Accident Analysis and Prevention 194 (2024) 107366

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Investigating transportation safety in disadvantaged communities by integrating crash and Environmental Justice data

A. Latif Patwary [a], Antora Mohsena Haque [a], Iman Mahdinia [b], Asad J. Khattak [a,*]

[a] *Department of Civil and Environmental Engineering, University of Tennessee Knoxville, Knoxville, TN 37996, USA*
[b] *Safe Transportation Research & Education Center, The University of California Berkeley, CA 94704, USA*

ARTICLE INFO

Keywords:
Disadvantaged communities
Justice40
Fatal crashes
Fatalities
Safety
environmental justice

ABSTRACT

Recent efforts to identify disadvantaged communities (DACs) on a census tract level have evoked possibilities of attaining transportation justice and vision zero goals in these areas. To identify DACs, the United States Department of Transportation (USDOT) has developed six comprehensive indicators: economy, environment, equity, health, resilience, and transportation access. The indicators are used to explore the associations between DACs (in 71,728 census tracts) and five years of fatal crashes, providing a comprehensive understanding of safety risks. Specifically, using data on DACs and linking it with census and crash data, this study aims to understand the complex connections between safety (captured through fatal crashes) and disadvantages that communities confront due to a convergence of multiple challenges and burdens using Zero-Hurdle Negative Binomial models. The results reveal that health, resilience, and transportation-disadvantaged tracts are associated with more fatal crashes. The study also found the presence of a higher percentage of the population with bachelor's degrees and increased use of public transportation are correlated with fewer fatal crashes. Also, a higher fatal crash rate is observed in disadvantaged census tracts where a high proportion of the Hawaiian or other Pacific Islander, and American Indian or Alaska Native populations live. This implies that targeted interventions can be explored further in tracts that show high correlations with fatal crashes. The findings contribute to traffic safety by highlighting the risks in DACs, which can help design and implement traffic safety interventions. The insights gained from this study can inform decision-making and help to guide the development of more equitable traffic safety programs in disadvantaged communities.

1. Introduction

To achieve the Vision Zero goal, the United States Department of Transportation (USDOT) has formulated National Roadway Safety Strategy (NRSS), which contains strategies to reduce injuries and fatalities from the US road network in a comprehensive manner. One of the core objectives of NRSS is to achieve safer people (USDOT 2022). Hence, USDOT needs to ensure that the transportation sector is not unfair to anyone. USDOT is taking initiatives to resolve disproportionate safety impacts that affect people of color and other minority groups who are historically disadvantaged and marginalized (USDOT 2023). Consequently, the Biden-Harris administration has been proactive about Environmental Justice (EJ), an initiative to achieve racial equity and address the climate crisis. Through the EJ initiative, 40 % of the overall benefits of federal investments are planned to be delivered to climate and clean energy, including sustainable transportation and disadvantaged communities (DACs) (White-House 2022). DACs need to be identified so that grant applicants of EJ can be assured that their projects will help DACs.

As part of these efforts, the USDOT has defined Disadvantaged Communities (DACs) as those that are affected by high poverty, low wealth, low educational attainment, insufficient local jobs, high inequality, and low homeownership. USDOT has also developed six categories of transportation disadvantages to identify census tracts that qualify as DACs, and provides a mapping tool to visualize these areas (USDOT 2023). The indicators include disadvantages in terms of economy, environment, equity, health, resilience, and transportation access. This study uses data on these unique indicators at the census tract level to investigate the relationship between fatal crashes and disadvantages that communities confront due to a convergence of multiple challenges

* Corresponding author.
E-mail addresses: apatwary@vols.utk.edu (A.L. Patwary), ahaque1@vols.utk.edu (A.M. Haque), imahdinia@berkley.edu (I. Mahdinia), akhattak@utk.edu (A.J. Khattak).

https://doi.org/10.1016/j.aap.2023.107366
Received 3 March 2023; Received in revised form 3 October 2023; Accepted 24 October 2023
Available online 2 November 2023

and burdens. By linking five years of fatal crash data with demographic information and developing crash-based count models, this study contributes to the USDOT's priority of improving safety in DACs and provides valuable insights into the complexity of fatal crashes in disadvantaged regions across the US. The use of USDOT-provided and publicly available disadvantage indicators adds to the intellectual merit and contribution of the paper, making it important for state and federal policymakers who need to allocate resources to improve disadvantaged regions.

2. Literature review

The main focus of safety improvement studies has been analyzing the relationship between crashes and various features like driver behavior (Liu and Khattak 2016, Khattak and Wali 2017, Kamrani et al. 2018, Wali et al. 2018, Patwary 2023, Gu et al. 2023, Khojastehpour et al. 2022), vehicle features (Kadilar 2016), roadway characteristics (Arvin et al. 2019, Patwary and Khattak 2022), traffic condition (Ahmed et al. 2018, Mohammadnazar et al. 2022), weather (Eisenberg and Warner 2005, Saha et al. 2016), land use (Pulugurtha et al. 2013), and many more. Specifically, the findings from previous studies suggest that greater alcohol consumption per capita and driver distraction due to mobile phones can potentially increase fatal crashes (Loeb and Clarke 2009, Baireddy et al. 2018, Schneider 2020, Patwary and Khattak 2023, Rudisill et al. 2023). Also, exposure measures such as high average daily traffic (ADT) and vehicle miles traveled (VMT) are positively associated with fatal crashes (Pei et al. 2012, Pour et al. 2012). Although these studies are helpful to many extents, socioeconomic factors and the minority status of the people can also play a critical role in understanding crash risks at the community level.

A significant portion of the existing literature has explored factors, including population, race, income, residence environment, transportation mode, and education attainment (Stamatiadis and Puccini 1999, Nantulya and Reich 2003, Traynor 2009, Lee et al. 2014, Chimba et al. 2018). For example, one of the studies found that a lower percentage of high school graduation and university attainment impact single-vehicle crashes in the Southeast region of the US (Stamatiadis and Puccini 1999). In residence characteristics of at-fault-drivers based study found that the proportion population who work from home, and commuters who commute for less than 15 min are negatively associated with the number of at-fault drivers. An increase in population is positively associated with at-fault drivers (Lee et al. 2014). Whereas studies found that high fatality and severe injuries are seen in lower-income regions in the US (Traynor 2009). Similar associations are observed in other countries. For instance, Christie et al. (2010) found in the UK that the residents of most deprived regions are 5 times more vulnerable to fatal crashes. According to the literature, this unequal distribution of traffic crashes or fatalities between low-income and higher-income communities can be attributed to various factors, including infrastructure disparities with limited access to well-maintained roads and traffic control devices, and limited emergency medical services (Traynor 2009, Azetsop 2010).

Furthermore, social and racial inequities and disproportionate vulnerability of certain groups (e.g., pedestrians and elderly populations) play a role in the differences in crashes and fatalities between different communities. Numerous studies have reported that fatalities differ among various racial and ethnic groups (Campos-Outcalt et al. 2003, Sanders and Schneider 2022). A recent study on all pedestrian fatalities in the US reported that Black and Native American pedestrians are killed more than White pedestrians (79 %, 83 %, and 72 %, respectively) in the darkness, and 65 years or older Asian pedestrians are 1.7 times more likely to be killed than the White pedestrians (Sanders and Schneider 2022). Another recent study in Texas found that census tracts with minority status are strongly correlated with overall crashes, and socioeconomic status is strongly correlated with fatal crashes (Li et al. 2022). Minority groups face racial discrimination in society, and it is important to realize whether these minority groups reside in DACs where fatal crashes occur. From 1999 to 2006, the Hispanic population was the highest among all age groups to suffer from alcohol-impaired driving deaths in the US (Roudsari et al. 2009). It was found that a 10 % point increase in the state alcohol policy of environment restrictiveness is associated with a reduction of 10 % odds of a crash being alcohol-related (Naimi et al. 2018).

There can be several reasons for the expectation of higher fatality rates in disadvantaged or underserved communities. First, these communities tend to have lower rates of vehicle ownership and higher levels of pedestrian activity (Noland et al. 2013). The increased pedestrian activity can lead to more conflicts between vehicles and pedestrians. Additionally, research by Jacobsen and Rutter (2012) indicates that bicycle crashes in lower socioeconomic status urban areas often result from interactions with vehicular traffic and vehicles operating at high speeds. Second, DACs are exposed to greater risks associated with unsafe traffic conditions. For instance, Giles-Corti and Donovan (2002) noted that individuals in low-income communities are more likely to be exposed to busy traffic environments with fewer supportive amenities for walking and biking. These neighborhoods often lack adequate pedestrian-friendly structures, e.g., sidewalks, unsafe intersections, and inadequate pedestrian signage (Noland et al. 2013), which expose vulnerable users to higher safety threats from vehicular traffic. Lower-income neighborhoods also tend to have more infrastructure disorder and unevenness than higher-income neighborhoods (Gibbs et al. 2012). Numerous studies have emphasized the role of sufficient bicycle infrastructure in reducing crash risks (Dill and Carr 2003, Buehler and Pucher 2012, Barajas 2018). However, historically, disadvantaged regions in the US have had less access to bicycling infrastructure due to lower investment in these facilities, ultimately resulting in higher crash risks (Cradock et al. 2009). Third, related to the lack of infrastructure, an increase in further conflicts, e.g., schools and drop/pick up zones, in disadvantaged communities can contribute to more crashes. For example, Yu et al. (2022) found that the presence of schools had a higher correlation with traffic crashes only in areas with high poverty rates and a predominantly people-of-color population. Fourth, disparities in roadway characteristics, e.g., narrower roads, built environment, e.g., high population density, can contribute to higher crashes. For instance, Shin (2023) identified a positive correlation between higher population density, a greater density of narrower lanes, and intersections with four or more ways, with increased bicycle crashes. Finally, cultural and language barriers can render new immigrants to be more susceptible to traffic crashes in such communities (Chen et al. 2012).

The US government allocates funds every year for safety improvements. Most of the time, those funds are disproportionately allocated, and DACs do not receive equitable funding. Therefore, identifying DACs is necessary to assist federal and local governments in allocating resources for safety improvement. A state-level study consisting of 48 adjacent states of the US was conducted to show the association between capital expenditures and highway capital stock on highway fatalities. It was found that there can be a 0.056 % decline in highway fatalities if states with lower capital stock increase their highway capital expenditure by 1 % (Nguyen-Hoang and Yeung 2014). However, a study on 50 states of the US's CO_2 emissions, fatalities, and truck transport value applied direction output distance function and found that the transfer of resources to CO_2 emission reduction is associated with an increase in fatalities (Rogers and Weber 2011).

Different count models, i.e., Poisson and negative binomial models, have widely been used to analyze vehicular crash frequency. Negative binomial models are adopted over Poisson models when the crash data shows overdispersion (i.e., variance is greater than mean). Previous studies have used zero-inflated models in case of excess zeros in the data (Lord et al. 2005). However, zero-inflated models may inaccurately estimate the coefficients since these models include both sampling and structural zeros. Researchers have argued that crashes usually result from sampling zeros, not structural zeros. Therefore, to accurately account for these aspects, Zero-Hurdle models should be adopted for crash analysis (Lord et al. 2007). Despite the potential advantages of employing Hurdle models in traffic safety research, their exploration has not been as extensive as anticipated. Zero-inflated models have often been estimated without thorough consideration, sometimes selected solely based on the assumption of an excessive number of zeros (Son et al. 2011). Nonetheless, several studies have incorporated Hurdle models in the crash investigation (Son et al. 2011, Hosseinpour et al. 2013, Hosseinpour et al. 2014, Cai et al. 2016). For instance, Hosseinpour et al. (2013) utilized Hurdle models to analyze the total number of pedestrian-vehicle crashes at the segment level on Malaysian federal roads. Their findings revealed that the Hurdle models outperformed zero-inflated models in comparative measures, such as AIC and log-likelihood at convergence values. Additionally, they identified significant associations of ADT, land use, and area type with pedestrian-involved crashes. In another study, Son et al. (2011) delved into crash occurrence by integrating individual vehicular data and crash records, and developed various count models. They concluded that zero-Hurdle models handle excessive zeros more effectively than zero-inflated models. Overall, when it comes to addressing excessive zeros in crash modeling, Hurdle models can be considered a more suitable choice compared to the conventional zero-inflated models.

This study's primary contribution lies in meticulously analyzing traffic safety in diverse, disadvantaged communities. Notably, safety is examined at the granular level of census tracts. Remarkably, in previous studies, these issues are lightly researched, i.e., the complex connections between safety (captured through fatal crashes) and disadvantages that communities confront due to a convergence of multiple challenges and burdens. The disadvantages are quantified using economic, environmental, equity, health, resilience, and transportation indicators (USDOT 2023). Harnessing the data helps us delve into an unexplored aspect of safety. Previous research has explored transportation fatalities based on certain sociodemographic and economic characteristics in selected counties, cities, and occasionally multiple states. However, a comprehensive study has yet to be undertaken for the entire US at a granular level, such as census tracts. This study has sought to fill this gap by conducting a thorough examination at the census tract level.

2.1. Data

Three types of datasets are used in this paper. The disadvantage indicator-based dataset is retrieved as a GIS shapefile from the USDOT's designated website for displaying DACs (USDOT 2023). Fatal crash information is retrieved from Fatality Analysis Reporting System (FARS) website, and socio-demographic information is collected from the US Census Bureau's website. A brief overview of the datasets is provided below:

2.2. Disadvantage indicator-based dataset

USDOT has developed an interactive GIS-based mapping tool for public use, and it is launched as a dashboard primarily for grant applicants who can confirm that their projects address DACs. The definition of DAC is based on indicators that are collected at the US Census tract level. These indicators formed six broad themes of transportation disadvantages: economy, environment, equity, health, resilience, and transportation. These six broad themes are termed disadvantage indicators throughout the paper. The short definitions of them are provided below:

2.2.1. Economic disadvantage

This theme identifies communities affected by high poverty, low wealth, low educational attainment, insufficient local jobs, high inequality, and low homeownership. The data source of this indicator includes CDC Social Vulnerability Index, Census America Community Survey, FEMA Resilience Analysis and Planning Tool.

2.2.2. Environmental disadvantage

This theme identifies communities that suffer from an unbalanced pollution burden and substandard environmental quality. The data is gathered from EPA EJ Screen.

2.2.3. Equity disadvantage

This theme identifies communities where a high percentile of the population's English speaking skill is "less than well" per the CDC Social Vulnerability Index.

2.2.4. Health disadvantage

This theme identifies communities affected by adverse health outcomes, disability, and environmental exposures. The data are collected from the CDC Social Vulnerability Index.

2.2.5. Resilience disadvantage

This theme identifies communities vulnerable to climate change hazards. The data are collected from FEMA National Risk Index.

2.2.6. Transportation disadvantage

This theme identifies communities affected by more prolonged and expensive ways of transportation. The data for this theme is collected from CDC Social Vulnerability Index, Census America Community Survey, EPA Smart Location Map, and HUD Location Affordability Index.

Each theme's relevant indicators are averaged to form an aggregated disadvantage indicator in percentiles. USDOT considers a census tract a disadvantaged community (DAC) if it surpasses the 50th percentile (75th percentile for resilience) across more than three of the six themes. The list of those indicators is provided in Table 1. For details on these indicators, refer to USDOT's website (USDOT 2023).

The Disadvantage indicator-based dataset has information for 72,843 census tracts but no tract information for the state of American Samoa, Guam, Northern Mariana, Puerto Rico, and the Virgin Islands. These states have 18, 57, 25, 945, and 32 census tracts, and any fatal crashes falling in these tracts are excluded from this study. The attribute table of the GIS shapefile of the dataset also contains names and FIPS codes of the states, counties, tracts, and tract sizes.

Table 2 shows that 47.28 % of census tracts are economically disadvantaged, 49.61 % are environmentally disadvantaged, 49.97 % are equity disadvantaged, 50.04 % are health disadvantaged, 25.19 % are resilience disadvantaged, and 48.38 % are transportation disadvantaged, and 30.43 % are overall disadvantaged. The disadvantaged tracts are shown in Figs. 1 and 2. Environmental DACs are concentrated in the southeastern and western regions of the US, while transportation DACs are distributed throughout the country.

2.3. Demographic information-based dataset

All the latest demographic information was collected at the census tract level. Tract population, gender, race, means of transport to work, median income, employment, and educational attainment were collected from 2021 census data. Any income above $250,000 is coded as $250,000 in the processed data.

2.4. Fatal crash-based dataset

FARS dataset contains crash-related information, e.g., geographic location (latitude-longitude), time of the crash, number of fatalities, etc. As the fatal crash frequency is limited compared to the rest of the crashes each year, aggregating five years of data is deemed reasonable. The fatal crash information is collected from 2017 to 2021. A total of 177,409 fatal crashes occurred in the US in these five years.

2.5. Data on other variables

Several other variables can provide valuable insights into the

Table 1
List of Disadvantage Indicators under Six themes (USDOT 2023).

Attribute Alias	Disadvantage Indicator
Average of Transportation Indicator Percentiles (calculated)	Transportation Theme
Total workers 16 or older in a census tract	Transportation Cost Burden
Percentage of non-transit households who have zero vehicles	Transportation Cost Burden
Percentile percentage households with no vehicle available estimate	Dependency on a single form of transportation (i.e. no personal vehicle)
Percentage of non-transit households who have one vehicle	Transportation Cost Burden
Percentage of non-transit households who have two vehicles	Transportation Cost Burden
Percentage of non-transit households who have 3+ vehicles	Transportation Cost Burden
Number of non-transit workers	Transportation Cost Burden
Number of transit users 16 and over	Transportation Cost Burden
Average weekday vehicle miles traveled per state	Transportation Cost Burden
Calculated Average Annual Vehicle Miles Traveled	Transportation Cost Burden
Average Annual Median Earnings	Transportation Cost Burden
Five Year average price of gas per state	Transportation Cost Burden
Five-year average gas mileage per state	Transportation Cost Burden
Calculated average number of cars per household	Transportation Cost Burden
Calculated average cost of owning a car	Transportation Cost Burden
Calculated national average annual cost of using transit	Transportation Cost Burden
Calculated average annual cost of transportation	Transportation Cost Burden
Annual Travel Time in Minutes	Transportation Cost Burden
Percentile of Mean commute time to work (in minutes)	Longer commute times
Annual Travel Time in Hours	Transportation Cost Burden
Travel Time Cost	Transportation Cost Burden
Calculated average annual cost of transportation as a % of income	Transportation Cost Burden
Percentile of Transportation Cost Burden	Transportation Cost Burden
National Walkability Index	Walkability
National Walkability Index Percentile	Walkability
Average of Health Indicator Percentiles (calculated)	Health Theme
Percentile percentage of persons aged 65 and older estimate	Age (over 65)
Adjunct variable - Percentage uninsured in the total civilian noninstitutionalized population estimate, 2014-2018 ACS	Uninsured
Percentile percentage uninsured in the total civilian noninstitutionalized population estimate, 2014-2018 ACS	Uninsured
Percentile percentage of civilian noninstitutionalized population with a disability estimate	Disability
Average of Economy Indicator Percentiles (calculated)	Economy Theme
Percentile Percentage of persons with no high school diploma (age 25+) estimate	Education
Overall Renter Rate: Percent of Occupied Housing Units that are Renter-Occupied	Rentership
Percentile Overall Renter Rate: Percent of Occupied Housing Units that are Renter-Occupied	Rentership
Percentile Percentage of civilian (age 16+) unemployed estimate	Unemployment Rate
Percentile per capita income estimate	Income
Percentile Percentage of persons below poverty estimate	Areas of Persistent Poverty
GINI Index Percentile (calculated)*	GINI
Total housing units	Housing Cost Burden
Total Owner-occupied housing units	Housing Cost Burden
Owner-occupied housing units - Less than $20,000 - 30% or more	Housing Cost Burden
Owner-occupied housing units - $20,000 to $34,999 - 30% or more	Housing Cost Burden
Owner-occupied housing units - $35,000 to $49,999 - 30% or more	Housing Cost Burden
Owner-occupied housing units - $50,000 to $74,999 - 30% or more	Housing Cost Burden
Owner-occupied housing units - $75,000 or more - 30% or more	Housing Cost Burden
Renter-occupied housing units	Housing Cost Burden
Renter-occupied housing units - Less than $20,000 - 30% or more	Housing Cost Burden
Renter-occupied housing units - $20,000 to $34,999 - 30% or more	Housing Cost Burden
Renter-occupied housing units - $35,000 to $49,999 - 30% or more	Housing Cost Burden
Renter-occupied housing units - $50,000 to $74,999 - 30% or more	Housing Cost Burden
Renter-occupied housing units - $75,000 or more - 30% or more	Housing Cost Burden
Percent of Household Units with 30% or more income towards housing cost	Housing Cost Burden
Percentile Percent of Household Units with 30 percent or more income towards housing cost	Housing Cost Burden
Average of Social and Equity Indicator Percentiles (calculated)	Equity Theme
Percentile percentage of persons (age 5+) who speak English "less than well" estimate	Linguistic Isolation
Resilience Indicator (NRI)	Resilience Theme
Average of Environmental Indicators	Environmental Theme
Percentile for % pre-1960 housing (lead paint indicator)	Environmental
Percentile for Diesel particulate matter level in air	Environmental
Percentile for Air toxics cancer risk	Environmental
Percentile for Air toxics respiratory hazard index	Environmental
Percentile for Ozone level in air	Environmental
Percentile for PM2.5 level in air	Environmental
Transportation Disadvantage Indicator	Transportation
Health Disadvantage Indicator	Health
Economy Disadvantage Indicator	Economy
Equity Disadvantage Indicator	Equity
Resilience Disadvantage Indicator	Resilience
Environmental Disadvantage Indicator	Environmental
Sum of Disadvantage Indicators	Overall
Overall Disadvantage Indicator	Overall

analysis. The variables include percent binge drinking, alcohol consumption per capita, seat belt usage, mobile phone use law, total lane miles, and traffic volume. Percent binge drinking data for 2020 is available at the census tract level and collected from the Center for Disease Control (CDC) 's Behavioral Risk Factor Surveillance System (BRFSS) database (BRFSS 2022). ADT data for 2019 is also available at the census tract level and collected from the National Neighborhood Data Archive (NaNDA) 's traffic volume database (Finlay et al. 2022). The other variables are gathered from various reputable sources at the state level, as census tract data for these variables are unavailable. Total lane miles data for 2019 is obtained from the Federal Highway Administration (FHWA 2023). Alcohol consumption per capita is obtained from the National Institute of Alcohol Abuse and Alcoholism (NIAAA) of the National Institute of Health (NIH) (NIAAA 2021). Mobile phone use law data by state is gathered from the Insurance Institute of Highway Safety (IIHS) (IIHS 2023). Finally, seat belt usage information for 2021 is collected from the National Traffic Safety Administration (NHTSA 2021). By incorporating these state-level variables, a more comprehensive understanding of the analysis can be achieved.

3. Methods

3.1. Data processing

The study framework is illustrated in Fig. 3. In the Disadvantaged indicators dataset, GEO_IDs are developed similarly to the GEO_IDs observed in US Census Bureau's database to join the files with a common identification number. This results in a cleaned dataset for 71,728 census tracts. The fatal crash-based dataset is then merged with this dataset, but due to some tracts not being present in the first dataset, less than 2% crash data are lost. The final cleaned dataset includes 175,169 fatal crashes in 71,728 census tracts. ArcMap is used for spatial visualization of data. Point density of fatal crashes is developed, and fatality rates across races are shown spatially. STATA and RStudio are used for statistical analysis.

3.2. Zero-Hurdle negative binomial regression

Different types of count models have been estimated to create statistical models for predicting roadway crashes. Zero-inflated count models are often estimated to address the issue of excessive zero counts in crash prediction modeling. These models assume the presence of two types of zeros: sampling zeros and structural zeros. Structural zeros represent inherently safe conditions that naturally result in zero crashes, while sampling zeros indicate potential crash situations where zero crashes occur merely by chance. However, considering that traffic crashes can happen under various conditions, assuming the existence of structural zeros may not be entirely realistic. Zero-Hurdle models are considered more suitable for crash analysis since it is unrealistic to assume the existence of structural zeros in traffic safety. Structural zeros refer to crash-free conditions or locations, and such assumptions do not align with the reality of traffic incidents (Lord et al. 2007). The utilization of Zero-Hurdle models assumes that every road segment has the potential for crashes, acknowledging that crashes can happen at any segment. This belief in the possibility of crashes occurring in any segment makes the Zero-Hurdle models more suitable compared to the zero-inflated models.

A zero-inflated Hurdle negative binomial model is considered for over-dispersed count data. It combines two distributions: a Hurdle component and a count component. The Hurdle component models the probability of observing a zero count versus a non-zero count, while the count component models the distribution of the non-zero counts. Zero-Hurdle negative binomial (ZHNB) improves over Zero-Hurdle Poisson model when the count data shows overdispersion. The ZHNB allows for overdispersion and can be used to measure different types of parameters more effectively. The probability distribution of a ZHNB random variable y_i is depicted in Equation (1) below:

$$f(y_i|X_i,\beta,\alpha) = \begin{cases} P_i if y_i = 0 \\ (1-P_i)g(y_i|\mu_i,\alpha), if y_i > 0 \end{cases} \quad (1)$$

Table 2
Area and population in census tracts.

Disadvantage Indicator	Census Tract Status	Census Tracts	Census Tract (%)	Area (Sq. Miles)	Area (%)	Population	Pop. (%)
Economy	Disadvantaged	33,910	47.28	1509429.68	43.12	144,249,443	44.26
	Not-Disadvantaged	37,818	52.72	1991226.54	56.88	181,674,216	55.74
Environment	Disadvantaged	35,585	49.61	250434.07	7.15	159,668,459	48.99
	Not-Disadvantaged	36,143	50.39	3250222.15	92.85	166,255,200	51.01
Equity	Disadvantaged	35,839	49.97	980326.65	28.00	172,857,828	53.04
	Not-Disadvantaged	35,889	50.03	2520329.57	72.00	153,065,831	46.96
Health	Disadvantaged	35,895	50.04	2694498.44	76.97	145,660,714	44.69
	Not-Disadvantaged	35,833	49.96	806157.78	23.03	180,262,945	55.31
Resilience	Disadvantaged	18,065	25.19	1736227.30	49.60	98,391,211	30.19
	Not-Disadvantaged	53,663	74.81	1764428.92	50.40	227,532,448	69.81
Transportation	Disadvantaged	34,702	48.38	2184432.34	62.40	158,930,306	48.76
	Not-Disadvantaged	37,026	51.62	1316223.89	37.60	166,993,353	51.24
Overall	Disadvantaged	21,827	30.43	891104.22	25.46	99,439,604	30.51
	Not-Disadvantaged	49,901	69.57	2609552.00	74.54	226,484,055	69.49
Total		71,728	100	3500656.2	100	325,923,659	100

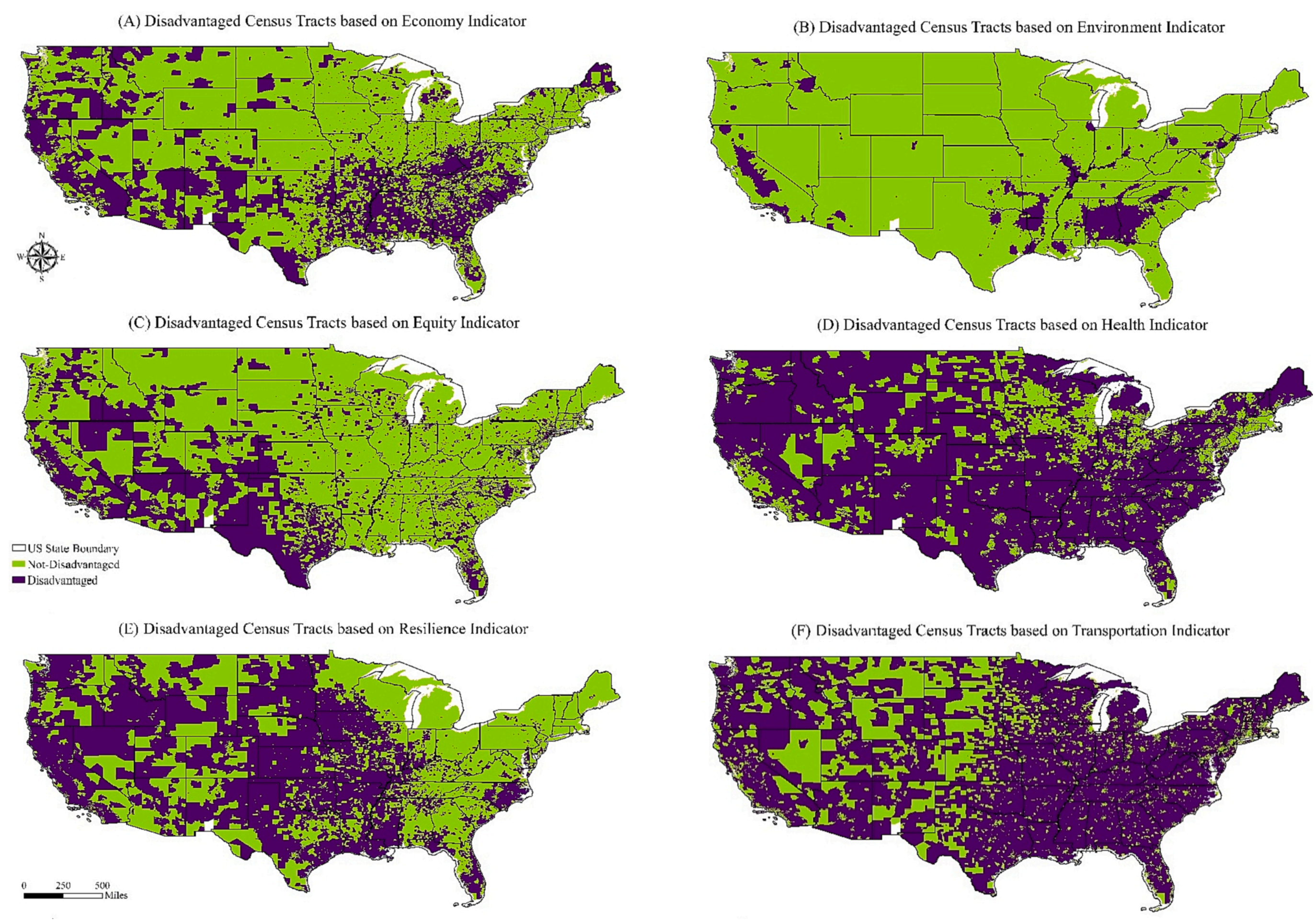

Fig. 1. Disadvantage Census Tracts based on the Six Disadvantage Indicators.

$$P_i = \frac{e^{\delta' \omega_i}}{1 + e^{\delta' \omega_i}} \tag{2}$$

$$g(y_i = 0, 1, 2 \cdots | \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \frac{(1 + \alpha\mu_i)}{1 - (1 + \alpha\mu_i)^{-\alpha^{-1}}}^{-\alpha^{-1} - y_{i\alpha} y_i \mu_i^{y_i}} \tag{3}$$

$$\ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{ik} = X_i \beta \tag{4}$$

where α is denoted as the rate of over-dispersion parameter, X_i denotes the set of independent variables for the NB model, β is the set of coefficients of independent variables, and μ_i denotes the mean crash frequency. Moreover, in Equation (1), P_i is the logistic link function that represents the probability of being a sampling zero. P_i can further be defined with Equation (2), where δ' is a vector of coefficients and ω_i is the covariate of *i*. ZHNB can capture two outcomes, as shown in Equation (1). $y_i = 0$ shows the probability of zero fatal crash occurrences, and

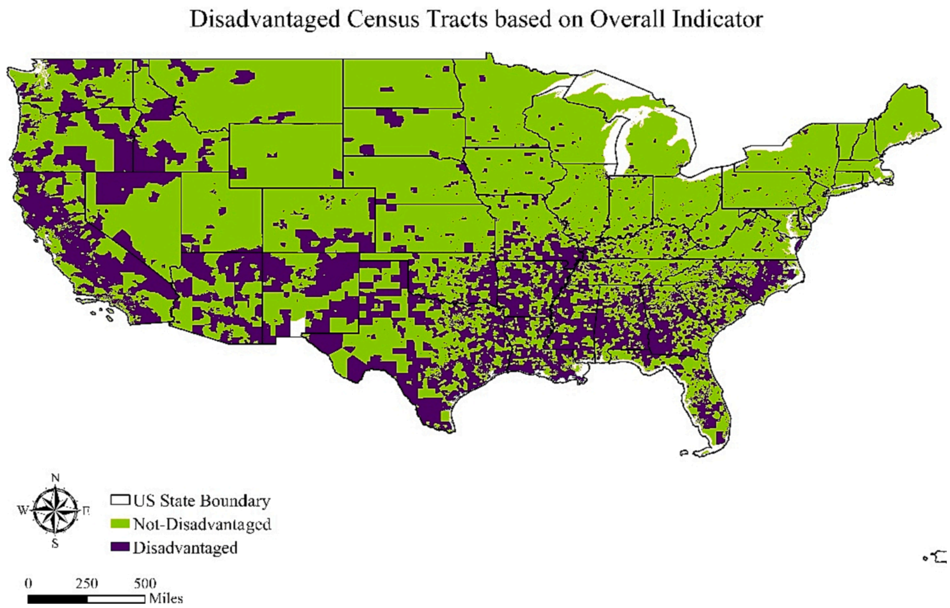

Fig. 2. Disadvantage Census Tracts based on the Overall Disadvantage Indicators.

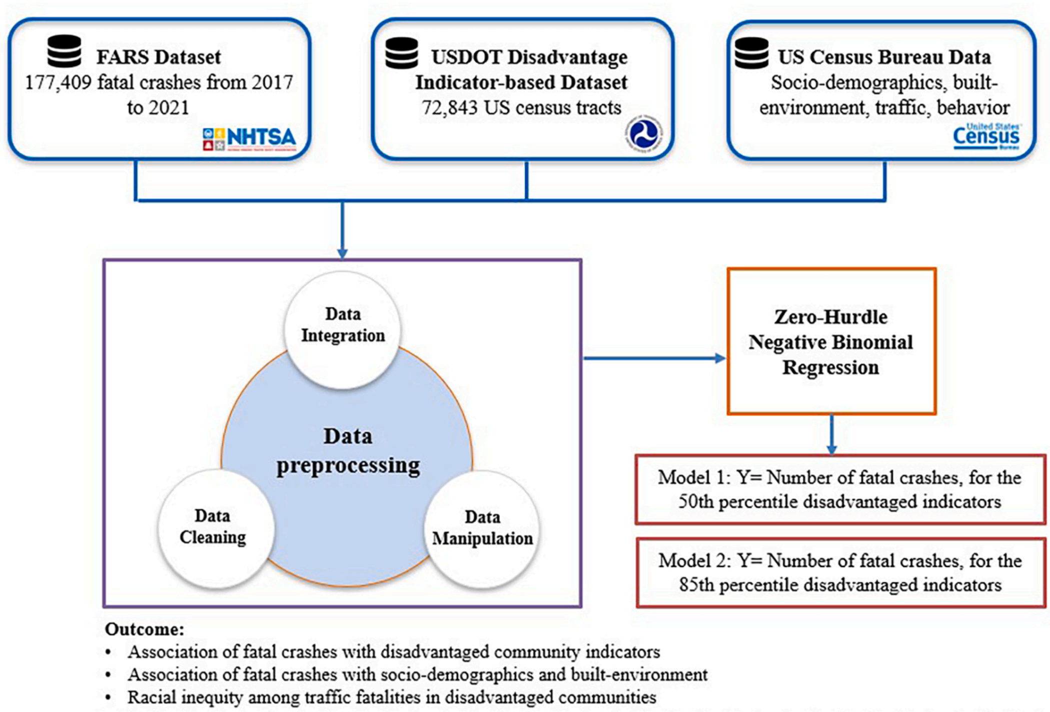

Fig. 3. Study Framework.

$y_i > 0$ denotes the probability of at least one fatal crash occurrence. $g(.)$ is the negative binomial distribution function defined in Equation (3). Notably, $\Gamma(.)$ is the Gamma function. Finally, Equation (4) represents the NB regression model adopted in this study. The *Hurdle* function of *pscl* package in *"R"* software program is used to estimate the ZHNB regression models.

The maximum likelihood estimation method is generally used while estimating ZHNB models. For model selection, the following equations (5) and (6) for Akaike information criterion (AIC) and the Bayesian information criteria (BIC) can be applied to find the fit of the models. A smaller AIC/BIC value indicates a better-performed model.

$$\mathrm{AIC} = -2\mathrm{LL} + 2\mathrm{K} \tag{5}$$

$$\mathrm{BIC} = -2\mathrm{LL} + 2(\ln N)\mathrm{K} \tag{6}$$

where LL is the log-likelihood, K is the number of estimated parameters, and N is the number of observations.

This study uses the entire dataset for model estimation instead of segregating it into calibration and validation subsets. This presents several compelling advantages: 1) the model uses the entire national data to understand the role of DAC indicators, representing a comprehensive set of variations in the entire US. This national-level perspective ensures that the model harnesses the full range of data from all the census tracts, leading to more generalized results. 2) Given that inference (as opposed to prediction with AI methods) is a critical issue in this analysis, maximizing the use of the entire data is important. Splitting the data into training and validation will remove (say) 30 % of the US census tracts, affecting the model's ability to capture key relationships with the available data. Since calibration in our study relies on having sufficient data nationally to estimate accurate probabilities for each location, removing a part of the data can lead to unstable or unreliable calibration estimates in those localities (Šinkovec et al. 2021). So restricting data will hinder the model's statistical power and capacity to uncover relationships across disadvantaged and non-disadvantaged communities. Having said this, to evaluate the performance of the ZHNB models, we generated various metrics, including goodness-of-fit measures (i.e., AIC, BIC, Pseudo-R^2), visual representation of the observed and predicted probabilities of crash occurrence, and overall model significance that shed light on model validity. Overall, this study uses the entire dataset in this context and does not split it into training and validation subsets, avoiding the non-coverage of DACs and non-DACs.

4. Results and discussion

4.1. Descriptive Statistics

Table 3 presents the descriptive statistics of the variables. The mean of the response variable "*Total Fatal Crashes*" is 2.44, indicating an average of 2.44 fatal crashes that occur over five years at the census tract level. A minimum value of zero and a maximum value of 55 indicate that there are census tracts where no fatal crashes occurred, and 55 fatal crashes occurred in one tract. Identification of these tracts can be insightful.

Independent variables consist of demographic characteristics and disadvantage indicators. The *Total Population* of Census tracts has a minimum of 3 and a mean of 4543.82, which indicates that there are tracts where a few people reside, as reported by the US census, but fatal crashes happened there. *Population density* is calculated by dividing the total population by census tract area. The mean population density in the dataset is 5505 persons/census tract. Race or ethnicity type represents the total population of each race or ethnicity in each tract, and their mean values indicate that *Percent White* is the most dominant race (60.89). All races have a minimum value of zero, meaning there are tracts where people of these races are absent. Considering the mean percent values, *Percent Hawaiian or other Pacific Islander*, *Percent American Indian or Alaska Native*, and *Percent Other* are the more minor groups compared to the *Percent Asian*, *Percent Black*, or *Percent Hispanic* populations.

The variables under the Means of Transportation to Work represent the percentage of people who use these modes to reach their workplace. The mean (74.89) value of *Percent Drive Alone* shows that most people drive alone. The maximum (100%) value indicates that people in some census tracts only drive rather than using other modes. People who choose other options are limited compared to those who drive alone to work. According to the mean values, the *Percent Drive Alone* option is followed by *Percent Carpooled*, *Percent Work from Home*, and *Percent Public Transportation*. The mean values of the Education category represent that, on average, the tracts have more *Percent College or Associate Degree* holders (23.89) compared to *Percent Bachelor's Degree or Higher* degree holders (22.96). The mean values of *Percent Less than Highschool Degree* and *Percent Highschool Graduates* indicate such degrees obtained by 9 % and 22 % of the census tract population. Tracts with people of less educational qualification can indicate a disadvantaged community.

The census tracts' income information shows the maximum *median income* is $250,000, and the mean median income is $86,904. *Employment density* is generated by dividing the total employed population by census tract area. The mean employment density is 4,454, indicating, on

Table 3
Descriptive statistics.

Variables	Mean	Std. Dev.	Min.	Max.
Crash Information (5 Yrs. Aggregated)				
Total Fatal Crashes	2.44	3.18	0.00	55.00
Total Fatalities	2.65	3.59	0.00	67.00
Race or Ethnicity				
Percent American Indian or Alaska Native	0.80	4.60	0.00	98.73
Percent Asian	5.21	9.39	0.00	93.77
Percent Black	13.23	20.67	0.00	100.00
Percent Hawaiian or other Pacific Islander	0.17	1.02	0.00	62.48
Percent Other	1.60	4.25	0.00	78.58
Percent White	60.89	28.80	0.00	100.00
Percent Hispanic	17.41	21.15	0.00	100.00
Means of Transportation to Work				
Percent Drove Alone	74.89	14.96	0.00	100.00
Percent Carpooled	9.04	5.34	0.00	58.24
Percent Public Transportation	4.91	10.98	0.00	100.00
Percent Walked	2.67	5.01	0.00	100.00
Percent Taxicab/Bicycle/ Motorcycle/Other	1.88	2.60	0.00	53.97
Percent Work from Home	6.62	5.28	0.00	95.84
Education				
Percent Less than Highschool Degree	9.34	7.46	0.00	666.67
Percent Highschool Graduate	21.99	9.40	0.00	500.00
Percent College or Associate Degree	23.89	26.23	0.00	6700.00
Percent Bachelor's Degree or Higher	22.96	24.29	0.00	5133.33
Geographic Information				
Shape Area (Sq. Miles)	48.80	546.58	0.01	85554.34
Demographic Information				
Total Population	4543.82	2336.62	3.00	72041.00
Population Density (total population/area)	5505.15	12223.75	0.03	271729.60
Economic Information				
Median Income (Dollars)	86904.41	41686.08	2500.00	250000.00
Employment Density (employed population/ area)	4454.64	10059.81	0.00	245705.50
Other Variables				
Average daily traffic (ADT)	14732.41	23266.33	50.00	354000.00
Percent binge drinking	16.83	3.21	2.70	35.00
Alcohol consumption per capita	32.66	4.42	17.70	59.50
Seat belt usage percentage	90.00	4.91	75.50	97.20
Ban on Hand-held Mobile Phone Usage While Driving (1 = Yes, 0 = No)	0.52	0.50	0.00	1.00
Total Lane Miles	260109.80	148957.60	3445.00	683533.00

N (Number of Census Tracts in the Study) = 71,728
Std. Dev. = Standard Deviation

average, 4,454 people are employed per square mile in a census tract. Employment density refers to the concentration of employment opportunities within a given area. It reflects the economic activity and job availability in a specific region. Understanding the impact of employment density on fatal crashes can provide insights into how the availability and proximity of job opportunities may influence traffic patterns, commute patterns, and overall road safety conditions.

The average *ADT* of all the census tracts is 14,732. ADT measures the traffic flow on all roads of a census tract, providing an estimate of the number of vehicles traveling on the roadways daily. The data indicates an average of 260,109 lane miles. The mean *percent binge drinking* is 16.83, ranging from a minimum of 2.7 to a maximum of 35. Binge drinking is defined as consuming a large amount of alcohol within a short period, typically resulting in a blood alcohol concentration (BAC) of 0.08 % or higher. It captures the percentage of individuals living in a census tract who drink such large amounts of alcohol quickly. Besides, the average *alcohol consumption per capita* is 32.99 gallons, ranging from a minimum of 17.7 to a maximum of 59.5 gallons. Alcohol consumption per capita is a measure used to estimate the average amount of alcohol consumed by an individual. The average *seat belt usage* percentage in the dataset is 90, indicating that most people wear seat belts. This high percentage reflects positive safety behavior and contributes to reducing the risk of injuries in accidents. Additionally, the average *ban on hand-held mobile phone usage while driving* is 0.52, indicating that approximately 52 % of the analyzed locations have implemented regulations prohibiting mobile phone use while driving.

4.2. Hurdle models on All-Fatal crashes in the US

Table 4 reports the Hurdle model for the full dataset. Of 71,728 tracts, 20,667 (28.81 %) contain zero observations (no fatal crash), and 51,061 (71.19 %) are non-zero observations (fatal crash present). *Fatal Crash (5 yrs.)* is the response variable predicted by the full model. The Count part of the Hurdle model gives the distribution of the fatal crashes as a negative binomial process. The Zero-Hurdle part is a logistic model predicting whether a census tract will have fatalities. The correlations among the independent variables were examined, specifically checking for correlation coefficients greater than ± 0.5. In cases where two variables exhibited a high correlation, one variable was removed from the analysis based on theoretical and empirical considerations. Examples of such variables include median income and population density. Notably, disadvantaged indicators don't exhibit high correlations. The correlation coefficients between these indicators are less than ± 0.25, as shown in Table 5.

The disadvantaged indicators used in the analysis are binary variables indicating a disadvantaged census tract. The 50th percentile indicators use the USDOT disadvantaged definition, where if the percentage ranking average value of the tracts is greater than and equals 50 % (75 % for the resilience disadvantaged category), it is defined as a disadvantaged tract. This definition may distribute the disadvantaged and non-disadvantaged somewhat evenly across the US. To mitigate such instances and identify substantially disadvantaged communities, we defined highly disadvantaged tracts by ranking percentage values greater than or equal to 85 %. Using this new definition, the 85th percentile-based model contains coefficients with similar signs of the key variables as we found in the 50th percentile model. The AIC value of the 85th percentile-based model is higher than that of the 50th percentile model, indicating a relatively worse fit. Also, the pseudo-R^2 is slightly lower in the 85th percentile model (18 %) than the 50th percentile model (19 %). The 50th percentile-based model is selected for further discussion. The Hurdle model with 50th percentile-based disadvantage indicators is explained. Fig. 4 visually shows the performance of the 50th percentile model, plotting the differences between the observed and predicted probabilities of crash occurrence. When the difference is minimal (i.e., close to zero), it signifies a strong alignment between the model and the data. In the graphical representation, accurate predictions (approximating zero difference) are observed for high crash counts, whereas a moderate level of variability is evident for low crash counts (i.e., the difference is less than 0.011).

The positive coefficient of 0.0663 for *health* indicates that a health-

Table 4
Hurdle Models Results (N = 71728).

Hurdle Models	50th Percentile			85th Percentile		
	Coeff.	z	P>|z|	Coeff.	z	P>|z|
Fatal Crash (5 yrs.)						
Constant	0.2236	3.81	0	1.2580	31.09	0
Disadvantage Indicator						
Environment	−0.0186	−0.75	0.39	0.1723	4.27	0
Health	0.0824	7.05	0	0.1419	4.74	0
Resilience	0.4520	42.35	0	0.4856	38.69	0
Transportation	0.3316	28.83	0	0.1903	3.08	0
Race or Ethnicity						
Percent American Indian or Alaska Native	0.0062	7.16	0	0.0060	6.80	0
Percent Black	0.0014	4.88	0	0.0014	4.95	0
Means of Transportation to Work						
Percent Public Transportation	−0.0037	−3.65	0	−0.0055	−5.35	0
Percent Taxi/ motorcycle/ bicycle/other	−0.0026	−1.78	0.07	−0.0059	−2.65	0.01
Education Information						
Percent Bachelor degree	−0.0054	−9.79	0	−0.0097	−18.52	0
Percent Less than highschool degree	0.0168	16.47	0	0.0187	17.93	0
Other Variables						
Logarithm (ADT)	0.0789	18.16	0	0.0343	7.85	0
Percent binge drinking	0.0045	2.54	0.01	0.0010	0.53	0.6
Ban on Hand-held Mobile Phone Usage While Driving, Yes	−0.0529	−5.18	0	−0.0585	−5.67	0
Employment Density	−0.0001	−19.95	0	−0.0001	−25.01	0
Shape Area (Square Miles)	0.0004	24.97	0	0.0004	23.90	0
Zero Hurdle Model						
Constant	1.4660	63.22	0	1.4660	63.22	0
Population density	−0.0001	−9.25	0	−0.0001	−9.25	0
Work from home percentage	−0.0419	−26.92	0	−0.0419	−26.92	0
Model Fit Statistics						
AIC	280,304			282,592		
BIC	280,497			282,785		
Log-likelihood at convergence	−140631			−141275		
Log-likelihood at null	−172552			−172552		
Pseudo R^2	0.19			0.18		
Degrees of freedom	21			21		
N (number of observations)	71,728			71,728		
Non-zero observations	51,061			51,061		
Zero Observations	20,667			20,667		

Table 5
Correlation coefficients between the disadvantage indicators.

	Transport	Health	Economy	Equity	Resilience	Environment
Transport	1.00					
Health	0.18	1.00				
Economy	0.17	0.19	1.00			
Equity	−0.06	−0.12	0.24	1.00		
Resilience	0.03	0.03	−0.05	0.00	1.00	
Environment	−0.09	−0.12	0.21	0.24	−0.06	1.00

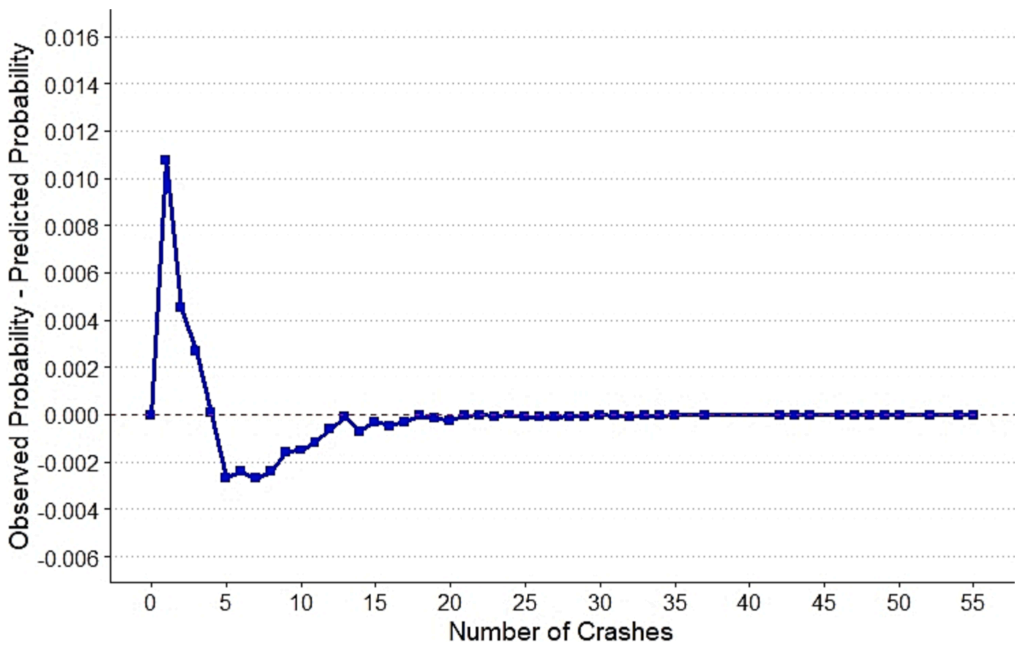

Fig. 4. Difference between observed and predicted probabilities of crash occurrence.

wise disadvantaged community is associated with (exp (-0.0824) or (1.0859–1)*100) = 8.59 % more fatal crashes than a not-disadvantaged community. As per the definition of health-wise disadvantaged tracts, these tracts have more elderly, disabled, and uninsured people prone to crashes while driving or walking to cross the street. *Resilience* has a positive coefficient of 0.4520, which indicates that a resilience-wise disadvantaged tract is associated with 57.14 % more fatal crashes. Resilience-disadvantaged communities are often characterized by a lack of resources, social support, and infrastructure, which can contribute to a higher risk of fatal crashes. For example, these communities may have inadequate public transportation systems, forcing residents to rely on personal vehicles that may not be well-maintained or safe. *Transportation* has a positive coefficient of 0.3316, indicating that transportation-wise disadvantaged tracks have 39.3 % higher fatal crashes. As per the definition of transportation-disadvantaged indicator, people living in transportation-disadvantaged tracts bear high traveling costs, more commute time, and lower vehicle ownership. Moreover, the disproportionate focus on automobile-centered planning and design could have led to a lack of safe and reliable pedestrian, bicycle, and public transportation options, negatively impacting the safety and well-being of low-income and minority populations and senior citizens (Krapp 2020). As a result, transportation-DACs may have a higher risk of fatal crashes.

We found that percent binge drinking is positively correlated with fatal crashes. A one-unit increase in binge drinking in a census tract is associated with a 0.45 % increase in fatal crashes. This is consistent with Voas et al. (2000), who found evidence indicating a strong correlation between higher levels of alcohol consumption in states and an increased proportion of drinking drivers involved in fatal crashes. The percentage binge drinking variable captures the proportion of individuals within a census tract who engage in this risky behavior. Alcohol consumption can significantly impair an individual's judgment, coordination, and motor skills. These impairments can affect a person's ability to make informed decisions, react quickly to potential hazards, and operate a vehicle safely. Binge drinkers can be more inclined to engage in dangerous activities, including reckless driving and speeding, increasing the likelihood of crashes and fatalities. Additionally, a ban on hand-held mobile phone use while driving is negatively associated with 5.15 % fewer fatal crashes. This is similar to the findings by Lim and Chi (2013), who found mobile phone bans significantly reduce fatal crashes in the US. Using a hand-held mobile phone while driving can be a significant distraction, taking the driver's attention away from the road and impairing their ability to react to sudden changes or hazards. By prohibiting this behavior, drivers can focus more on their surroundings and maintain better control of their vehicles, reducing the risk of accidents. Furthermore, this study found a positive association between logarithm (ADT) and the number of fatal crashes. ADT measures traffic volume, representing the number of vehicles traveling on a specific roadway within a given period. Therefore, higher ADT values generally indicate greater exposure to traffic. Specifically, the analysis shows a 1 % increase in ADT increases the likelihood of fatal crashes by 7.9 %. Mohammadnazar et al. (2022) found a similar association between traffic and crashes. With more vehicles on the road, there is an increased potential for interactions between vehicles, pedestrians, and cyclists, which can elevate the risk of fatal crashes.

It is expected that disadvantaged or underserved communities are shelters for minority groups. Among all the race or ethnicity-based variables, only the sign and significance of *Percent American Indian or*

Alaska Native and *Percent Black* are according to expectation. The coefficient of 0.0062 indicates that an additional American Indian or Alaskan person in a tract is associated with higher fatal crashes in the tract by 0.06 %. This is consistent with previous studies (Naimi et al. 2008, Pollack et al. 2012, Murphy et al. 2014). According to Murphy et al. (2014), the fatality rate for American Indian or Alaska Native (AI/AN) individuals is 2.4 times higher than that of Whites. The leading causes of road fatalities in this population are driving while under the influence of alcohol and failing to use seat belts (Pollack et al. 2012). In addition, the coefficient of 0.0001 for the black population suggests that the presence of a more racially black population in the census tract is associated with more fatal crashes by 0.01 %. This indicates that fatal crashes are expected to be higher in communities with more minority groups. It is crucial to prioritize interventions that focus on road safety and invest in the minority communities with the highest fatality rates.

The *Percent Public Transportation* variable has a negative coefficient of 0.0037, which indicates that additional commuters who take public transportation to work are associated with fewer fatal crashes of the tract by 0.3 %. Dharmaratne et al. (2015) suggested a similar association between public transportation and crash fatalities. This makes sense because public transit is a safer mode that uses local roads at low speeds and has fewer chances of involvement in fatal crashes. Likewise, in the 85th percentile case, an increase in commuters who take a taxi/motorcycle/bicycle/other mode to commute is associated with a decrease in fatal crashes by 0.26 %. Although these modes may be associated with lower fatal crashes, their magnitudes are relatively small.

The positive coefficient for lower educational qualification and the negative coefficient for higher education indicate the association with more fatal crashes in tracts with less educated people and fewer crashes in tracts with highly educated people. These findings are consistent with the research conducted by Munteanu et al. (2014), which also showed that individuals with higher education qualifications are less likely to be involved in traffic crashes compared to those with lower educational attainment. Moreover, the study indicates that higher employment density is associated with a lower frequency of fatal crashes. Stamatiadis and Puccini (2000) found lower fatality crash rates in areas with higher employment rates. They suggested a lower prevalence of risky driving behaviors in these areas may contribute to lower rates of crashes.

In the *Zero-Hurdle* model, the *population density* has a negative coefficient. This indicates that if the population per square mile increases, the odds of a zero fatal crash decreases. Work from home percentage has a negative coefficient of −0.0419. It implies that an increase in working from home population is associated with lower odds of observing a zero fatal crash count. Working from home eliminates the need for commuting but encourages non-work travel, including shopping trips, restaurant trips, and recreational trips, among others (Patwary and Khattak 2022).

4.3. Safety in disadvantaged communities by race

Fig. 5 represents the number of fatalities per year per 100,000 population by race. It shows that in the disadvantaged tracts, the highest fatality per year per 100,000 is among the Hawaiian or Other Pacific Islander (HPI) (54.98), followed by American Indian or Alaska Native (29.88), White (22.58), Black (17.37), and Hispanic (12.70) population. The lowest fatality/per 100,000 is observed among Asians. Fig. 6 shows the fatality rates of the minority groups and indicates that the higher fatality rates are spatially distributed in the disadvantaged areas. According to the definition of the DAC indicators, this may be due to a lack of access to safe and reliable vehicles, poor road conditions, and other systematic inequalities. Therefore, these communities can be the target areas for safety improvement projects. Addressing the higher fatality rate in DACs may require a multifaceted approach that includes improving infrastructure, expanding access to safe vehicles, strengthening healthcare access, providing comprehensive safety education, and addressing systemic inequities.

4.4. Limitations

Any potential measurement errors and non-coverage errors are recognized. Integrating different data sources resulted in data loss (less than 2%). Furthermore, a small amount of census tracts (less than 1%) in the GIS shapefile and disadvantage indicator-based dataset did not match. Specifically, the disadvantage database did not have information for some of the census tracts. Those tracts were excluded from the analysis. Some fatal crashes (1.5%) from each year were lost during the spatial joining process in ArcGIS. They were on the boundaries of the tracts and were not counted. Despite these limitations, the study successfully identies safety risks in disadvantaged communities.

5. Conclusions

This study comprehensively investigates traffic safety by exploring the role of different disadvantaged community indicators, sociodemographics, and built environments using unique and high-quality data. The data collected by the US Department of Transportation on

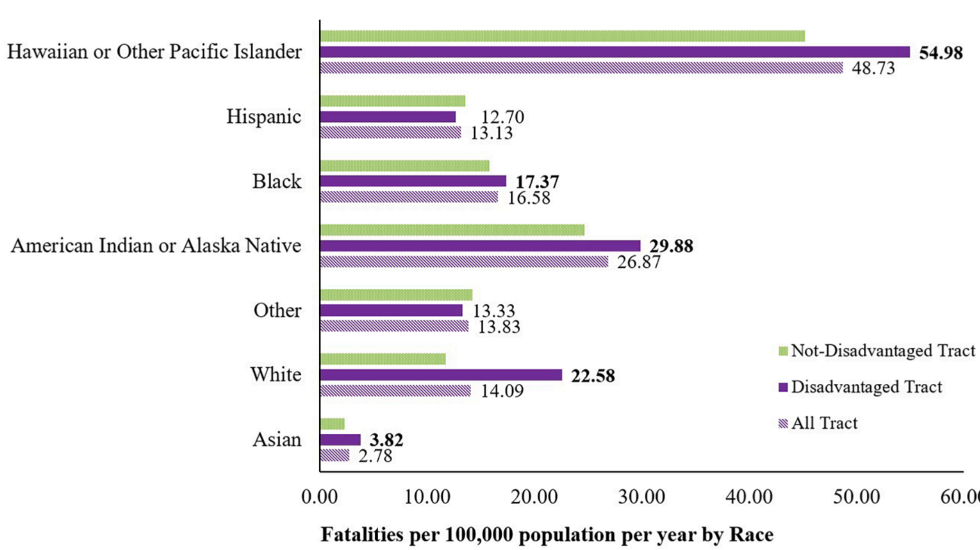

Fig. 5. Fatalities per year per 100,000 population across different races in the US.

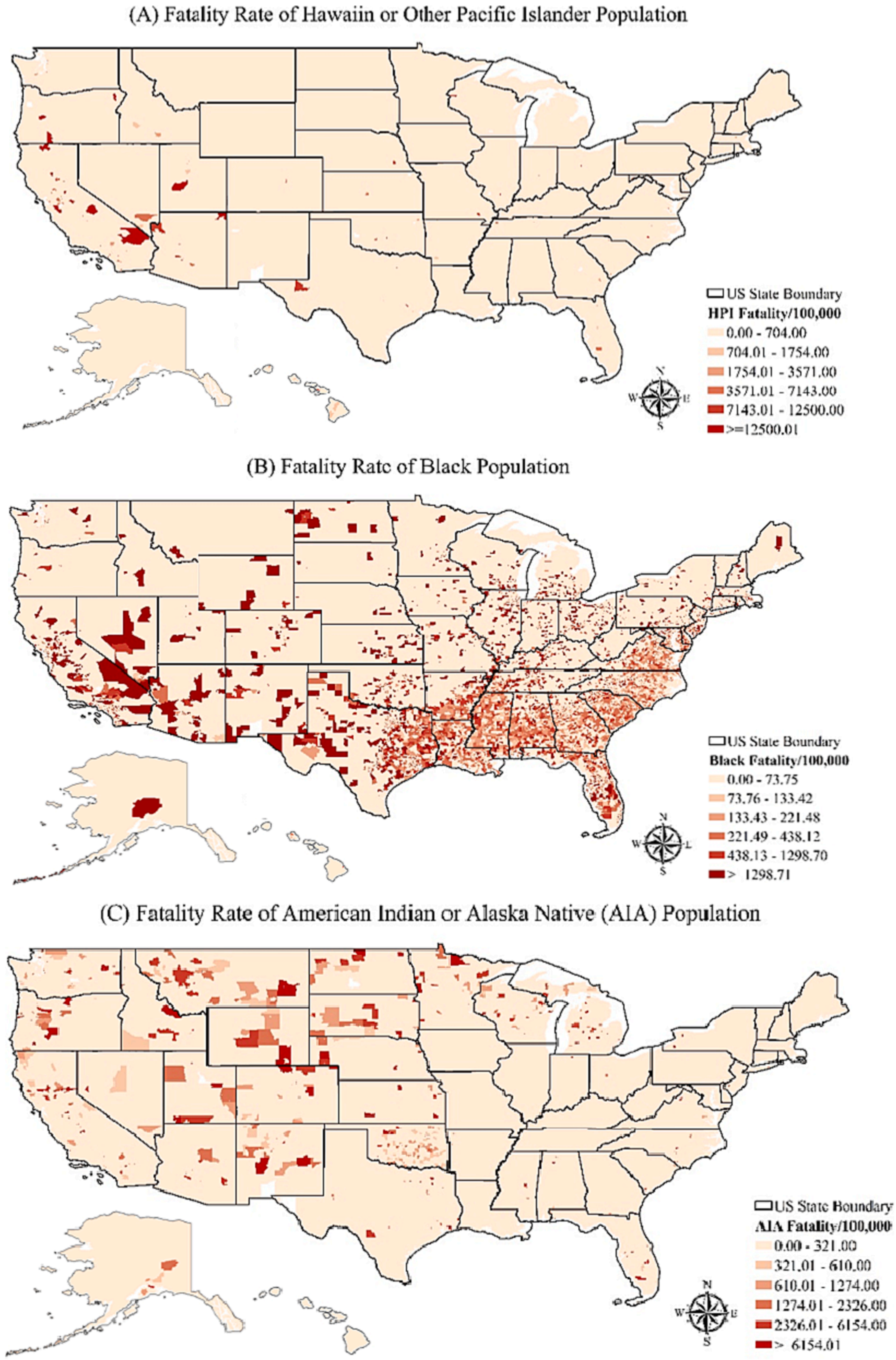

Fig. 6. Fatality Rate of (a) HPI (b) Black and (c) AIA Population in the US.

disadvantaged communities at the census tract level was linked with fatal crash data from FARS. The analysis, conducted at the census tract level with five years of data, estimated inference-based Zero-Hurdle negative binomial models. The study controls commonly known confounding factors by including diverse variables. The analysis was done at a disaggregated and granular spatial level, capturing local variations.

The findings revealed several important insights. Firstly, health, resilience, and transportation-disadvantaged tracts are associated with more fatal crashes in the US, highlighting the impact of multiple dimensions of disadvantage on safety outcomes. Specifically, census tracts with health, resilience, and transportation disadvantages have an 8.59 %, 57.14 %, and 39.30 % higher rate of fatal crashes, respectively. Additionally, higher rates of fatal crashes were associated with census tracts characterized by high traffic volume. Also, higher levels of binge drinking, and no mobile phone law are associated with increased fatal crashes. The study also found that a higher percentage of the population with bachelor's degrees and increased public transportation use correlate with fewer fatal crashes. Conversely, a higher proportion of Black, American Indian or Alaska Native populations were associated with a greater number of fatal crashes. Subsequently, in the DACs, the highest fatality per 100,000 is among the Hispanic or other Pacific Islander, and American Indian or Alaska Native populations. This implies that a comprehensive approach is required to address the elevated fatalities in DACs, including infrastructure improvements, increased access to safe vehicles, improved healthcare access, comprehensive safety education, and addressing systemic inequalities.

This study contributes new knowledge about safety in diverse contexts characterized by disadvantaged communities, socio-demographics, and built environments. The results generated in this study can support national-level planning. The findings provide valuable information for policymakers, helping them allocate resources effectively, invest in more equitable safety measures, and prioritize improvements in DACs. Since transportation and housing costs are becoming increasingly burdensome for low and medium-income households, investments should be devoted to improving access to a range of high-quality, safe, and affordable mobility options, i.e., transit, shared mobility, and active transportation options within disadvantaged communities. Specifically, since we found more use of public transportation improves overall safety, investments can be directed to enhancing access to public transportation options, including bus stops and transit stations, to reduce the reliance on private vehicles in DACs. Moreover, disadvantaged communities should be supported in playing an active and direct role in transportation planning, engagement, and decision-making processes, ensuring historically excluded voices are centered in the transportation decision-making process. This will not only involve community members in the planning and design of road safety interventions but also will ensure that solutions are culturally sensitive and address local concerns. Notably, implementing policies, e.g., banning hand-held mobile phone use while driving and enhancing traffic enforcement, are crucial for reducing traffic safety risks in DACs. Ultimately, this research aims to enhance safety and equity in transportation and guide policymakers toward evidence-based decision-making. Future studies can incorporate spatial analysis to observe whether the relations vary over space. Besides, future research should be conducted on different types of crashes, e.g., pedestrian-involved, large truck-involved, and rear-end crashes in DACs. Moreover, future studies can include more socio-demographic information on the drivers. This can assist policymakers in deciding which disadvantaged community might need transportation improvement on a priority basis, which can ultimately take the United States one step closer to the vision zero goals.

CRediT authorship contribution statement

A. Latif Patwary: Conceptualization, Methodology, Validation, Investigation, Writing – review & editing, Formal analysis, Writing – original draft, Visualization, Data curation. **Antora Mohsena Haque:** Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Visualization, Data curation. **Iman Mahdinia:** Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Visualization. **Asad J. Khattak:** Conceptualization, Validation, Investigation, Funding acquisition, Visualization, Supervision, Project administration, Resources, Writing – review & editing.

Declaration of Competing Interest

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

Data availability

The data used for analysis in this study are publicly available. The authors will also make the data available on request.

Acknowledgments

The authors would like to thank the Collaborative Sciences Center for Road Safety (CSCRS), a U.S. DOT-funded National Transportation Center, under Grant No. 69A3551747113, for providing financial support for this article's research, authorship, and/or presentation/publication.

References

Ahmed, M.M., Franke, R., Ksaibati, K., Shinstine, D.S., 2018. Effects of truck traffic on crash injury severity on rural highways in wyoming using bayesian binary logit models. Accid. Anal. Prev. 117, 106–113.
Arvin, R., Kamrani, M., Khattak, A.J., 2019. How instantaneous driving behavior contributes to crashes at intersections: extracting useful information from connected vehicle message data. Accid. Anal. Prev. 127, 118–133.
Azetsop, J., 2010. Social justice approach to road safety in kenya: Addressing the uneven distribution of road traffic injuries and deaths across population groups. Public Health Ethics 3 (2), 115–127.
Baireddy, R., Zhou, H., Jalayer, M., 2018. Multiple correspondence analysis of pedestrian crashes in rural illinois. Transp. Res. Rec. 2672 (38), 116–127.
Barajas, J.M., 2018. Not all crashes are created equal. J. Transp. Land Use 11 (1), 865–882.
Buehler, R., Pucher, J., 2012. Cycling to work in 90 large american cities: new evidence on the role of bike paths and lanes. Transportation 39, 409–432.
Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M., 2016. Macro-level pedestrian and bicycle crash analysis: incorporating spatial spillover effects in dual state count models. Accid. Anal. Prev. 93, 14–22.
Campos-Outcalt, D., Bay, C., Dellapena, A., Cota, M.K., 2003. Motor vehicle crash fatalities by race/ethnicity in Arizona, 1990–96. Inj. Prev. 9 (3), 251–256.
Chen, C., Lin, H., Loo, B.P., 2012. Exploring the impacts of safety culture on immigrants' vulnerability in non-motorized crashes: a cross-sectional study. J. Urban Health 89, 138–152.
Chimba, D., Musinguzi, A., Kidando, E., 2018. Associating pedestrian crashes with demographic and socioeconomic factors. Case Studies on Transport Policy 6 (1), 11–16.
NIAAA, 2021. Alcohol Consumption by State. Retrieved June 25, 2023 from https://wisevoter.com/state-rankings/alcohol-consumption-by-state/.
NHTSA, 2021. Seat Belt Use in 2021 - Use Rates in the States and Territories. Retrieved June 25, 2023 from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813307.
BRFSS, 2022. Behavioral Risk Factor Surveillance System. Places: Census Tract Data (GIS Friendly Format), 2022 release. Retrieved June 25, 2023 from https://data.cdc.gov/500-Cities-Places/PLACES-Census-Tract-Data-GIS-Friendly-Format-2022-/shc3-fzig.
Christie, N., Ward, H., Kimberlee, R., 2010. Road traffic injury risk in disadvantaged communities: Evaluation of the neighbourhood road safety initiative.
Cradock, A.L., Troped, P.J., Fields, B., Melly, S.J., Simms, S.V., Gimmler, F., Fowler, M., 2009. Factors associated with federal transportation funding for local pedestrian and bicycle programming and facilities. J. Public Health Policy 30, S38–S72.
Dharmaratne, S.D., Jayatilleke, A.U., Jayatilleke, A.C., 2015. Road traffic crashes, injury and fatality trends in Sri lanka: 1938–2013. Bull. World Health Organ. 93, 640–647.
Dill, J., Carr, T., 2003. Bicycle commuting and facilities in major us cities: if you build them, commuters will use them. Transp. Res. Rec. 1828 (1), 116–123.
Eisenberg, D., Warner, K.E., 2005. Effects of snowfalls on motor vehicle collisions, injuries, and fatalities. Am. J. Public Health 95 (1), 120–124.
FHWA, 2023. Highway statistics series. Retrieved June 25, 2023 from https://www.fhwa.dot.gov/policyinformation/statistics/2019/hm60.cfm.
Finlay, J., Robert, M., Michael, E., Anam, K., Mao, L., Iris, G.-L., Philippa, C., Megan, C., 2022. National neighborhood data archive (nanda): Traffic volume by census tract,

united states, 1963-2019. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].
Gibbs, K., Slater, S.J., Nicholson, L., Barker, D., Chaloupka, F.J., 2012. Income disparities in street features that encourage walking. Bridging the Gap Program.
Giles-Corti, B., Donovan, R.J., 2002. Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment. Prev. Med. 35 (6), 601–611.
Gu, Y., Liu, D., Arvin, R., Khattak, A.J., Han, L.D., 2023. Predicting intersection crash frequency using connected vehicle data: A framework for geographical random forest. Accid. Anal. Prev. 179, 106880.
Hosseinpour, M., Prasetijo, J., Yahaya, A.S., Ghadiri, S.M.R., 2013. A comparative study of count models: application to pedestrian-vehicle crashes along Malaysia federal roads. Traffic Inj. Prev. 14 (6), 630–638.
Hosseinpour, M., Yahaya, A.S., Sadullah, A.F., 2014. Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: case studies from Malaysian federal roads. Accid. Anal. Prev. 62, 209–222.
IIHS, 2023. Cellphone use laws. Retrieved June 25, 2023 from https://www.iihs.org/topics/distracted-driving/cellphone-use-laws.
Jacobsen, P.L., Rutter, H., 2012. Cycling safety MIT Press Cambridge, MA.
Kadilar, G.O., 2016. Effect of driver, roadway, collision, and vehicle characteristics on crash severity: a conditional logistic regression approach. Int. J. Inj. Contr. Saf. Promot. 23 (2), 135–144.
Kamrani, M., Arvin, R., Khattak, A.J., 2018. Extracting useful information from basic safety message data: an empirical study of driving volatility measures and crash frequency at intersections. Transp. Res. Rec., 0361198118773869
Khattak, A.J., Wali, B., 2017. Analysis of volatility in driving regimes extracted from basic safety messages transmitted between connected vehicles. Transportation Research Part C: Emerging Technologies 84, 48–73.
Khojastehpour, M., Sahebi, S., Samimi, A., 2022. Public acceptance of a crowdsourcing platform for traffic enforcement. *Case studies on transport*. policy 10 (4), 2012–2024.
Krapp, A., 2020. Transportation equity project prioritization criteria. MUP Capstone.
Lee, J., Abdel-Aty, M., Choi, K., 2014. Analysis of residence characteristics of at-fault drivers in traffic crashes. Saf. Sci. 68, 6–13.
Li, X., Yu, S., Huang, X., Dadashova, B., Cui, W., Zhang, Z., 2022. Do underserved and socially vulnerable communities observe more crashes? A spatial examination of social vulnerability and crash risks in texas. Accid. Anal. Prev. 173, 106721.
Lim, S.H., Chi, J., 2013. Cellphone bans and fatal motor vehicle crash rates in the united states. J. Public Health Policy 34 (2), 197–212.
Liu, J., Khattak, A.J., 2016. Delivering improved alerts, warnings, and control assistance using basic safety messages transmitted between connected vehicles. Transportation Research Part c: Emerging Technologies 68, 83–100.
Loeb, P.D., Clarke, W.A., 2009. The cell phone effect on pedestrian fatalities. Transportation Research Part e: Logistics and Transportation Review 45 (1), 284–290.
Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accid. Anal. Prev. 37 (1), 35–46.
Lord, D., Washington, S., Ivan, J.N., 2007. Further notes on the application of zero-inflated models in highway safety. Accid. Anal. Prev. 39 (1), 53–57.
Mohammadnazar, A., Patwary, A.L., Moradloo, N., Arvin, R., Khattak, A.J., 2022. Incorporating driving volatility measures in safety performance functions: improving safety at signalized intersections. Accid. Anal. Prev. 178, 106872.
Munteanu, P.L., Rosu, M., Panaitescu, V., Pungă, A., 2014. Human and environmental factors contributing to fatal road accidents in a Romanian population. Rom J Leg Med 22 (2), 97–100.
Murphy, T., Pokhrel, P., Worthington, A., Billie, H., Sewell, M., Bill, N., 2014. Unintentional injury mortality among american indians and Alaska natives in the United States, 1990–2009. Am. J. Public Health 104 (S3), S470–S480.
Naimi, T.S., Cobb, N., Boyd, D., Jarman, D.W., Brewer, R., Nelson, D.E., Holt, J., Espey, D., Snesrud, P., Chavez, P., 2008. Alcohol-attributable deaths and years of potential life lost among american indians and alaska natives–united states, 2001–2005. MMWR: Morbidity and mortality weekly report 57 (34), 938-941.
Naimi, T.S., Xuan, Z., Sarda, V., Hadland, S.E., Lira, M.C., Swahn, M.H., Voas, R.B., Heeren, T.C., 2018. Association of state alcohol policies with alcohol-related motor vehicle crash fatalities among us adults. JAMA Intern. Med. 178 (7), 894–901.
Nantulya, V.M., Reich, M.R., 2003. Equity dimensions of road traffic injuries in low-and middle-income countries. Inj. Control Saf. Promot. 10 (1–2), 13–20.
Nguyen-Hoang, P., Yeung, R., 2014. Dollars for lives: the effect of highway capital investments on traffic fatalities. J. Saf. Res. 51, 109–115.
Noland, R.B., Klein, N.J., Tulach, N.K., 2013. Do lower income areas have more pedestrian casualties? Accid. Anal. Prev. 59, 337–345.
Patwary, A.L., 2023. Bus safety during covid-19: analyzing bus-involved crashes and economic harm. Int. J. Crashworthiness 1–8.
Patwary, A.L., Khattak, A.J., 2022. Interaction between information and communication technologies and travel behavior: Using behavioral data to explore correlates of the COVID-19 pandemic. Transp. Res. Rec., 03611981221116626
Patwary, A.L. and Khattak, A., 2022. A Study of Truck-Involved Crashes and Fatalities Before and During the COVID-19 Pandemic: Evidence of Traffic Safety from Tennessee.(Ed.),ˆ. In *101st Annual Meeting of Transportation Research Board (TRB)*.
Patwary, A.L., Khattak, A.J., 2023. Crash harm before and during the covid-19 pandemic: evidence for spatial heterogeneity in tennessee. Accid. Anal. Prev. 183, 106988.
Pei, X., Wong, S.C., Sze, N.N., 2012. The roles of exposure and speed in road safety analysis. Accid. Anal. Prev. 48, 464–471.
Pollack, K.M., Frattaroli, S., Young, J.L., Dana-Sacco, G., Gielen, A.C., 2012. Motor vehicle deaths among American Indian and Alaska native populations. Epidemiol. Rev. 34 (1), 73–88.
Pour, M.H., Prasetijo, J., Yahaya, A.S., Ghadiri, S.M.R., 2012. Modeling vehicle-pedestrian crashes with excess zero along Malaysia federal roads. Procedia. Soc. Behav. Sci. 53, 1216–1225.
Pulugurtha, S.S., Duddu, V.R., Kotagiri, Y., 2013. Traffic analysis zone level crash estimation models based on land use characteristics. Accid. Anal. Prev. 50, 678–687.
Rogers, M.M., Weber, W.L., 2011. Evaluating Co_2 emissions and fatalities tradeoffs in truck transport. Int. J. Phys. Distrib. Logist. Manag.
Roudsari, B., Ramisetty-Mikler, S., Rodriguez, L.A., 2009. Ethnicity, age, and trends in alcohol-related driver fatalities in the united states. Traffic Inj. Prev. 10 (5), 410–414.
Rudisill, T.M., Barbee, L.O., Hendricks, B., 2023. Characteristics of fatal, pedestrian-involved, motor vehicle crashes in west virginia: a cross-sectional and spatial analysis. Int. J. Environ. Res. Public Health 20 (7), 5251.
Saha, S., Schramm, P., Nolan, A., Hess, J., 2016. Adverse weather conditions and fatal motor vehicle crashes in the united states, 1994–2012. Environ. Health 15 (1), 1–9.
Sanders, R.L., Schneider, R.J., 2022. An exploration of pedestrian fatalities by race in the United States. Transp. Res. Part D: Transp. Environ. 107, 103298.
Schneider, R.J., 2020. United states pedestrian fatality trends, 1977 to 2016. Transp. Res. Rec. 2674 (9), 1069–1083.
Shin, E.J., 2023. Decomposing neighborhood disparities in bicycle crashes: a gelbach decomposition analysis. Transp. Policy 131, 156–172.
Šinkovec, H., Heinze, G., Blagus, R., Geroldinger, A., 2021. To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets. BMC Med. Res. Method. 21, 1–15.
Son, H.D., Kweon, Y.-J., Park, B.B., 2011. Development of crash prediction models with individual vehicular data. Transportation Research Part c: Emerging Technologies 19 (6), 1353–1363.
Stamatiadis, N., Puccini, G., 1999. Fatal crash rates in the southeastern united states: why are they higher? Transp. Res. Rec. 1665 (1), 118–124.
Stamatiadis, N., Puccini, G., 2000. Socioeconomic descriptors of fatal crash rates in the southeast USA. Inj. Control Saf. Promot. 7 (3), 165–173.
Traynor, T.L., 2009. The relationship between regional economic conditions and the severity of traffic crashes. Traffic Inj. Prev. 10 (4), 368–374.
USDOT, 2022. National Roadway Safety Strategy. Retrieved May 10, 2022 from https://www.transportation.gov/NRSS.
Voas, R.B., Tippetts, A.S., Fell, J., 2000. The relationship of alcohol safety laws to drinking drivers in fatal crashes. Accid. Anal. Prev. 32 (4), 483–492.
Wali, B., Khattak, A.J., Bozdogan, H., Kamrani, M., 2018. How is driving volatility related to intersection safety? A bayesian heterogeneity-based analysis of instrumented vehicles data. Transportation Research Part c: Emerging Technologies 92, 504–524.
USDOT, 2023. Transportation disadvantaged census tracts (historically disadvantaged communities) interim definition methodology. Retrieved January 25, 2023 from https://www.transportation.gov/priorities/equity/justice40/transportation-disadvantaged-census-tracts-historically-disadvantaged.
White-House, 2022. Justice40 initiative - environmental justice - the white house. Retrieved December 30, 2022 from https://www.whitehouse.gov/environmentaljustice/justice40/.
Yu, C.-Y., Zhu, X., Lee, C., 2022. Income and racial disparity and the role of the built environment in pedestrian injuries. J. Plan. Educ. Res. 42 (2), 136–149.