.888
Proportion of >= 3lane roads	--	--	-0.179	-1.842
Average inside shoulder width	0.016	2.024	--	--
Proportion of institutional area	-0.667	-1.650	--	--
Proportion of recreational area	--	--	0.232	1.650
Number of financial centers	-0.075	-2.211	--	--
Over dispersion parameter	3.863	34.054	--	--
Macro Level				
Variables	TAZ NB		TAZ OPFS	
	Est.	t-stat	Est.	t-stat
Constant	-1.391	-5.192	--	--
Threshold between PNI-PPI	--	--	0.590	10.564
Threshold between PPI-PNII	--	--	1.271	24.968
Threshold between PNII-PFII	--	--	2.034	29.141
Ln (Truck AADT+1)	0.228	5.736	--	--

Proportion of ≥ 3 lane roads	0.511	3.660	--	--
Average inside shoulder width	--	--	-0.016	-4.113
Traffic signal density	0.706	2.647	-0.213	-1.823
Number of restaurants	--	--	-0.056	-3.902
Proportion of commercial area	0.311	3.224	--	--
Proportion of African American population	0.141	1.655	--	--
Over dispersion parameter	0.798	5.692	--	--
Parameters for Propensity Sum				
	Count Component		Fraction Component	
Segment crash record level (within micro) ($\rho_{c_{ds}}$)	0.536	13.244	--	--
Segment facility level (ρ_{c_s})	0.251	7.588	--	--
Intersection crash record level (within micro) ($\rho_{c_{di}}$)	0.519	34.574	--	--
Intersection facility level (ρ_{c_i})	0.584	9.263	--	--
Segment crash record level (within macro) ($\rho_{f_{dsz}}$)	--	--	0.047	2.743
Unobserved Heterogeneity (Spatial Dependency)				
Correlation between segments in a zone ($\theta_{z(s)}$)	0.344	4.665	0.344	4.665
Correlation between intersections in a zone ($\theta_{z(i)}$)	0.404	8.813	0.404	8.813
Correlation between crash records at a segment (θ_s)	0.178	5.731	0.178	5.731
Correlation between crash records at an intersection (θ_i)	0.126	7.256	0.126	7.256
Log-likelihood: -29,401.650; Number of parameters: 66; BIC: 59,172.796; AICc: 58,978.867				

Note: "--" indicates variables are not statistically significant at 90% confidence level; NI = no-injury crashes, PI = possible injury crashes, NII = non-incapacitating injury crashes, and FII = fatal and incapacitating injury crashes; PNI = proportion of no-injury crashes, PPI = proportion of possible injury crashes, PNII = proportion of non-incapacitating injury crashes, and PFII = proportion of fatal and incapacitating injury crashes.

Among road, environment and weather attributes considered, the time of the day variables show that late night increases the likelihood of severe intersection crashes. The lower traffic volume and higher operating speeds during the period are likely to contribute to this higher severity (Pervaz et al., 2023). The results also show that compared to daylight and dawn/dusk conditions, dark conditions irrespective of light have a positive impact on severity across the facility types. This is because dark conditions often reduce visibility and increase reaction time on the roads (see Marcoux et al., 2018; Pervaz et al., 2023; Wang and Kim, 2019 for similar findings). The results highlight the role of lighting in improving safety and solutions such as improvement of intersection lighting, advanced warning signages for low visibility areas, educating drivers on the importance of using headlights, and staying alert could be effective countermeasures to reduce crashes during dark conditions across the facility types (FHWA, 2009). Interestingly, no weather variables are found to be significant in our selected integrated model.

5.4 Micro Level Attributes

5.4.1 Segment facility level

In the segment facility level count component, the results show that the parameter associated with AADT is found to be positively associated with segment crash frequency. The results indicate that the segments with higher AADT have higher likelihood of crashes (as found in Alarifi et al., 2017; Cai et al., 2019; Pervaz et al., 2022; Wang et al., 2020). In addition, the parameter associated with the segment length has a positive impact on crash frequency. The result is plausible as the longer road segments typically indicate a higher exposure to traffic (Alhomaidat et al., 2020; Yu et al., 2019; Zeng et al., 2020). The parameter for the number of lanes also shows positive association

with crash frequency. This is intuitive as the roads with higher number of lanes usually have higher traffic volume, higher lane changing rates and conflict risk resulting in higher number of crashes. Improving roadway markings and delineations, providing medians in multi-lane roadways, redistributing roadway space through road diet concept, driver education on lane changing risks and promoting safe driving are some proven effective safety countermeasures in this context (FHWA, 2022). The parameter associated with average sidewalk width shows negative effect on segment level crash count (as found in Bhowmik et al., 2019). Further, an increase in traffic signal per unit road length increases the likelihood of crashes. This is intuitive as a higher number of traffic signals may lead to an increase in certain types of crashes (such as rear-end crashes) in dilemma zones (Abdel-Aty and Wang, 2006; Lee et al., 2017; Park et al., 2020; Pervaz et al., 2022). Providing adequate signal visibility, appropriate signal timing based on real time traffic conditions, installation of advanced warning signs, enforcement to reduce red-light running, and promoting attentive driving habits could be some potential countermeasures to reduce these crashes.

In the segment level severity proportion component, the results found that the parameter associated with the segment length has a positive impact on crash severity proportion. The longer road segment with similar geometric attributes might encourage speeding of the vehicles, hence crash severity may increase. On the other hand, as expected, road segments with speed limit ≤ 40 mph show negative association with the severity of the segment crashes. Following these findings, installing appropriate speed limits for all road users, automated speed enforcement in high-speed zones, speed safety cameras, and variable speed limits could be considered as potential countermeasures against speeding crashes and crashes on high-speed roadways (FHWA, 2022).

5.4.2 *Intersection facility level*

In the intersection level count component, the results indicate that a higher proportion of arterial roads within intersection influence zone increase the crash frequency for intersection facility while average inside shoulder width decreases the crash frequency. Alternatively, higher proportion of institutional area and higher number of financial centers within the intersection influence zone decrease the intersection crash frequency.

In the intersection level severity proportion component, we found that higher proportion of minor roads and proportion of roads with more than 2 lanes decrease the likelihood of the intersection crash severity. On the other hand, intersections in recreation areas are found to have a higher likelihood of severe intersection crashes.

5.5 **Macro Level Attributes**

In the macro level count component, the parameter associated with the truck AADT has a positive impact on zonal crash frequency. An increased presence of trucks in the zone can affect traffic flow, visibility and increase speed variance leading to higher number of crashes (Pervaz et al., 2022). Further, a higher proportion of roads with more than 2 lanes and higher intersection density increases the likelihood of zonal crash frequency (as found in Pervaz et al., 2023).

With regards to the land-use and sociodemographic attributes, proportion of commercial areas and proportion of African American population show positive impacts on zonal crash count. These findings are intuitive as commercial areas have commerce related activities such as loading/unloading, movement of heavy vehicles and increased traffic conflicts that might contribute to higher crash risk (Cui and Xie, 2021; Mohammadnazar et al., 2021; Pervaz et al., 2023). Providing adequate turning areas for commercial vehicles, separate lanes for heavy

vehicles, informative display boards and warning signages for other road users, and safe and controlled loading/unloading activities could increase safety in these areas. For the proportion of African American population, the result might be a potential manifestation of inadequate facilities in low-income and minority neighborhoods in the region (Pervaz et al., 2022).

In the severity proportion component, wider inside shoulder width parameter indicates a negative impact on zonal crash severity. This is intuitive as wider shoulders provide additional safety margin on the road, thus, contributing to reduced severity (see Chen et al., 2017). Further, higher traffic signal density is associated with lower severity (as found in Bhowmik et al., 2021; Pervaz et al., 2023). Among the built environment attributes, an increased presence of restaurants in a zone reduces the severity of the crashes. Similar findings are reported in previous studies (Pervaz et al., 2023; Yasmin and Eluru, 2018).

5.6 Parameter for Propensity Sums

As mentioned earlier, our integrated framework tested the significant effect of micro level (segment and intersection facility) and disaggregate level (segment crash record and intersection crash record) attributes and twelve additional scalar parameters (six coefficients for count component and six for severity proportion component) for propensity sums (composite score values) along with the macro level attributes to estimate macro level crash frequency by severity. For the proposed integrated model approach 3 with unobserved heterogeneity, the coefficients for the propensity sum variables from the micro level and disaggregate level models in the count component and severity component are presented in the lower row panel of **Table 4**. The positive sign of the parameter for facility types in count and severity proportion components indicates that a higher value of micro level and disaggregate level model propensity is likely to increase the number of crashes and the crash severity in the macro level.

Table 4 shows that segment crash record level propensity sum (within segment facility), segment facility level propensity sum, intersection crash record level propensity sum (within intersection facility), and intersection facility level propensity sum show statistically significant effect with positive linkage for the count component while segment crash record level propensity sum (within macro) shows significant positive effect in the severity proportion component. The results clearly highlight that the higher propensities for crash frequency from segment and intersection facility types and the higher propensities from severity at the crash record level for the facilities are significantly associated with the increase of the crash frequency by severity at the zonal level.

5.7 Unobserved Heterogeneity

The proposed model system can capture unobserved heterogeneity in the form of spatial dependency among the micro-facilities (segments or intersections) in a zone as well as among the crash records within a micro-facility, correlations between count and severity components and random parameter effects of the variables across the observational resolutions. The spatial dependencies indicate that segments or intersections in the same zone and all the crash records at a segment or intersection are spatially correlated, and ignoring these correlations might lead to inaccurate estimates. The unobserved heterogeneity variables presented in **Table 4** correspond to these common spatial correlations. The significant effects of these correlation parameters ($\theta_{z(s)}$, $\theta_{z(i)}$, θ_s , θ_i) clearly highlight the presence of common unobserved factors across segments and intersections in the same zone and all the crash records in the same facility type. These common spatial correlations also highlight the unobserved interconnectedness in the hierarchical

resolutions considered in this study. Further, we attempted to capture the unobserved correlation (η_z) between total crashes and crash proportions by severity levels, and random parameter effects (ζ and ρ) in our proposed model system. We found that the effect of the traffic signal density variable in the segment facility level is not same across the segments as highlighted by the significant standard deviation parameter in **Table 4**. However, no statistically significant correlation between crash count and severity proportion component was recovered in our dataset.

5.8 Predictive Performance of the Model

We conduct a comparison exercise between the proposed integrated model approach 3 and independent macro level model (NB-OPFS) by testing model performance on estimation and holdout samples. The exercise involves comparing the performance of the models by employing four statistical predictive measures including mean percentage bias (MPB), mean absolute deviation (MAD), mean squared prediction error (MSPE) and Root Mean Square Error (RMSE) (please see Bhowmik et al., 2018; Pervaz et al., 2023 for a detailed definition of these measures). The model with the values of MPB, MAD, MSPE, and RMSE closer to zero provides better predictions for the observed data. The results are presented in **Table 5**.

Table 5: Predictive Performance of the Models

Dataset	Models	Measures	NI	PI	NII	FII	Total
Estimation (N=270 TAZs)	Integrated approach 3 with unobserved heterogeneity	MPB	<u>1.28</u>	<u>0.06</u>	<u>0.18</u>	<u>0.27</u>	<u>1.79</u>
	Non-integrated NB-OPFS model		2.36	0.53	0.46	0.36	3.71
	Integrated approach 3 with unobserved heterogeneity	MAD	<u>14.98</u>	<u>4.49</u>	<u>2.53</u>	<u>1.18</u>	<u>20.01</u>
	Non-integrated NB-OPFS model		21.33	6.55	3.60	1.33	30.28
	Integrated approach 3 with unobserved heterogeneity	MSPE	<u>503.70</u>	<u>43.61</u>	<u>14.03</u>	<u>2.99</u>	<u>907.48</u>
	Non-integrated NB-OPFS model		1,160.50	106.04	29.09	4.26	2,379.53
	Integrated approach 3 with unobserved heterogeneity	RMSE	<u>22.44</u>	<u>6.6</u>	<u>3.75</u>	<u>1.73</u>	<u>30.12</u>
	Non-integrated NB-OPFS model		34.07	10.3	5.39	2.06	48.78
Validation (N=30 TAZs)	Integrated approach 3 with unobserved heterogeneity	MPB	<u>-4.78</u>	<u>-1.21</u>	<u>-1.86</u>	0.06	<u>-7.80</u>
	Non-integrated NB-OPFS model		-6.66	-1.62	-2.02	<u>0.01</u>	-10.29
	Integrated approach 3 with unobserved heterogeneity	MAD	<u>27.57</u>	<u>4.75</u>	<u>4.01</u>	<u>1.42</u>	<u>35.06</u>
	Non-integrated NB-OPFS model		39.77	7.61	6.08	1.43	52.02
	Integrated approach 3 with unobserved heterogeneity	MSPE	<u>2,172.68</u>	<u>57.07</u>	<u>35.66</u>	4.37	<u>3,305.37</u>
	Non-integrated NB-OPFS model		3,781.79	130.84	59.15	<u>3.71</u>	6,222.42
	Integrated approach 3 with unobserved heterogeneity	RMSE	<u>46.61</u>	<u>7.55</u>	<u>5.97</u>	2.09	<u>57.49</u>
	Non-integrated NB-OPFS model		61.50	11.44	7.69	<u>1.93</u>	78.88

Note: NI = no-injury crashes, PI = possible injury crashes, NII = non-incapacitating injury crashes, and FII = fatal and incapacitating injury crashes.

The results clearly highlight that the proposed integrated model performs better than traditional independent NB-OPFS model across almost all fit measures computed for both estimation and validation datasets (3 exceptions, see underlined values).

6. ELASTICITY EFFECT ANALYSIS

The results presented in **Table 4** represent a joint interaction of disaggregate, micro and macro level variables and do not directly provide the actual magnitude of the effects of the variables on crash counts. To quantify the actual effects across three dimensions, we compute aggregate level elasticity effects. This study follows the elasticity effects estimation procedure demonstrated in Eluru and Bhat (2007) (see Eluru and Bhat, 2007 for a discussion on the methodology for computing elasticities). Following this procedure, the percentage change in the expected zonal crash counts by severity caused by the change in the disaggregate, micro and macro level exogenous variables are computed. For continuous variables, we obtain these changes in response to the increase of the explanatory variables by 10%. For indicator variables, we obtain the changes by changing the value of the variable to one for the subsample of observations for which the variable takes a value of zero and to zero for the subsample of observations for which the variable takes a value of one. The computed elasticities are presented in **Table 6**.

Table 6: Elasticity Effects of the Variables

Parameters	%Total	%NI	%PI	%NII	%FII
<i>Segment Record Level (Disaggregate)</i>					
DUI related	3.80	2.97	5.24	6.42	7.88
Distraction related	2.73	2.14	3.77	4.60	5.63
Single vehicle	3.10	3.71	6.54	8.00	9.81
Driving with passengers	3.93	3.08	5.42	6.62	8.09
Divided road	-4.45	-3.48	-6.15	-7.53	-9.25
Dark lighted	0.91	0.71	1.26	1.53	1.87
Dark not lighted	3.15	2.47	4.35	5.32	6.52
<i>Intersection Record Level (Disaggregate)</i>					
DUI related	17.44	17.45	17.41	17.40	17.38
Distraction related	5.86	5.87	5.85	5.84	5.83
Driving with passengers	6.36	6.37	6.35	6.35	6.34
Late night	2.21	2.22	2.21	2.21	2.21
Dark lighted	2.33	2.33	2.33	2.33	2.32
Dark not lighted	3.21	3.21	3.20	3.20	3.19
<i>Segment Facility Level (Micro)</i>					
Ln (AADT)	0.51	0.51	0.51	0.51	0.51
Ln (Segment length)	1.48	1.48	1.48	1.48	1.48
Number of lanes	1.85	1.85	1.84	1.84	1.83
Ln (Sidewalk width+1)	-0.36	-0.36	-0.36	-0.36	-0.36
Traffic signal density	0.25	0.25	0.25	0.25	0.25
<i>Intersection Facility Level (Micro)</i>					
Proportion of arterial roads	0.71	0.71	0.71	0.71	0.71
Average inside shoulder width	0.42	0.42	0.42	0.42	0.42
Proportion of institutional area	-0.25	-0.25	-0.25	-0.25	-0.25
Number of financial centers	-0.01	-0.01	-0.01	-0.01	-0.01
<i>Zonal Level (Macro)</i>					
Ln (Truck AADT+1)	2.25	2.25	2.25	2.25	2.25
Proportion of >= 3lane roads	1.79	1.79	1.77	1.77	1.75
Average inside shoulder width	0.00	0.33	-0.60	-1.02	-1.51
Traffic signal density	0.50	0.58	0.35	0.24	0.11
Number of restaurants	0.00	0.22	-0.41	-0.66	-0.93
Proportion of commercial area	0.79	0.80	0.78	0.77	0.76
Proportion of African American population	0.39	0.39	0.39	0.39	0.39

In this study, we consider a subset of exogenous variables to estimate the elasticity effects. The table shows the percentage change in the number of crashes by different severities due to the changes in the disaggregate, micro and macro level exogenous variable of interest. For example, the elasticity estimate for the DUI related variable at segments and intersections indicates that driving under influence increases total crashes by 3.80% and 17.44% respectively. The elasticity estimates for the variable associated with segment facility level AADT indicates that the expected mean total crashes will increase by 0.51% for a 10% increase in the AADT of the segment. The effects of all the variables presented in **Table 6** can be interpreted in a similar fashion. By analyzing these effects, several observations can be drawn. First, all the significant variables across the analysis resolutions are notably affecting the zonal level crash counts and severities. Second, there are differences in the elasticity effects across the expected number of crashes for different severities. Third, the most significant variables at the disaggregate level with respect to an increase in the expected number of zonal total crashes are driving with passengers, DUI related, dark not lighted, single vehicle and distraction related at the segment level and DUI related, driving with passengers, distraction related, dark not lighted, dark lighted and late night at the intersection level. Fourth, the most significant variables at micro level with respect to an increase in the expected number of fatal and incapacitating crashes are number of lanes, segment length, AADT, and traffic signal density at the segment facility level and proportion of arterial roads and average inside shoulder width at the intersection level. Finally, zonal truck AADT, proportion of 3-lane roads in the zone, proportion of commercial area and proportion of African American population in the zone are notably contributing to an increase in total, fatal and incapacitating crashes at the zonal level.

7. CONCLUSIONS

Transportation safety modelling techniques currently do not allow us to estimate the impact of observed and unobserved variables from different observation resolutions simultaneously within the same system. The integrated model systems overcome this limitation by capturing variable impacts from different analysis resolutions simultaneously within a unified framework while also accounting for unobserved heterogeneity. These integrated systems augment the macro level crash frequency estimation with the rich information flow from either micro level or disaggregate level. The current study proposes a unified framework that allows for the information flow of observed and unobserved variables from the micro level and disaggregate level into the macro level crash frequency by severity estimation. The approach involves summing up the crash propensities from micro level crash frequency computation and disaggregate level crash severity computation and incorporates the composite summed scores as independent variables within the macro level model estimation.

For independent model system, the NB-OPFS framework can be employed at micro (for segment and intersection facility) and macro level (TAZ) to jointly estimate crash frequency by severity where the NB component models total crashes and the OPFS component determines the proportion of crashes by severity class for each spatial analysis unit. At disaggregate level, the crash severity variable can be examined using an OP model for each facility type (segment and intersection). The integrated system incorporates these building blocks from various observational resolutions into a unified framework with four model structures based on how the lower-level parameters are estimated. The proposed model system is estimated using data drawn from the City of Orlando, Florida for the year 2019. The study obtained 21,189 crash records (5,669 segment and 15,520 intersection crashes) for the disaggregate level model analysis. The records contain

crash-specific factors, driver and vehicle factors, roadway attributes, road environmental and weather information of each crash record. For macro and micro level analysis, the study aggregated the crash records over 300 traffic analysis zones (TAZs), and 1,818 segments and 4,184 intersections, respectively. An exhaustive set of independent variables including roadway and traffic factors, land-use and built environment attributes, and sociodemographic characteristics are considered in these levels.

A series of models was estimated for the empirical analysis of the proposed framework, including ordered probit model for facility-specific disaggregate level severity analysis, NB-OPFS model for micro level and macro level crash frequency by severity analysis, and four integrated models to jointly estimate crash frequency by severity. We compare the model systems and select the best model using Bayesian Information Criterion (BIC) and corrected Akaike Information Criterion (AICc). Based on the BIC and AICc values, we found that all the integrated systems provide improved data fit as evidenced by the lower BIC and AICc values in comparison to the non-integrated model system. Further, within the integrated systems, our proposed integrated approach 3 provides the lowest BIC and AICc indicating the best data fit in comparison to other integrated approaches. Finally, we captured unobserved heterogeneity in the form of spatial dependency of the segments and intersections in the same zone as well as all the crash records at a segment or intersection in the integrated approach 3 and find improved model performance. We also compared the performance of the proposed integrated model with the non-integrated model system by using several predictive performance measures using estimation and holdout samples. The measures also clearly highlighted the superiority of our proposed integrated model over the non-integrated model system.

This study is not without limitations. The proposed integrated approach requires substantial data compilation and processing efforts for different observation resolutions as well as coding resources for model estimation. In addition, the model framework requires systematic analysis, i.e., the approach should start from disaggregate level analysis and involves step-by-step integration with micro level analysis followed by integration with zonal level analysis. The current study considered crash data for one year from Orlando city, Florida for the empirical analysis. It would be useful to consider data from multiple cities and/or states for multiple years while also accounting for potential spatial and temporal heterogeneity of the parameter estimates within the proposed integrated framework in future research efforts. It would also be useful to exhaustively test the impact of incorporating additional unobserved heterogeneity across all integrated approaches.

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the Signal Four Analytics (S4A), Florida Department of Transportation (FDOT) and other data sources for providing access to Florida crash and geospatial data.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: Shahrior Pervaz, Naveen Eluru, Tanmoy Bhowmik; data collection: Shahrior Pervaz; model estimation and validation: Shahrior Pervaz, Naveen Eluru, Tanmoy Bhowmik; analysis and interpretation of results: Shahrior Pervaz, Naveen Eluru, Tanmoy Bhowmik; draft manuscript preparation: Shahrior Pervaz, Naveen Eluru, Tanmoy Bhowmik. All authors reviewed the results and approved the final version of the manuscript.

8. REFERENCES

- Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research* 34(5), 597–603.
- Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: Analyzing spatial effect and identifying significant factors. *Transportation Research Record* 1953, 98–111.
- Afghari, A.P., Haque, M.M., Washington, S., 2020. Applying a joint model of crash count and crash severity to identify road segments with high risk of fatal and serious injury crashes. *Accident Analysis and Prevention* 144, 105615.
- Ahmad, N., Gayah, V. V, Donnell, E.T., 2023. Copula-based bivariate count data regression models for simultaneous estimation of crash counts based on severity and number of vehicles. *Accident Analysis and Prevention* 181, 106928.
- Alarif, S.A., Abdel-Aty, M., Lee, J., 2018a. A Bayesian multivariate hierarchical spatial joint model for predicting crash counts by crash type at intersections and segments along corridors. *Accident Analysis and Prevention* 119, 263–273.
- Alarif, S.A., Abdel-Aty, M.A., Lee, J., Park, J., 2017. Crash modeling for intersections and segments along corridors: A Bayesian multilevel joint model with random parameters. *Analytic Methods in Accident Research* 16, 48–59.
- Alarif, S.A., Abdel-Aty, M.A., Lee, J., Wang, X., 2018b. Exploring the Effect of Different Neighboring Structures on Spatial Hierarchical Joint Crash Frequency Models. *Transportation Research Record* 2672, 210–222.
- Alhomaiddat, F., Kwigizile, V., Oh, J.-S., Houten, R. Van, 2020. How does an increased freeway speed limit influence the frequency of crashes on adjacent roads? *Accident Analysis and Prevention* 136, 105433.
- Aptech, 2015. Aptech Systems Inc.
- Bhat, C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B* 35(7), 677–693.
- Bhowmik, T., Yasmin, S., Eluru, N., 2021. A New Econometric Approach for Modeling Several Count Variables: A Case Study of Crash Frequency Analysis by Crash Type and Severity. *Transportation Research Part B* 153, 172–203.
- 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.
- Bhowmik, T., Yasmin, S., Eluru, N., 2018. A joint econometric approach for modeling crash counts by collision type. *Analytic Methods in Accident Research* 19, 16–32.
- Briz-Redón, Á., Martínez-Ruiz, F., Montes, F., 2019. Investigation of the consequences of the modifiable areal unit problem in macroscopic traffic safety analysis: A case study accounting for scale and zoning. *Accident Analysis and Prevention* 132, 105276.

- Cai, Q., Abdel-Aty, M., Lee, J., 2017. Macro-level vulnerable road users crash analysis: A Bayesian joint modeling approach of frequency and proportion. *Accident Analysis and Prevention* 107, 11–19.
- Cai, Q., Abdel-Aty, M., Lee, J., Huang, H., 2019. Integrating macro- and micro-level safety analyses: a Bayesian approach incorporating spatial interaction. *Transportmetrica A* 15(2), 285–306.
- Chen, S., Saeed, T.U., Labi, S., 2017. Impact of road-surface condition on rural highway safety: A multivariate random parameters negative binomial approach. *Analytic Methods in Accident Research* 16, 75–89.
- Cheng, W., Gill, G.S., Ensich, J.L., Kwong, J., Jia, X., 2018. Multimodal crash frequency modeling: Multivariate space-time models with alternate spatiotemporal interactions. *Accident Analysis and Prevention* 113, 159–170.
- Cui, H., Xie, K., 2021. An accelerated hierarchical Bayesian crash frequency model with accommodation of spatiotemporal interactions. *Accident Analysis and Prevention* 153, 106018.
- Das, A., Abdel-Aty, M., Pande, A., 2009. Using conditional inference forests to identify the factors affecting crash severity on arterial corridors. *Journal of Safety Research* 40(4), 317–327.
- Ding, H., Lu, Y., Sze, N.N., Antoniou, C., Guo, Y., 2023. A crash feature-based allocation method for boundary crash problem in spatial analysis of bicycle crashes. *Analytic Methods in Accident Research* 37, 100251.
- Eluru, N., Bhat, C.R., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis and Prevention* 39, 1037–1049.
- Federal Highway Administration (FHWA), 2022. Proven Safety Countermeasures. Available at <https://highways.dot.gov/safety/proven-safety-countermeasures>, Accessed on April 25, 2025.
- Federal Highway Administration (FHWA), 2009. Reducing Late-Night/Early-Morning Intersection Crashes By Providing Lighting (FHWA-SA-09-017). Available at https://highways.dot.gov/sites/fhwa.dot.gov/files/dot_49237_DS1.pdf, Accessed on April 25, 2025.
- Haddad, A., Mondal, A., Eluru, N., Bhat, C.R., 2024. A novel integrated approach to modeling and predicting crash frequency by crash event state. *Analytic Methods in Accident Research* 41, 100319.
- Han, C., Huang, H., Lee, J., Wang, J., 2018. Investigating varying effect of road-level factors on crash frequency across regions: A Bayesian hierarchical random parameter modeling approach. *Analytic Methods in Accident Research* 20, 81–91.
- Hosseinpour, M., Sahebi, S., Zamzuri, Z.H., Yahaya, A.S., Ismail, N., 2018. Predicting crash frequency for multi-vehicle collision types using multivariate Poisson-lognormal spatial model: A comparative analysis. *Accident Analysis and Prevention* 118, 277–288.
- Huang, H., Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis and Prevention* 42(6), 1556–1565.

- Huang, H., Zhou, H., Wang, J., Chang, F., Ma, M., 2017. A multivariate spatial model of crash frequency by transportation modes for urban intersections. *Analytic Methods in Accident Research* 14, 10–21.
- Ingle, A., Gates, T.J., 2021. Safety performance of rural offset-T intersections. *Transportation Research Record*, 2675(2), 40–52.
- Islam, S.M., Washington, S., Kim, J., Haque, M.M., 2023. A hierarchical multinomial logit model to examine the effects of signal strategies on right-turn crash injury severity at signalised intersections. *Accident Analysis and Prevention* 188, 107091.
- Jahan, M.I., Bhowmik, T., Eluru, N., 2024. Enhanced aggregate framework to model crash frequency by accommodating zero crashes by crash type. *Transportation Research Record*, 2678(2), 506–519.
- Kim, M., Kho, S.-Y., Kim, D.-K., 2017. Hierarchical ordered model for injury severity of pedestrian crashes in South Korea. *Journal of Safety Research* 61, 33–40.
- Lee, J., Abdel-Aty, M., Cai, Q., 2017. Intersection crash prediction modeling with macro-level data from various geographic units. *Accident Analysis and Prevention* 102, 213–226.
- Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A* 44(5), 291–305.
- Marcoux, R., Pervaz, S., Eluru, N., 2024. Assessing non-motorist safety in motor vehicle crashes – a copula-based approach to jointly estimate crash location type and injury severity. *Analytic Methods in Accident Research* 42, 100322.
- Marcoux, R., Yasmin, S., Eluru, N., Rahman, M., 2018. Evaluating temporal variability of exogenous variable impacts over 25 years: An application of scaled generalized ordered logit model for driver injury severity. *Analytic Methods in Accident Research* 20, 15–29.
- Mohammadnazar, A., Mahdinia, I., Ahmad, N., Khattak, A.J., Liu, J., 2021. Understanding how relationships between crash frequency and correlates vary for multilane rural highways: Estimating geographically and temporally weighted regression models. *Accident Analysis and Prevention* 157, 106146.
- Paleti, R., Eluru, N., Bhat, C.R., 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. *Accident Analysis and Prevention* 42(6), 1839–1854.
- 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 and Prevention* 106, 305–314.
- Park, H.-C., Park, B.-J., Park, P.Y., 2022. A multiple membership multilevel negative binomial model for intersection crash analysis. *Analytic Methods in Accident Research* 35, 100228.
- Park, H.-C., Yang, S., Park, P.Y., Kim, D.-K., 2020. Multiple membership multilevel model to estimate intersection crashes. *Accident Analysis and Prevention* 144, 105589.
- Pervaz, S., Bhowmik, T., Eluru, N., 2024. An integrated multi-resolution framework for jointly estimating crash type and crash severity. *Analytic Methods in Accident Research* 42, 100321.

- Pervaz, S., Bhowmik, T., Eluru, N., 2023. An econometric framework for integrating aggregate and disaggregate level crash analysis. *Analytic Methods in Accident Research* 39, 100280.
- Pervaz, S., Bhowmik, T., Eluru, N., 2022. Integrating macro and micro level crash frequency models considering spatial heterogeneity and random effects. *Analytic Methods in Accident Research* 36, 100238.
- Pew, T., Warr, R.L., Schultz, G.G., Heaton, M., 2020. Justification for considering zero-inflated models in crash frequency analysis. *Transportation Research Interdisciplinary Perspectives* 8, 100249.
- Sharafeldin, M., Farid, A., Ksaibati, K., 2022. Examining the risk factors of rear-end crashes at signalized intersections. *Journal of Transportation Technologies* 12(4), 635–650.
- Wang, C., Chen, F., Cheng, J., Bo, W., Zhang, P., Hou, M., Xiao, F., 2020. Random-Parameter Multivariate Negative Binomial Regression for Modeling Impacts of Contributing Factors on the Crash Frequency by Crash Types. *Discrete Dynamics in Nature and Society*, 2020(1), 1–13.
- Wang, J., Huang, H., 2016. Road network safety evaluation using Bayesian hierarchical joint model. *Accident Analysis and Prevention* 90, 152–158.
- Wang, K., Bhowmik, T., Zhao, S., Eluru, N., Jackson, E., 2021. Highway safety assessment and improvement through crash prediction by injury severity and vehicle damage using Multivariate Poisson-Lognormal model and Joint Negative Binomial-Generalized Ordered Probit Fractional Split model. *Journal of Safety Research* 76, 44–55.
- Wang, X., Kim, S.H., 2019. Prediction and Factor Identification for Crash Severity: Comparison of Discrete Choice and Tree-Based Models. *Transportation Research Record* 2673, 640–653.
- Xie, K., Ozbay, K., Yang, H., 2019. A multivariate spatial approach to model crash counts by injury severity. *Accident Analysis and Prevention* 122, 189–198.
- 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., Huang, H., Dong, N., 2018. The modifiable areal unit problem in traffic safety: Basic issue, potential solutions and future research. *Journal of Traffic and Transportation Engineering (English edition)* 5, 73–82.
- Xu, P., Huang, H., Dong, N., Abdel-Aty, M., 2014. Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem. *Accident Analysis and Prevention* 70, 110–120.
- Yasmin, S., Eluru, N., 2018. A joint econometric framework for modeling crash counts by severity. *Transportmetrica A* 14(3), 230–255.
- Yasmin, S., Eluru, N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Accident Analysis and Prevention* 59, 506–521.
- Yasmin, S., Momtaz, S.U., Nashad, T., Eluru, N., 2018. A Multivariate Copula-Based Macro-Level Crash Count Model. *Transportation Research Record*, 2672(30), 64–75.

- Yu, R., Wang, Y., Quddus, M., Li, J., 2019. A marginalized random effects hurdle negative binomial model for analyzing refined-scale crash frequency data. *Analytic Methods in Accident Research* 22, 100092.
- Yue, H., 2024. Investigating the influence of streetscape environmental characteristics on pedestrian crashes at intersections using street view images and explainable machine learning. *Accident Analysis and Prevention* 205, 107693.
- Zeng, Q., Wen, H., Huang, H., Wang, J., Lee, J., 2020. Analysis of crash frequency using a Bayesian underreporting count model with spatial correlation. *Physica A* 545, 123754.
- Zhai, X., Sze, N.N., Lee, J.J., Xu, P., Huang, H., 2025. Multi-scale approaches to cope with scale effect issues in macroscopic crash analysis. *Accident Analysis and Prevention* 214, 107971.
- Zhu, M., Sze, N.N., Li, H., 2024. Influence of walking accessibility for metro system on pedestrian safety: A multiple membership multilevel model. *Analytic Methods in Accident Research* 43, 100337.

APPENDIX

Table A1: Estimation Results of the Non-integrated Model System

Disaggregate Level				
Variables	Segment OP		Intersection OP	
	Est.	t-stat	Est.	t-stat
Threshold between NI-PI	0.713	7.114	0.734	44.528
Threshold between PI-NII	1.402	13.708	1.460	75.057
Threshold between NII-FII	2.185	20.253	2.299	80.107
DUI related	0.360	2.111	0.667	7.401
Distraction related	0.271	5.952	0.260	9.763
Single vehicle	0.454	8.412	--	--
Driving with passengers	0.365	9.210	0.277	12.671
Divided road	-0.185	-1.879	--	--
Late night	--	--	0.090	2.104
Dark lighted	0.087	1.853	0.103	3.561
Dark not lighted	0.320	3.113	0.150	2.153
Foggy weather	--	--	0.757	1.819
Log-likelihood	-3,724.584		-11,585.204	
Number of parameters	10		10	
Micro Level				
Variables	Segment NB		Segment OPFS	
	Est.	t-stat	Est.	t-stat
Constant	-0.464	-0.790	--	--
Threshold between PNI-PPI	--	--	0.493	10.476
Threshold between PPI-PNII	--	--	1.062	20.394
Threshold between PNII-PFII	--	--	1.746	25.455
Ln (Segment length)	0.645	7.747	0.043	1.927
Ln (AADT)	0.224	3.813	--	--
Speed limit <=40mph	--	--	-0.217	-2.981
Number of lanes	0.303	4.029	--	--
Ln (Sidewalk width+1)	-0.158	-2.088	--	--
Traffic signal density	0.031	5.191	--	--
Over dispersion parameter	1.797	10.541	--	--
Log-likelihood	-2,863.301		-642.898	
Number of parameters	7		5	
Variables	Intersection NB		Intersection OPFS	
	Est.	t-stat	Est.	t-stat
Constant	1.161	13.024	--	--
Threshold between PNI-PPI	--	--	0.405	8.303
Threshold between PPI-PNII	--	--	1.094	21.267
Threshold between PNII-PFII	--	--	1.933	32.844
Proportion of arterial roads	0.143	1.650	--	--
Proportion of minor roads	--	--	-0.107	-1.888
Proportion of >= 3lane roads	--	--	-0.179	-1.842
Average inside shoulder width	0.016	2.024	--	--
Proportion of institutional area	-0.667	-1.650	--	--
Proportion of recreational area	--	--	0.232	1.650
Number of financial centers	-0.075	-2.211	--	--
Over dispersion parameter	3.863	34.054	--	--
Log-likelihood	-7,917.508		-1,700.927	
Number of parameters	6		6	

Macro Level				
Variables	TAZ NB		TAZ OPFS	
	Est.	t-stat	Est.	t-stat
Constant	-2.500	-5.671	--	--
Threshold between PNI-PPI	--	--	0.917	3.977
Threshold between PPI-PNII	--	--	1.597	7.213
Threshold between PNII-PFII	--	--	2.361	10.568
Ln (Truck AADT+1)	0.638	12.609	0.050	1.902
Proportion of >= 3lane roads	0.309	1.734	--	--
Average inside shoulder width	-0.020	-1.762	-0.017	-3.499
Proportion of divided roads	0.635	3.454	--	--
Intersection density	1.283	2.659	--	--
Traffic signal density	1.060	1.914	-0.272	-2.476
Number of restaurants	--	--	-0.037	-2.535
Proportion of residential area	0.610	3.877	--	--
Proportion of commercial area	0.454	2.203	-0.100	-1.650
Proportion of African American population	0.362	2.200	0.132	1.854
Over dispersion parameter	1.053	7.453	--	--
Log-likelihood	-1,316.534		-218.404	
Number of parameters	11		9	
Total log-likelihood: -29,969.460; Total number of parameters: 64; BIC: 60,297.219; AICc: 60,107.505				

Note: Note: "--" indicates variables are not statistically significant at 90% confidence level; NI = no-injury crashes, PI = possible injury crashes, NII = non-incapacitating injury crashes, and FII = fatal and incapacitating injury crashes; PNI = proportion of no-injury crashes, PPI = proportion of possible injury crashes, PNII = proportion of non-incapacitating injury crashes, and PFII = proportion of fatal and incapacitating injury crashes.