12-2023

Transportation System Performance and Traveler Behavior in the Context of a Systemwide Shock: Applications of Data Science Toward a Sustainable Future

A. Latif Patwary
apatwary@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by A. Latif Patwary entitled "Transportation System Performance and Traveler Behavior in the Context of a Systemwide Shock: Applications of Data Science Toward a Sustainable Future." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Civil Engineering.

Asad J. Khattak, Major Professor

We have read this dissertation and recommend its acceptance:

Asad J. Khattak, Candace Brakewood, Kevin Heaslip, Russell Zaretzki

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Transportation System Performance and Traveler Behavior in the Context of a Systemwide Shock: Applications of Data Science Toward a Sustainable Future

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

A. Latif Patwary
December 2023

This dissertation is dedicated to my parents, and my beloved family who constantly encouraged me to pursue my dreams and finish my dissertation.

ACKNOWLEDGMENTS

First and most of all, I am profoundly grateful to the Almighty God for the strength, wisdom, and grace bestowed upon me, which have been indispensable throughout my doctoral journey.

I would like to express my deepest appreciation to my major adviser, Dr. Asad J. Khattak. His unwavering support, expert guidance, and invaluable mentorship have been pivotal in my academic and personal growth. His leadership and mentoring were instrumental in my ability to participate in and make meaningful contributions to the research program at the University of Tennessee. Working under his supervision has been my honor, and I will always be proud of my association with him. My heartfelt thanks extend to the esteemed members of my dissertation committee: Dr. Candace Brakewood, Dr. Russell Zaretzki, and Dr. Kevin Heaslip. Their insights, feedback, and support have significantly enriched my research experience. Also, special thanks to Dr. Zaretzki for supervising my M.S. statistics degree.

I am also immensely thankful to my co-authors, Dr. Asad J. Khattak, Dr. Iman Mahdinia, and Dr. Antora Mohsena Haque. Their profound knowledge and valuable insights have been instrumental in shaping and enhancing my research. I also want to thank all my lab mates for their continuous support and memorable experiences. A special note of gratitude goes to the faculty and staff of the Department of Civil and Environmental Engineering. Their support and encouragement have been a cornerstone of my academic journey. I am also grateful to the Collaborative Sciences Center for Road Safety, the US Department of Transportation, and the Tennessee Department of Transportation for greatly supporting my research and academic endeavors.

Last but certainly not least, I extend my deepest gratitude to my beloved wife, Dr. Tonny, whose love, patience, and unwavering support have been my constant source of encouragement. To my mother, father, and family members, whose sacrifices, guidance, and unconditional love have shaped the person I am today, I owe an eternal debt of gratitude. Their belief in me and their invaluable support have been the bedrock of my strength and perseverance.

ABSTRACT

The COVID-19 pandemic, a systemwide shock, has left a long-lasting and significant impact on transportation systems. It has contributed to a shift in travel behavior, with many people turning to work from home (WFH) and online shopping. This shift has led to a reduction in vehicular travel. However, the pandemic witnessed increased crash fatalities despite a reduction in overall crashes, disproportionately affecting disadvantaged communities (DACs). The main question arising from these pandemic-related issues is what we can learn to improve transportation systems and shape future travel behavior. Therefore, this dissertation aims to investigate how the transportation system changed during COVID-19 and explore the future implications while examining the travel behavior, technology adoption behavior, and road safety aspects in DACs compared with non-DACs during COVID-19. As such, this dissertation first explores the interaction between WFH, online shopping, and in-person shopping behaviors, revealing nuanced relationships that have evolved amidst the pandemic. Second, comprehensive safety data are utilized to dissect why crash fatalities increased during COVID-19. Third, transportation safety in DACs is investigated by leveraging safety data covering COVID-19 periods and the comprehensive DAC indicators developed by the US Department of Transportation. Fourth, DACs' shopping behavior during COVID-19 is analyzed by focusing on the interplay of emerging online delivery components (retail, grocery, and food) and in-person activities. Finally, the study compares technology adoption behaviors between DACs and non-DACs by exploring infrastructure and socio-economic barriers. Methodologically speaking, this dissertation employs various state-of-the-art statistical and explainable artificial intelligence techniques. Overall findings indicate that compared to pre-COVID-19, the surge in WFM and e-commerce trends was associated with a substantial reduction in physical shopping trips during COVID-19. Speeding and reckless behaviors were strongly associated with the increased road fatalities. DACs experienced heightened adversity than non-DACs, associated with a higher rate of fatal crashes (an increase of 8% to 57%). Online orders were considerably less frequent in DACs than non-DACs (2% to 7%), emphasizing disparity in digital infrastructure. Additionally, technology adoption rates were significantly lower in DACs. These findings underscore the importance of better preparedness and planning for such communities to be equipped to handle future systemic shocks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

Although the COVID-19 pandemic, a systemwide shock, may be a once-in-a-lifetime event, it has long-lasting and significant impacts on various aspects of our daily lives, including transportation. It has contributed to a shift in travel behavior, with many people turning to work from home (WFH) and online shopping, leading to decreased trips taken for commuting and non-essential purposes. This shift is reflected in the fact that Americans drove 13% fewer miles in 2020 than in 2019, resulting in a positive impact on reducing air pollution and carbon emissions. Moreover, the decrease in the number of vehicles on the road had the potential to reduce the total number of crashes and fatalities during the pandemic. However, reports suggest that while there were fewer crashes during COVID-19, crash fatalities increased substantially. This also affected disadvantaged communities (DACs), such as those with lower incomes or limited access to transportation options, and they may have experienced different impacts than non-DACs. This has raised concerns about the equity implications of pandemic-related travel behavior and road safety changes. A graphical abstract showing the motivation of the study is provided in Figure 1-1.

The main question arising from these pandemic-related issues is what we can learn to improve transportation safety and shape future travel behavior. As such, this dissertation aims to answer two specific research questions: 1) how the transportation system changed during COVID-19 compared to the pre-COVID-19 periods, and 2) how DACs were affected compared with non-DACs during COVID-19 in terms of travel behavior, technology adoption behavior, and road safety aspects. The overall framework of the study is presented in Figure 1-2. The study employs various state-of-the-art statistical and explainable artificial intelligence techniques and utilizes unique and comprehensive databases. It also addresses methodological issues like spatial heterogeneity and unobserved endogeneity. Additionally, Figure 1-3 helps visualize the design of the study chapters. "X" denotes the introduction or change (e.g., the onset of COVID-19), O_1 and O_2 represent the observation or measurement of the relevant output variables (e.g., traffic fatalities, work from home) before and during the COVID-19 pandemic, and O_2' and O_2'' signify measurements for DACs and non-DACs. Chapters 2 and 3 address the first research question by exploring the changes that occurred during COVID-19 compared to the pre-COVID-19 periods. Chapters 4, 5, and 6 address the second research question by examining the differences between DAC and non-DACs regarding travelers' behavior and road safety during COVID-19. By addressing these research questions, this dissertation creates new knowledge for the post-pandemic era. It provides significant implications for future transportation planning models and the development of more equitable traffic safety programs.

In summary, ***Chapter 2*** of this dissertation utilizes survey data to examine WFH, online shopping, and in-person shopping behavior before and during COVID-19. Generally, online shopping can lead to fewer shopping trips; similarly, WFH may reduce work-related trips. However, more WFH has the potential to generate other non-work trips, including shopping trips. To find answers and explore interdependencies, this study analyzes WFH and online shopping together to find their impacts on shopping trip generation. It further explores whether the relationships are similar before and during the COVID-19 pandemic. This chapter jointly analyzes the relationships among shopping trips, online shopping, and WFH with a conditional mixed process model that addresses unobserved endogeneity. The implications of the results are discussed with consideration of improving existing travel demand models in the post-pandemic era.

Chapter 3 adopts a comprehensive analysis of safety-critical events on roadways during COVID-19 by using Trips-by-Distance data and safety data from the Tennessee Integrated Transportation Network (TITAN), which collects police-reported crash data. This chapter examines the factors associated with the increased crash fatalities during COVID-19. In addition to crash fatalities, total monetary harm is used as an additional safety measure. Data statistics show that while crash fatalities increased by 8.2%, total crashes decreased by 15.3%, and the total harm cost was lower by about $1.76 billion during COVID-19 (2020) compared with pre-COVID-19 conditions (2019). Several models, including generalized least squares linear, Poisson, and geographically weighted regression models using the differences between 2020 and 2019 values, are estimated to rigorously quantify the correlates of crash fatalities and crash harm. The results and implications of this study are discussed from the lens of traffic law enforcement and safety countermeasures.

Chapter 4 uses census and safety data to understand transportation safety in DACs. The data periods include the COVID-19 timeframe. To identify DACs, the US department of transportation (USDOT) has established six comprehensive indicators: economy, environment, equity, health, resilience, and transportation access. These indicators are utilized to explore the associations between DACs and fatal crashes, providing a comprehensive understanding of safety risks in DACs. The study identifies the common contributing factors to safety in DACs using Zero-Hurdle Negative Binomial (ZINB) models. The findings and implications of the study are discussed by highlighting the importance of implementing more equitable safety programs in DACs.

Chapter 5 examines the existence of a digital divide in DACs by analyzing the emerging components of online shopping, i.e., retail, grocery, and food, during COVID-19. These components of online shopping and their interactions are relatively unexplored in DACs that are underserved, marginalized, and overburdened by pollution. Given the limited access to digital resources and the disproportionate impact of COVID-19 on DACs, it is crucial to study these components to ensure that DACs are not left behind in the transition to digital commerce. Hence, this chapter aims to provide a comprehensive understanding of travel behavior changes by analyzing the interconnectedness of online shopping and in-person activities in both DACs and non-DACs. A unique household-level database is created by linking the 2021 Puget Sound Household Travel Survey and the US Department of Transportation's Justice40 databases, and a conditional mixed process model is estimated to account for unobserved endogeneity. The findings and implications are discussed, highlighting the importance of considering trip frequencies by purpose to accurately estimate the travel demand model in DACs.

Chapter 6 explores the adoption of alternative fuel vehicles (AFVs), leading to decarbonization, in DACs by applying statistical and explainable artificial intelligence (XAI) techniques to understand the factors associated with AFV adoption in these communities. The study harnesses a unique and comprehensive database of surveys and public databases collected during COVID-19. The XAI techniques, specifically the Extreme Gradient Boosting (XGBoost) algorithm with Shapely Additive Explanations (SHAP), provide interpretable and understandable explanations of factors associated with AFV adoption in DACs. The study findings provide an understanding of the social and economic factors and challenges of DACs. This contributes to the literature on AFV adoption and suggests opportunities for improvements in DACs transitioning to AFVs. The results and infrastructure-level implications are discussed.

Overall, the main contribution of this dissertation is to understand the changes the pandemic brought to road safety and travel behavior and look at the future by harnessing unique and comprehensive databases. This can help advance knowledge of how systemwide shocks or other major disruptions affect transportation systems and contribute to improving the system overall. This is critical to managing the public health crisis, ensuring road safety, and informing planning and policy decisions. Importantly, the findings of this dissertation underscore the importance of better preparedness and planning for underserved communities to be equipped to handle future systemic shocks. It is crucial to guide policy and planning decisions to promote safer and more equitable traffic safety programs in DACs in the future. Improving the digital infrastructure in DACs is essential for bridging the digital divide. Additionally, the study is critical for promoting electrification in DACs and supporting the ongoing decarbonization efforts.

The key analyses conducted in this dissertation are expected to result in a minimum of five publications in prominent transportation journals as follows:

1. Patwary, A. L., & Khattak, A. J. *Interaction Between Information and Communication Technologies and Travel Behavior: Using Behavioral Data to Explore Correlates of the COVID-19 Pandemic*
    - Peer-review conference paper: **Presented** at the 101st Transportation Research Board Annual Meeting 2022, Washington D.C.
    - Journal article: **Published** in *Transportation Research Record*, 2022. 03611981221116626.
2. Patwary, A. L., & Khattak, A. J. *Crash harm before and during the COVID-19 pandemic: Evidence for spatial heterogeneity in Tennessee*.
    - Peer-review conference paper: **Presented** at the 101st Transportation Research Board Annual Meeting 2022, Washington D.C.
    - Journal article: **Published** in *Accident Analysis & Prevention*, 2023. 106988.
3. Patwary, A.L., Haque A.M., Mahdinia, I., Khattak, A.J. *Investigating Transportation Safety in Disadvantaged Communities by Integrating Crash and Environmental Justice Data*
    - Peer-review conference paper: **Presented** at the 102nd Transportation Research Board Annual Meeting 2023, Washington D.C.
    - Journal article: **Published** in *Accident Analysis & Prevention*
4. Patwary, A.L., Khattak, A.J. *Interaction between the Emerging Components of Online Shopping and In-Person Activities: Insights from Behavioral Survey and Justice40 Initiative Data*
    - Journal article: **Under Review** for Publication
5. Patwary, A.L., Khattak, A.J. *Explainable Artificial Intelligence for Decarbonization: Alternative Fuel Vehicle Adoption in Disadvantaged Communities*
    - Peer-review conference paper: **Presented** at the 102nd Transportation Research Board Annual Meeting 2023, Washington D.C.
    - Journal article: **Accepted for publication** in the *International Journal of Sustainable Transportation*

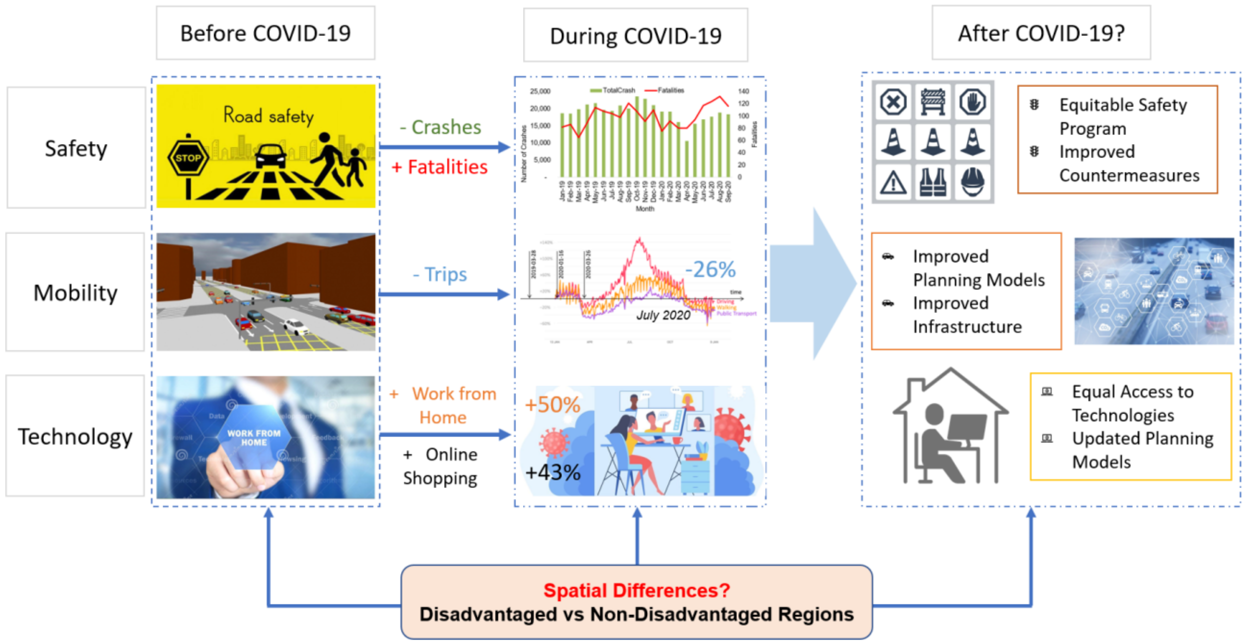


Figure 1-1: Graphical Abstract of the Dissertation

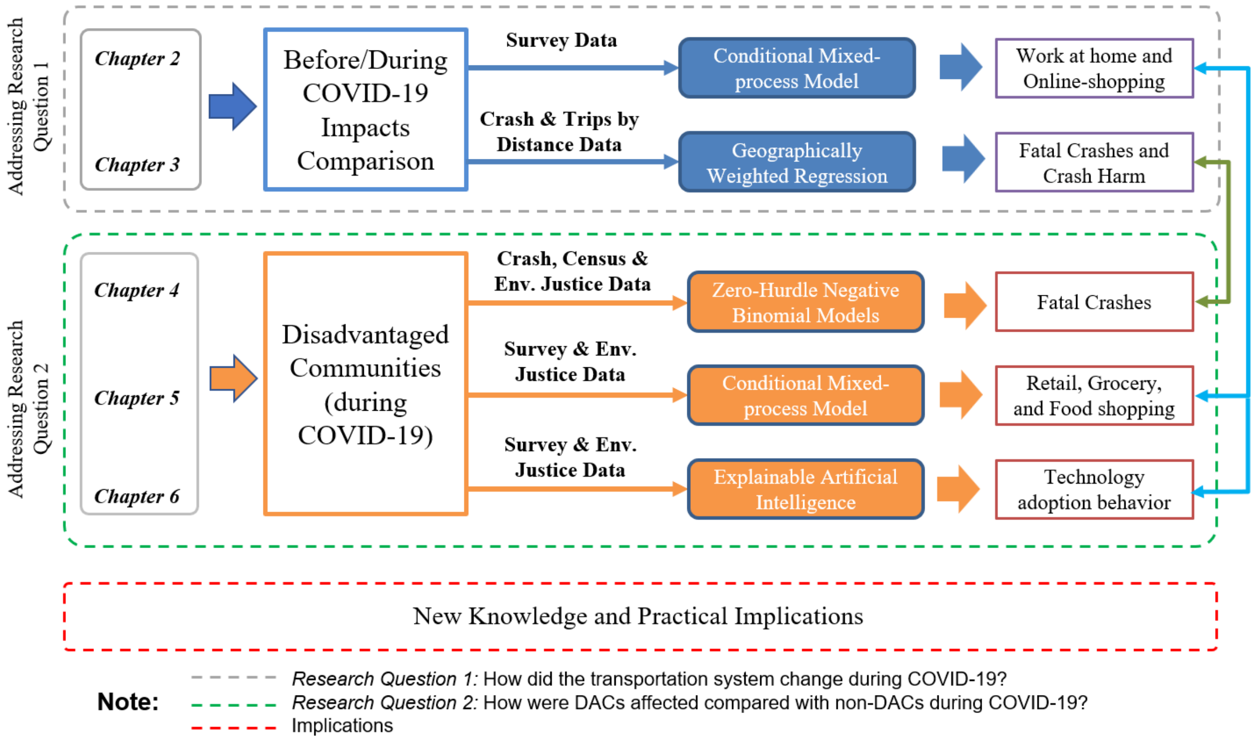


Figure 1-2: Overall Framework of the Dissertation

	Before	COVID-19	During	After?
Chapter 2: Analyzes WFH and online shopping together to find their impacts on shopping trip generation before and during COVID-19	O_1	X	O_2	
Chapter 3: Provides a comprehensive analysis of safety-critical events on roadways during COVID-19	O_1	X	O_2	
Chapter 4: Investigates transportation safety in Disadvantaged Communities		X	O_2' O_2''	
Chapter 5: Examines the existence of a digital divide in DACs by analyzing the emerging components of online shopping		X	O_2' O_2''	
Chapter 6: Explores the adoption of alternative fuel vehicles (AFVs), leading to decarbonization, in DACs		X	O_2' O_2''	

Note:
O_1 = Before-COVID-19 Outputs
O_2 = During-COVID-19 Outputs
O_2' = During-COVID-19 Outputs for DAC
O_2'' = During-COVID-19 Outputs for Non-DAC

Figure 1-3: Overview of the Study Design

CHAPTER 2 INTERACTION BETWEEN INFORMATION AND COMMUNICATION TECHNOLOGIES AND TRAVEL BEHAVIOR: USING BEHAVIORAL DATA TO EXPLORE CORRELATES OF THE COVID-19 PANDEMIC

A version of this chapter was originally published by A. Latif Patwary and Asad J. Khattak in the Journal of Transportation Research Record:

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.* Transportation Research Record, 03611981221116626

2.1 Abstract

The COVID-19 Pandemic has highlighted the importance of information and communication technologies (ICTs), providing virtual engagement. The question for planners and engineers is whether cities will see reductions in travel demand, given the increasing use of ICT technologies. Notably, ICTs facilitate online shopping and working-from-home (WFH). Generally, online shopping can lead to fewer shopping trips; similarly, WFH may reduce work-related trips. However, more WFH has the potential to generate other non-work trips, including shopping trips. To find answers and explore interdependencies, this study integrates pre-pandemic behavioral data with during-pandemic travel data. In our framework, WFH and online shopping are considered together. By harnessing the pre-pandemic 2017 National Household Travel Survey data, this study jointly analyzes the relationships among shopping trips, online shopping, and WFH with a conditional mixed process model that can address unobserved endogeneity and selection bias. The results suggest that pre-pandemic online shopping was associated with lower in-person shopping trips. Furthermore, WFH was associated with more shopping trips. The role of socio-demographic, locational, and travel-related factors is also explored. The during-pandemic data and analysis capture how COVID-19 impacts travel behavior. The results show that the relationships among the key variables found in pre-pandemic data are similar but differ in magnitude from the during-pandemic periods. WFH went up from 12% to 61% during COVID-19, admittedly an unusual situation. In the next new normal, planners may improve travel demand models by treating WFH explicitly as an alternative to traveling to work in the trip generation and time-of-day models.

2.2 Introduction

The world is changing through the advancement of information and communication technology (ICT). With affordable ICT appliances and widely available internet, many activities that needed a fixed location and time are no longer as bounded as before (*1*). Therefore, partial decoupling of virtual and physical activity spaces is expected (*2*). These changes challenge the traditional belief that an individual's daily activity is limited to fixed spaces and time (*3*). Virtual communication behaviors could substitute activities that previously required physical travel, as well as can stimulate more virtual communication (*4*). It can be in the form of e-commerce (e.g., online shopping) and teleworking (e.g., working from home), among others.

The COVID-19 Pandemic and the ever-changing ICT landscape have further prompted changes in daily activity travel. During the Pandemic, ICT has played a crucial role in fulfilling daily necessities through working-from-home (WFH) and online shopping. The E-commerce surge triggers more and more individuals to buy from online stores. In the United States (US), e-commerce sales grew by 17.3% in 2019, and sales escalated by 36.7% in the 3rd quarter of 2020. E-commerce comprised up to 14.3% of total retail sales, and this growth is expected to continue in the coming years (*5*). Hence, the surge in online shopping has the potential to affect travel behavior. Moreover, ICT has given the option to work virtually, leading to change in the workforce and economy. It enables companies to permit workers to work from remote locations, which can bring significant changes in workers' travel behavior. It is estimated that in 2017, 5.2% of people were engaged in full-time remote working (*6*); however, this number increased to 51% in April and 33% in September of 2020 (*7*). These new routines and alternate activities are also expected to be continued in the post-COVID-19 periods.

Presumably, WFH generates numerous benefits, including reduced congestion, emissions, office space saving costs, and flexibility (*8*). Moreover, policymakers are promoting WFH to substitute the commute for work (*9*). They stand on the traditional belief that WFH and online shopping might reduce physical shopping trips. However, the relationships among these activities do not provide any clear-cut results as expected. Does ICT usage substitute travel, especially in uncertainties like COVID-19, or can these incite new opportunities for people to engage in other travel-required activities? For example, Cao et al. (*10*) discovered that online shopping could increase shopping trips instead of reducing them. Besides, WFH can intervene in this relationship, as it has the potential to generate other non-work trips like shopping trips. Moreover, WFH and online shopping can also be influenced by socio-demographic behavior, location, and travel-specific factors. However, empirical evidence indicates no significant prior work on the integration of these activities. It is warranted to discover the potential relationships among the mentioned ICT uses together and their overall policy-driven implications, as understanding these relationships is important for considering and planning for future travel behavior patterns.

Therefore, this study aims to analyze WFH and online shopping together to find their impacts on the shopping trip generation. It further explores whether the relationships are similar or not before and during the COVID-19 pandemic. The study also aims to discover the exogenous factors that may encourage people to work from home, shop online, and make shopping trips. The structure of this paper is as follows: a literature review, which includes the previous works based on which the objectives were formulated, a conceptual framework, the methodology used, the results, and a discussion of the analysis.

2.3 Literature Review

ICT can interact with travel in four different ways: substitution, modification, neutrality, and complementarity (*11*). In general, ICT use can reduce travel costs and time (substitute). Nevertheless, these reductions may be used for participating in other activities where travel might be needed (complement). This rebound eventually offsets the initial travel reduction gains (*11-13*). A major share of research has focused on online shopping and WFH among all the ICT uses (*14*). Online shopping (also known as e-shopping) has experienced substantial growth over the years. Moreover, online sales soared amid the COVID-19 pandemic. This surge certainly can affect shopping trip generation, which accounts for nearly 20% of all US trips in 2017 (*15*). Therefore, online shopping is of interest to many engineers and planners. A long-lasting debate is going on: whether the relationship between online shopping and shopping trips is substitutive or complementary. Several studies have explored this relationship and found a complementary effect, implying the increase in overall trips due to increased online shopping frequency (*10; 16*). Wilson et al. (*17*) surveyed people in three cities in the United States and found that e-shopping for the last purchase replaced 79% of shopping trips. They also found that 55.5% of the shoppers made new trips after obtaining information online. Other studies found the relationship to be a substitution (*18-20*). People substitute online shopping time for travel time, leading online shoppers to take fewer trips and travel to nearby places for shopping purposes. Moreover, online shopping has the potential to reduce longer shopping trips (*19*). Some studies found both substitution and complementary effects on travel. For example, Weltevreden and Rietbergen (*21*) found that more than 20% of online buyers made fewer trips to city center stores in the long run, whereas they also found a complementary effect in the short run. In addition, Tonn & Hemrick (*22*) noticed that certain internet users reduced trips to stores, with a minor percentage of them making new trips in Knoxville, US.

WFH (also known as teleworking or telecommuting) helps to achieve travel reduction (substitute) and urban sustainability goals based on policy implementation (*23; 24*), while it is also found that the impact might be smaller than expected or complementary (*25; 26*). WFH exerts complex travel substitution rather than the common assumption of its reduction to overall travel demand. It can promote more dispersed, decentralized, and car-dependent patterns of working while reducing congestion (*27*). Importantly, WFH provides workplace and work time flexibility that may ease peak-hour congestion (*28*). There also might be an overall increase in mobility in terms of personal travel attributes. For example, it is observed that people with higher education levels are more likely to work from home and take fewer commute trips but more insignificant trips (*20*). Silva et al. (*29*) suggested that WFH is an approach used to get through long and costly travel, especially for workers living in remote areas. Overall, while the benefits of WFH in terms of sustainability, flexibility, or reducing commute trips have been explored, the potential for WFH to promote alternate activities like more online shopping or making shopping trips is less explored in previous studies. Furthermore, the COVID-19 pandemic may intervene in these relationships, as suggested by the recent literature on COVID-19 and ICT uses (*30-33*). Therefore, it is necessary to investigate the effects of COVID-19 as well.

Online shopping and WFH are largely influenced by socio-demographic characteristics, regional features, and travel attributes (*10; 16; 18; 21; 34*). Higher education and urban locations tend to increase online shopping frequency. Internet use and e-shopping are largely urban phenomena (*10*). People living in urban areas buy online more frequently than those in rural areas, which reveals a complementary effect between traveling and buying online (*34*). Higher household income, shopping attitude, and full-time employment (FTE) have both positive and negative effects on online shopping. For example, Zhou and Wang (*16*) & Cao et al. (*35*) found a positive impact on household income, whereas Farag et al. (*34*) found a negative result on online shopping. A summary of the impacts of some key variables is presented in

Table 2-1.

This research contributes to the literature in two ways while filling the gaps. First, we investigate the ICT uses (i.e., WFH and online shopping) in terms of their association with shopping trip generation while controlling for other exogenous factors using pre-pandemic data in a joint estimation framework. Second, using the same framework, we examine whether the relationships among WFH, online shopping, and shopping trip generation are similar or not during the pandemic.

2.4 Conceptual Framework and Hypothesis

The interaction between online shopping and shopping trip generation is not straightforward.

Table 2-1: Summary of Relevant Literature

Author	Study Approach		Findings	Some exogenous variables and their impacts on Online Shopping/ Telework			
Name and Year	***Location and Data***	***Analysis Method***	***Overall Effects on Travel***	***Household Income***	***Education***	***Urban Location***	***FTE***
Zhou and Wang, 2014	US, N= 85,663, 2009 NHTS	Structural Equation Modelling (SEM)	*Complementary*	+	+	+	+
Ferrell, 2005	San Francisco, N= 14563, Area Travel Survey for the year 2000	Two-stage Least squares (2SLS) Regression	*Substitution*	+	NA	+	+
Cao et al., 2012	Minneapolis, N=539, Online survey	SEM	*Complementary*	+	+	+	+
Farag et al., 2006	Netherlands, N= 2190, Online Survey	SEM	*Complementary*	-	NA	+	NA
Weltvreden and Rietbergen, 2007	Netherlands, N= 3200, online survey	Multinomial Logit	*Complementary, Substitution*	NA	+	+	NA
Tonn and Hemrick, 2004	Knoxville, N= 118, Web Survey	Trip Generation model (Regression)	*Complementary, Substitution*	+	+	NA	-
Loo and Wang, 2018	China, N= 608 FTE employees, Household Survey	Logit Regression	*Substitution*	+	+/-	NA	NA

Note: "+" and "-" denote positive and negative impacts, respectively.

Previous studies suggest that this relationship can be either complementary or substitutive. Besides, WFH can also intervene in this relationship since WFH itself can be influenced by the same exogenous features (e.g., socio-demographics) for defining online shopping and shopping trip association. On the other hand, recent developments in e-commerce suggest that online shopping can take many new and different forms, e.g., online shopping for durable goods or groceries. These components of online shopping can be interconnected with each other and also with physical shopping trips (*36*). However, such nuances could not be explored because of data limitations, i.e., these components were aggregated in a broad category of online shopping. To analyze the overall potential impacts, we identify the number of shopping trips, online shopping, and WFH as the outcome variables, which may be influenced by some specific person, household, location, and travel-related variables (Figure 2-1).

We anticipate that the number of shopping trips will decrease with more online shopping. As the products are delivered to the door, there is less need for physical travel for shopping (*18*). WFH may stimulate shopping trips. While people work from home, it is expected that they will desire to make other non-work trips, e.g., shopping trips, as opposed to shopping online. Nonetheless, they can also be encouraged to buy online, as they get more free time by not commuting to work. We can expect that a higher-aged individual with higher education is likely to make more physical shopping trips (*35*). However, they may also be more comfortable with WFH (*8*). In contrast, young people and females may spend more time online searching for discounts or deals and will, therefore, prefer online shopping more than shopping in person.

People with higher incomes are expected to shop more in-store and online than lower-income people (*16; 20; 34*). People living in urban areas have higher accessibility to newer technologies. Therefore, they may shop more online instead of in stores. However, they may be less interested in working from home, as most workplaces are in urban areas. Hence, travel time should be lower for those from urban areas as compared to people living in rural areas (*16; 34*). Higher travel time may also play a role in more online shopping, fewer shopping trips, and WFH. Higher gas prices may also be linked to this association (*16*). Travel day may influence the shopping trip generation as well. If people go out on weekends, it is generally anticipated that they may go shopping as well (*16*).

2.5 Methodology

2.5.1 Data

Pre-pandemic data for this study is collected from the 2017 US National Household Travel Survey (NHTS) (*15*). The survey covered 129,696 households, including 264,234 individuals over all the US states and the District of Columbia (DC). The survey produces four data files which are on household, person, vehicle, and trips. These files are merged, summarized, and averaged to generate person-specific variables. At first, samples are sorted out based on the person-specific shopping trips on each travel day. Then, the individuals who are less than 18 years old are dropped. Observations with missing values are also discarded. A total of 108,297 observations (N) finally remain after dropping the invalid observations. The cleaned dataset contains three endogenous variables and four types of exogenous variables: person, household, location, and travel pattern. Overall, the NHTS data is carefully collected, and error checking is performed using descriptive analysis (Table 2-2 and Table 2-3).

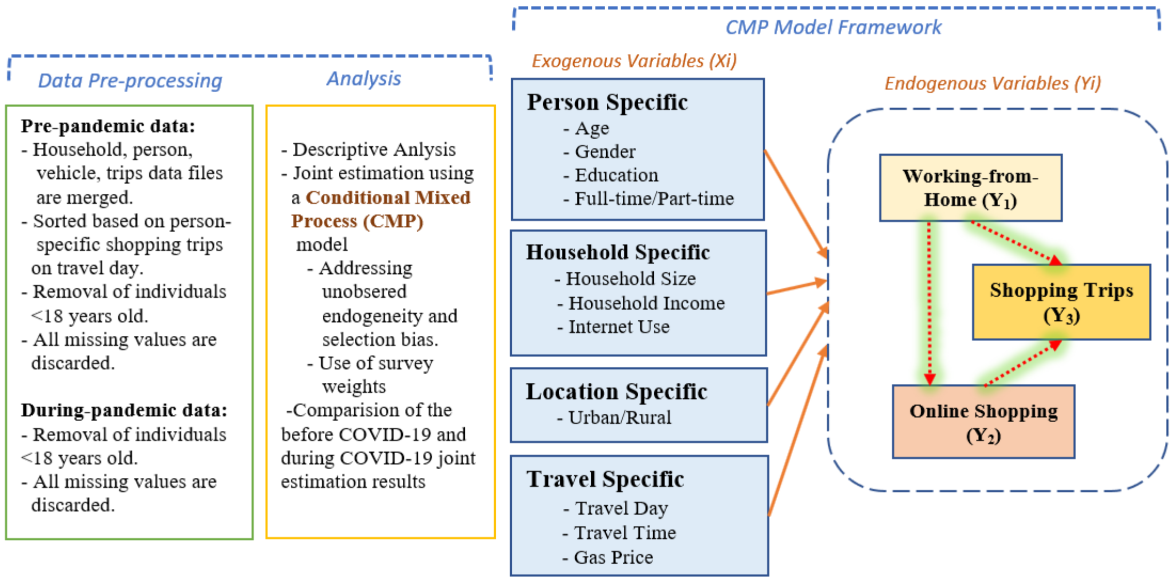


Figure 2-1: Study Framework

Shopping trips, online shopping, and WFH are the endogenous variables. It can be observed from the descriptive statistics that 59% of the sample do not make any shopping trips, whereas 34% make 1-2 trips, 6% make 3-4 trips, and 1% make more than four trips on a travel day. A travel day is specified in the NHTS user guide as a day that starts from 4:00 AM of one day until 3.59 AM of the next day (*15*). In the sample, 79% of travel days are on weekdays. It is seen that 83% of the sample shop 0-5 times online in a month. 12% of the respondents in the sample work from home. It is understood from the NHTS survey that hybrid workers (i.e., workers who work both from the office and home) fall into the "Yes" response category of WFH for working a few days a week from home. 51% are male, and 25% have a graduate or professional degree. Of the sample, 80% were full-time employees. Most individuals are from urban areas and affluent families (i.e., 74% of households have income greater than $50,000). It is not surprising that 96% of people sampled use the internet almost daily.

During-pandemic data is collected from the US census bureau, which has introduced an experimental household pulse survey to collect data to quickly and effectively capture how the ongoing COVID-19 pandemic is changing people's travel behavior (*37*). The survey has been collecting weekly data from all over the US. This study uses person-level data from week 23 (i.e., January 20th – February 1st, 2021) from the survey. The data are cleaned and checked for errors. The final dataset has a sample size of 69,905. In the sample, 60% are male, 30% have bachelor's degrees, and 60% are from households with an income of more than 50,000 (Table 2-2 and Table 2-3). The data shows that 61% of people aged 18 or older live in households where at least one person telework to substitute the work trips during COVID-19, while, before COVID-19, only 12% of people worked from home. However, a total of 70% of people took fewer trips to the store, and 53% of people made more online purchases in the last seven days of the data collection period due to COVID-19. Notably, the recent developments during COVID-19 highlight online grocery shopping as crucial as online shopping for other household and personal items. In our pre-COVID-19 and during-COVID-19 databases, online shopping is identified as a broad category, including grocery shopping and shopping for other personal or household items. The substitution vs. complementarity framework is flexible enough to incorporate these new developments.

2.5.2 Model

In this paper, for the pre-pandemic data, we define shopping trips as a categorical variable, where the frequency of the trips is ordered, ranging from "no trips" to ">4 trips". Online shopping is also defined as a categorical ordered variable, including "0-5 times" making an online purchase as the lowest category and ">15 times" as the highest category. WFH is a binary variable. On the other hand, WFH, online shopping, and shopping trips are all binary variables for the during-pandemic data. If we denote online shopping as OS_i, and shopping trips as ST_i, the following empirical models can be estimated:

$$WFH_i = \alpha_0 + \alpha_1 X_i + \mu_{1i} \qquad \text{(i)}$$

$$OS_i = \beta_0 + \beta_1 X_i + \delta WFH_i + \mu_{2i} \qquad \text{(ii)}$$

$$ST_i = \gamma_0 + \gamma_1 X_i + \varphi OS_i + \tau WFH_i + \mu_{3i} \qquad \text{(iii)}$$

Where X_i signifies a vector of the explanatory variables associated with the individual which are hypothesized to influence WFH, online shopping, and making shopping trips, μ_1, μ_2, and μ_3 are the corresponding random error terms, and α, β, γ, δ, τ and φ are the parameters to be estimated.

Table 2-2: Descriptive Statistics of Categorical Variables

Pre-pandemic Data (N=108,297)			
Variables	**Description**	**Frequency**	**%**
Shopping Trips*	Number of shopping trips on travel day		
	No Trips	60,737	59%
	1-2 Trips	34,888	34%
	3-4 Trips	5,871	6%
	>4 Trips	1,198	1%
Online shopping*	Number of Online Purchases in Past Month		
	0-5 times	90,240	83%
	6-10 times	12,454	11%
	11-15 times	4,815	4%
	>15 times	788	1%
WFH*	Yes	13,197	12%
	No	95,100	88%
Person Specific			
Age	1= "18-30"	17,893	17%
	0= ">30"	90,404	83%
Gender	1= Male	55,210	51%
	0= Female	53,087	49%
Education	1= Graduate or professional degree	27,174	25%
	2= Bachelor's degree	30,928	29%
	3= Some college or less	50,195	46%
Work Status	1= Full-time	86,243	80%
	0= Part-time	22,054	20%
Household Specific			
Household Income	1= Household income >= 50,000	80,646	74%
	0= Household income < 50,000	24,651	26%
Internet Use	1= Once a week or less	2,283	2%
	2= Several times a week	2,524	2%
	3= Daily	103,490	96%
Location Specific			
Urban	1= Urban	86,025	79%
	0= Rural	22,272	21%
Travel Pattern Specific			
Travel Day	1= Weekend	22,303	21%
	0= Weekdays	85,994	79%
During-pandemic Data (N=69,905)			
Variables	**Description**	**Frequency**	**%**
Shopping Trips*	Whether took fewer trips to the store in last 7 days during COVID-19		
	Yes	49,214	70%
	No	20,691	30%
Online Shopping*	Whether made more purchases online in last 7 days during COVID-19		
	Yes	37,196	53%

**: Endogenous Variable.* *Source: (NHTS, 2017; Census, 2021)*

Table 2-2 continued

	No	32,709	47%
WFH*	Whether at least 1 person in a household telework during COVID-19		
	Yes	42,589	61%
	No	27,316	39%
Age	1= "18-30"	6,479	10%
	0= ">30"	63,426	90%
Gender	1= Male	41,895	60%
	0= Female	28,010	40%
Education	1= Graduate or professional degree	18,299	26%
	2= Bachelor's degree	20,803	30%
	3= Some college or less	30,803	44%
Household Income	1= Household income >= 50,000	42,612	61%
	0= Household income < 50,000	27,293	39%

**: Endogenous Variable.* *Source: (NHTS, 2017; Census, 2021)*

Table 2-3: Descriptive Statistics of Continuous Variables

Pre-Pandemic Data (N=108,297)					
Variables	**Description**	**Min.**	**Mean**	**Max.**	**SD**
Household Size	Count of Household member	1	2.65	13	1.29
Travel Time	Avg. travel time (minute) per trip	0.5	26.26	1200	34.59
Gas Price	USD per gallon	2.01	2.4	2.96	0.23
During-pandemic Data (N=69,905)					
Household Size	Count of Household member	1	2.74	10	1.46

Source: (NHTS, 2017; Census, 2021)

The parameters δ, φ, and τ are the estimate of WFH on online shopping, the estimate of online shopping on shopping trips, and the estimate of WFH on shopping trips, respectively.

The estimation of the above three equations (i)-(iii) can be performed using conventional path analysis or structural equation modeling. The individual models may produce biased and unreliable estimates due to a potential issue of selection and unobserved endogeneity. To overcome these concerns, we can propose a framework that can jointly estimate the equations. Conventional models become inapplicable if mixed equations are used (*38*). The mentioned three equations are mixed for the pre-pandemic data; Equation (ii) and (iii) are ordered probit models, and equation (i) is a binary model. Whereas these three equations are binary models for the during-pandemic data. The conditional mixed process model (CMP) developed by Roodman (*39*) provides us with a unique opportunity to employ this mixed structural model with different equations while correcting bias and unobserved endogeneity. CMP jointly estimates two or more equations with associations among their error processes. These are individual equations with correlated errors. CMP can handle all types of dependent variables, e.g., binary and ordered (*39*).

The equations are restructured into the following in the CMP format:

$$y_1^* = \sigma_1 + \mu_1 \qquad \text{(iv)}$$
$$y_2^* = \sigma_2 + \mu_2 \qquad \text{(v)}$$
$$y_3^* = \sigma_3 + \mu_3 \qquad \text{(vi)}$$

Where,

$$\sigma_1 = \alpha_1 X_i, \sigma_2 = \beta_1 X + \delta y_1, \sigma_3 = \gamma_1 X + \varphi y_2 + \tau y_1$$
$$y = g(y^*) = (1\{|y_1^* > 0\}, y_2^*, 1\{y_3^* > 0\})'| \qquad \text{(vii)}$$
$$\mu = (\mu_1, \mu_2, \mu_3)' \sim N(0, \Sigma) \text{ and } \Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$

Here, $y_1^*, y_2^*, and\ y_3^*$ are latent factors for WFH, online shopping, and shopping trips, respectively. The terms ρ_{12}, ρ_{13}, and ρ_{23}, respectively, are the correlation between the error terms of WFH and online shopping, WFH and shopping trips, and online shopping and shopping trips. Presuming that $y_i = (0, y_{i2}, 0)'$ is observed, a consequent likelihood function can be denoted as follows:

$$L_i(\alpha_1, \beta_1, \gamma_1, \delta, \varphi, \tau, \Sigma;\ y_i|x_i) = \int_{-\infty}^{-\sigma_1} \int_{-\infty}^{-\sigma_2} \int_{-\infty}^{-\sigma_3} \emptyset_j\{\mu_1, y_{i2} - \sigma_{i2}, \mu_3)'; \Sigma\} d\mu_1 d\mu_2 d\mu_3 \qquad \text{(viii)}$$

This modeling framework is applied to produce results for both pre-pandemic and during-pandemic data. Survey estimation design is employed in the CMP modeling. The analysis is performed using the statistical software "STATA" version 17. STATA's sampling weight option (*pweights*) is adopted to generate justifiable population-level estimates for the survey data. Besides, direct marginal effects are produced in the CMP post-estimation.

2.6 Results and Discussion

2.6.1 Pre-pandemic

The results of the CMP modeling approach are reported in Table 2-4. Columns 1, 2, and 3 show the estimates for WFH, online shopping, and shopping trips, respectively. The model significance test indicates that the model fits the data well. We first discuss the results to explore the relationships among the endogenous variables. Then, we expound on the effect of exogenous variables on each of the endogenous variables in the model. The signs of expected and observed findings are summarized in Table 2-5.

The results indicate that structurally, WFH positively affects online shopping. Specifically, compared to not WFH, WFH is associated with being 18.5% less likely to shop online 0-5 times but 9.1% more likely to shop online 6-10 times, 7.2% more likely to shop online 11-15 times, and 2.1% more likely to shop online for >15 times. Online shopping is negatively associated with shopping trips (from the lower category "no trips" to the higher category ">4 trips"). The direct marginal effect of online shopping in the shopping trips model is negative for making shopping trips. For example, compared to the base of online shopping (0-5 times), 6-10 times online shopping is associated with being 8.5% less likely to make 1-2 trips, 2.9% less likely to make 3-4 trips, and 1% less likely to make more than 4 trips. Similarly, the effects are increased for the remaining categories of online shopping. WFH is positively associated with shopping trips (i.e., from making "no trips" to ">4 trips"). People who work from home are 13.3% more likely to make 1-2 trips, 9% to make 3-4 trips, and 4% to make >4 trips. The findings suggest that online shopping reduces shopping trips, which is consistent with our earlier assumption. However, this claim is in contrast with Cao et al. (*35*), who found a positive association between online purchase frequency and physical shopping trips. Mostly, people prefer online shopping instead of making shopping trips because of the convenience of online shopping. The products are delivered to the doorsteps. Online shopping has many benefits, e.g., people can shop any time they want and compare prices from different online stores. Thus, comparing with the prior studies on whether the association between online shopping and shopping trips is a complementary or substitution effect, our study suggests that the relationship is substitution. However, we should acknowledge that shopping is usually comprised of several stages that may be completely or partially completed online. In-store purchases can be associated with online searches. For example, people can make trips to stores to compare or experience the actual goods they browse online. In addition, shopping trips are influenced by WFH, as we found that it stimulates shopping trips. While people are working-from-home, they may desire to make other non-work and/or insignificant trips, and shopping trips are one of them.

We shift our discussion to other factors that influence people to work from home, shop online, and make shopping trips. It is observed that the probability of young people aged 18-30 buying online for the 0-5 times category is 2.5% lower than for the older people; however, this age group has a higher probability of buying online for other higher purchase categories. They are also 8.6% more inclined to make no shopping trips as compared to the older population. As we expected, higher-aged people prefer buying online less and making shopping trips more than younger ones. The findings are aligned with Zhou and Wang (*16*) and Cao et al. (*35*). However, older people are 7% more likely to work from home than younger ones. It is also observed that a male is more likely to work from home compared to a female. Males are less likely to shop online for 6-10, 11-15, & >15 categories than females, whereas males are more likely to purchase for lower "0-5" times category. As we assumed earlier, males prefer less online shopping as well as fewer shopping trips than females, which is also consistent with the results of Ferrell (*18*). Young people and females may spend more time online searching for discounts or deals; hence, they prefer to shop online rather than in person.

Our results further suggest that people with bachelor's degrees are 1.5% more likely, and people with some college degrees are 4.3% less likely to work from home than people with graduate or professional degrees. As we anticipated, higher education attainment motivates someone to work from home. Drucker & Khattak (*8*), among others, discovered similar findings (*29*). Higher education attainment is also associated with more online shopping. Besides, higher household income (>=$50,000) is associated with more online purchases for 6-10, 11-15, and >15 categories, whereas lower household income is more associated only with the 0-5 times online purchase category. This claim is supported by Farag et al. (*34*). Compared to those who use the internet once a week or less, using the internet daily increases the probability of buying online for 6-10, 11-15, and >15 categories. An increase in travel time and gas price increases the probability of WFH. An increase in gas price is associated with more online shopping for 6-10, 11-15, and >15 times categories and less for the 0-5 times category. As we anticipated earlier, higher travel times and higher gas prices discourage people from making shopping trips, and instead, they shop online and prefer to work from home. Daily internet use contributes to this association. This is also supported by Zhou and Wang (*16*).

People are 3% less likely to telework in urban areas compared to rural areas. In urban areas, people are less likely to shop online for the 0-5 category than in rural areas, whereas they are more likely to shop online for other higher categories. As we expected, people living in urban areas are more reluctant to work from home compared to people from rural areas. However, they buy online more times compared to people from rural areas, which aligns with our earlier assumptions. Farag et al. (*34*) discovered similar findings in their study. This may be because shopping accessibility and internet use in rural areas are limited. Hence, people from rural areas are less encouraged to buy online than those in urban areas.

2.6.2 During-pandemic

The study attempts to shed light on the relationships among the key variables during the COVID-19 pandemic. As COVID-19 is affecting the world economy, including the transportation sector, travel behavior has changed substantially. Commuting to work is a vital element of local travel and is related to many other aspects of local transportation. COVID-19 pushes people to telework as a way of substituting physical travel for work. Activities like trips to stores, which are a vital element of local travel, are also being substituted by online shopping.

The CMP joint estimation results for the during-pandemic data are presented in Table 2-6. The model significance test suggests that the model fits the data well. The results are examined to explore the relationships among WFH, shopping trips, and online shopping during COVID-19 and compare them with the pre-COVID-19 results. Firstly, the results reveal that WFH is negatively associated with purchasing more online, which is consistent with the 0-5 times online purchases in a month but contradicts the higher categories of purchases (more than 5 times in a month) in pre-COVID-19 findings. However, this relationship is not statistically significant for the during-COVID-19 data. Secondly, the findings suggest that WFH is negatively associated with the making of fewer shopping trips, which indicates that compared to not WFH, WFH is likely to increase the probability of making shopping trips by 14.2% during COVID-19. This finding is consistent with the pre-pandemic periods (29.6%), although the magnitude is 52% lower during COVID-19. Thirdly, online shopping is positively associated with fewer shopping trips; hence, it is more likely to reduce shopping trips. More specifically, online shopping lowers the probability of making shopping trips by 20.7% during COVID-19. This relationship is also aligned with the pre-pandemic findings. Due to the pandemic and its associated stay-at-home orders, more people are

Table 2-4: CMP Joint Estimation Results for Before COVID-19 Data (N= 108,297)

Variables	(1) WFH		(2) Online Shopping					(3) Shopping Trips				
	Coef.	Mar. Effect	Coef.	Marginal Effect				Coef.	Marginal Effect			
				0-5 times	6-10 times	11-15 times	>15 times		No Trips	1-2 Trips	3-4 Trips	>4 Trips
WFH (Base: No), Yes			**0.66*****	**-0.18**	**0.09**	**0.07**	**0.02**	**0.81*****	**-0.30**	**0.15**	**0.10**	**0.05**
Online Shopping (Base: 0-5 times)												
6-10 times								**-0.36*****	**0.12**	**-0.08**	**-0.03**	**-0.01**
11-15 times								**-0.59*****	**0.19**	**-0.14**	**-0.04**	**-0.01**
>15 times								**-1.03*****	**0.29**	**-0.22**	**-0.05**	**-0.02**
Person Specific												
Age (Base: >30), "18-30"	-0.49***	-0.07	0.11***	-0.03	0.02	0.01	0.00	-0.25***	0.09	-0.06	-0.02	-0.01
Gender (Base: Female), Male	0.05**	0.01	-0.18***	0.04	-0.02	-0.01	-0.01	-0.17***	0.06	-0.04	-0.02	-0.00
Education (Base: graduate or prof. degree)												
Bachelor's	0.07**	0.02	-0.13***	0.03	-0.02	-0.01	-0.00	-0.01	0.01	-0.01	-0.00	-0.00
Some college or less	-0.25***	-0.04	-0.33***	0.08	-0.04	-0.03	-0.01	-0.09***	0.03	-0.02	-0.01	-0.00
Work Status (Base: Part-time), Fulltime	-0.59***	-0.13	0.18***	-0.04	0.02	0.01	0.01	-0.01	0.03	-0.02	-0.01	0.00
Household Specific												
Household Size	-0.03***	-0.01										
Household Income (Base: <50,000), >=50,000	0.01	0.002	0.37***	-0.08	0.05	0.03	0.00	-0.11***	0.03	-0.02	-0.01	-0.00
Internet (Base: Once a week or less)												
Daily	0.002	0.00	0.93***	-0.13	0.08	0.04	0.01	-0.04	0.01	-0.01	-0.00	-0.00

Table 2-4 continued

Several Times a Week	-0.26**	-0.04	0.28**	-0.02	0.02	0.00	0.00	-0.02	0.01	-0.01	-0.00	-0.00
Location Specific												
Urban (Base: No), Yes	-0.16***	-0.01	-0.07**	-0.02	0.01	0.01	0.00	-0.01	0.01	-0.01	0.00	0.00
Travel Pattern Specific												
Travel Day (Base: Weekdays), Weekend								0.43***	-0.12	0.09	0.04	0.02
Avg. Travel Time	0.01***	0.00						-0.01***	0.01	-0.01	-0.00	-0.00
Gas Price	0.23***	0.04	0.19***	-0.04	0.02	0.02	0.00	-0.10***	0.04	-0.02	-0.01	-0.01
Constant	-0.99**											
Model Fit Statistics												
Number of observations	108,297.00											
Wald chi2 (39)	3249.24											
Model Significance Test	0.00											
Log Pseudolikelihood	-221400.85											
AIC	443925.70											
BIC	4438395.70											

**** p<.01, ** p<.05, * p<.1*

Table 2-5: Expected and Observed Signs of the Exogenous Variables

Exogenous Variables	Endogenous Variables					
	WFH		Online Shopping		Shopping Trips	
	Expected	Observed	Expected	Observed	Expected	Observed
Person Specific						
Higher age	+	+	-	-	+	-
Gender (Male)	+	+	-	-	-	-
Education (Bachelors)	+	+	+	-	+	NS
Full Time	-	-	+	+	-	NS
Household Specific						
Household size	+	-	NA	NA	NA	NA
Household Income	+	NS	+	+	+	-
Daily Internet Use	+	NS	+	+	-	-
Location Specific						
Urban	-	-	+	-	-	NS
Travel Pattern Specific						
Average Travel Time Per trip	+	+	NA	NA	-	-
Weekend Travel Day	NA	NA	NA	NA	+	+
Gas Price	+	+	+	+	-	-
Others						
Online Shopping	NA	NA	NA	NA	-	-
WFH	NA	NA	+/-	+/-	+	+

Note: *NA = Not Applicable, NS= Not Statistically Significant* working from home, and many of them buy online rather than making other non-essential trips, e.g., shopping trips. However, during COVID-19, online shopping also introduced several travel components that need to be acknowledged. COVID-19 saw the frequent use of the curbside pickup of goods ordered online, which warrants making physical trips. Moreover, the trips could be made to the courier office, locker facility, or other designated locations where orders are fulfilled through goods delivery.

The factors that influence WFH, online shopping, and shopping trips are also explored during COVID-19. CMP modeling results indicate that younger age people, compared to the old ones, are 10.2% more likely to work from home during COVID-19, whereas they are 7.3% less likely to work from home before COVID-19. It may be because younger people are forced to work at home, as some schools or companies are encouraging WFH during COVID-19. The findings are similar to Tahlyan et al. (*31*), who indicated that older people have lower satisfaction and higher obstacles to WFH than younger middle-aged people during COVID-19. Unlike the pre-pandemic periods, our findings suggest that younger people are 2.9% more likely to make shopping trips. However, they are more likely to shop online, like in the pre-COVID-19 periods, which is supported by recent research on COVID-19 and shopping behavior (*32*). Our results further indicate that compared to females, males are 1.4% more likely to work from home but 4.6% less likely to shop online during COVID-19 - similar to the pre-COVID-19 findings. On the other hand, males are 4.9% more likely to make shopping trips, which contradicts with pre-pandemic periods where males are less likely to buy in person than females. Compared to people with graduate or professional degrees, people with bachelor's degrees are less likely to work from home, shop online, and more likely to make shopping trips. The pre-pandemic findings on online shopping and shopping trips are consistent with the during-pandemic findings; however, they contradict the probability of working from home. An increase in the household size increased the probability of working from home by 0.7% during COVID-19, whereas it decreased by 0.5% before COVID-19. In addition, higher-income households' probability of making more shopping trips increased during COVID-19, whereas it decreased before COVID-19. It has been observed that higher-income households are 12.3% more likely to work from home during COVID-19. They are also more likely to shop online, and the relationship is similar to the pre-COVID-19 period. The results are consistent with Wang et al. (*32*), who found the association between higher-income households and online shopping during COVID-19 is positive and similar to the previous studies based on before-COVID-19 normal conditions.

While individuals work from home during COVID-19, they make more physical shopping trips. WFH increases the desire to do more online shopping, while more online shopping reduces physical shopping trips. People working from home create shopping needs for home offices and more space that may have affected typical shopping habits. The risk of contracting the virus, lockdowns, and the advancement of ICTs are some of the factors encouraging online shopping while people adapt to working from home. Since the pandemic has accelerated the shift to a more advanced digital era, ICT uses such as WFH and online shopping will have a lasting effect in the future with a resurgence of the economy. Moreover, WFH will continue as a hybrid work system with flexible working hours and places with positive unintended consequences for future transportation planning (*30*).

2.7 Limitations

This study ignores the bidirectional relationship among the endogenous outcome variables. Besides, this study only concentrates on the national-level estimates. State-level estimation could

Table 2-6: CMP Joint Estimation Results for During COVID-19 Data (N= 69,905)

Variables	WFH		Online Shopping[#]		Shopping Trips[##]	
	Coef.	**Marginal Effect**	**Coef.**	**Marginal Effect**	**Coef.**	**Marginal Effect**
Working from home			-0.133	-0.052	-0.450*	-0.142
Online Shopping[#]					0.636***	0.207
Age (Base: >30), "18-30"	0.285***	0.102	0.156***	0.061	-0.089**	-0.029
Gender (Base: Female), Male	0.038*	0.014	-0.119***	-0.046	-0.152***	-0.049
Education (Base: graduate or prof. degree)						
Bachelor's	-0.208***	-0.067	-0.113***	-0.044	-0.051*	-0.016
Some college or less	-0.781***	-0.282	-0.320**	-0.125	-0.072	-0.023
Household Size	0.020**	0.007				
Household Income (Base: <50,000), >=50,000	0.331***	0.123	0.263***	0.103	-0.160***	-0.051
Constant	0.435***		0.503		0.751***	
Model Fit Statistics:						
Number of Observations	*69,905*					
F (6, 69899)	*374.21*					
Model Significance Test	*0.00*					
Log Pseudolikelihood	*-3.87e-8*					
AIC	*7.74e8*					
BIC	*7.74e8*					

[#] *Whether made more purchases online in last 7 days during COVID-19,* [##] *Whether took fewer trips to the store in last 7 days during COVID-19*
****** p<.01, ** p<.05, * p<.1***

provide different levels of estimations since travel behavior can vary across the states. In the analysis presented, online shopping is used as a broad category, e.g., which includes grocery shopping and shopping for durable goods. The correlates of online grocery shopping may differ from those of online durable goods shopping, and the purchase intention may work differently for these two components. However, such correlates could not be explored because these are not included in the 2017 NHTS and 2020 household pulse survey databases. Furthermore, other variables regarding the local built information (e.g., transport accessibility, employment generation) that are not part of these databases could not be introduced in the models to measure the spatial variation of the impacts on the outcome variables.

2.8 Conclusions

Are the relationships between ICT uses and travel behavior similar before and during the COVID-19 pandemic? To answer this question, the paper focuses on two different forms of ICT use: online shopping and WFH. Online shopping generally reduces the urge to make shopping trips. People who work from home may contribute to this by making other non-work trips such as shopping trips. Besides, these activities may also be influenced by other exogenous factors. Previous studies mainly focused on one activity (either online shopping or WFH) in terms of shopping trip generation. This study attempts to model shopping trips with online shopping and WFH together by integrating pre- pandemic data with the during-pandemic data. By harnessing the pre-pandemic 2017 NHTS data, we analyze three different models for the three endogenous variables (i.e., WFH, online shopping, and shopping trips) in a joint framework with a conditional mixed process approach that can correct the unobserved endogeneity and selection bias. The during-pandemic analysis captures the impacts of COVID-19 on travel behavior by exploring the US Census Bureau's experimental household pulse survey database.

Overall results suggest that online shopping and physical shopping trips can be substitutes, suggesting that online shopping is associated with reductions in shopping trips. In contrast, WFH encourages people to undertake more shopping trips in pre-pandemic periods. These associations are found to be similar during the pandemic but differ in magnitude. Notably, the WFH percentage has increased during the pandemic, as expected. People who work from home during COVID-19 are less interested in making in-person shopping trips and more interested in shopping online than in pre-pandemic periods. These key relationships are correlated with socio-demographic, location, and individual travel-related factors. Hence, the results generated for the ICT uses are robust. Some explanatory factors are found to be different from the pre-COVID-19 results, e.g., younger people and higher household size are more likely to work from home during COVID-19 as opposed to the findings in pre-COVID-19 periods.

These results should be interpreted with caution because the vaccination was still in the early stages during the data collection periods (week 23 - January/February 2021). Very few people had the choice to return to work, and many people were still reluctant to make physical shopping trips. Besides, pandemic-related restrictions across blue and red states were different over time. Restrictions in red states were not as strict as in blue states. Guess et al. (*40*) found that blue states had significantly higher scores than red states regarding behavioral/mitigation efforts. These include the duration of lockdowns, mask mandates, and vaccination rates. Therefore, compared to the pre-COVID-19 periods, the travel behavior (e.g., shopping trips) during COVID-19 might be similar in red states but different in blue states. Nonetheless, the results have important policy implications. The findings of this study may benefit future transportation planning (e.g., trip generation forecasting) and policymaking with the progression of ICT. Currently, planners do not account fully for ICT uses (e.g., WFH and online shopping); they also do not account for uncertainty due to large-scale events like COVID-19. Furthermore, other complications, such as online shopping, have a delivery component whose impact may be large both due to volume and the nature of the vehicles used. These types of complex relationships between ICT uses and travel behavior can inform planners and decision-makers to formulate more comprehensive policy provisions on different levels of ICTs in different periods (e.g., uncertainty) that can be used as an effective travel demand management technique. For example, a reduction in commuter trips due to more people working from home needs to be integrated into the travel demand models, i.e., trip generation and time-of-day models. In addition, the study findings show that both ICT uses have the potential to reduce the negative effects of transport (i.e., congestion, pollution, and other related external factors that may compel individuals to make unnecessary trips in the event of a global emergency). The reduction in vehicular trips following the increased ICT uses can improve the accessibility of all active transportation modes, including walking and bicycling, promoting non-motorized transport and the local built environment (*41*). Overall, the behavioral changes explored in the paper have strong implications for future economic activity, safety, traffic congestion, energy consumption, emissions, etc.

Future research may emphasize the inclusion of variables that may effectively define the spatial variation in the parameter estimates. Future research can also conduct a bidirectional study to understand better the assumption that shopping trips, online shopping, and WFH influence each other. Newer developments in e-commerce (e.g., online grocery shopping) and the many ways in which online shopping and physical shopping trips interact should be investigated in future research with the availability of newer data. Furthermore, the changing landscape of pandemic-related restrictions both over time (i.e., different waves of the pandemic) and across space (i.e., blue/red states) can be explored in future studies.

2.9 Acknowledgments

This study was partially supported by the Collaborative Sciences Center for Road Safety, CSCRS, a US DOT funded National Transportation Center, under Grant No. 69A3551747113. The authors would like to acknowledge Ms. Meredith King for proofreading the paper. Finally, the authors would also like to thank Dr. Ramin Arvin for his support and help.

CHAPTER 3 CRASH HARM BEFORE AND DURING THE COVID-19 PANDEMIC: EVIDENCE FOR SPATIAL HETEROGENEITY IN TENNESSEE

A version of this chapter was originally published by A. Latif Patwary and Asad J. Khattak in the Journal of Accident Analysis and Prevention:

Patwary, A. L., & Khattak, A. J. (2022). *Crash Harm Before and During the COVID-19 Pandemic: Evidence for Spatial Heterogeneity in Tennessee*, Accident Analysis and Prevention, 03611981221116626.

3.1 Abstract

Major concerns have been raised about road safety during the COVID-19 pandemic in the US, as crash fatalities have increased despite a substantial reduction in traffic. However, a comprehensive analysis of safety-critical events on roadways based on a broader set of traffic safety metrics and their correlates is needed. In addition to fatalities, this study uses changes in total monetary harm as additional measures of safety. A comprehensive and unique time-series database of crashes and socio-economic variables is created at the county level in Tennessee. Statistics show that while fatal crashes increased by 8.2%, total crashes decreased by 15.3%, and the total harm cost was lower by about $1.76 billion during COVID-19 (2020) compared with pre-COVID-19 conditions (2019). Several models, including generalized least squares linear, Poisson, and geographically weighted regression models using the differences between 2020 and 2019 values, are estimated to rigorously quantify the correlates of fatalities and crash harm. The results indicate that compared to the pre-pandemic periods, fatal crashes that occurred during the pandemic are associated with more speeding & reckless behaviors and varied across jurisdictions. Fatal crashes are more likely to happen on interstates and dark-not-lighted roads and involve commercial trucks. These same factors largely contribute to crash harm. In addition, a greater number of long trips per person not staying home during COVID-19 is found to be associated with more fatalities and crash harm. These results can inform policymaking to strengthen traffic law enforcement through appropriate countermeasures, such as the placement of warning signs and the reduction of the speed limit in hotspots.

3.2 Introduction

The novel coronavirus 2019 (COVID-19) pandemic has undoubtedly impacted the whole world with unparalleled destruction. However, free-flowing traffic can be considered one of the silver linings of the pandemic. Stay-at-home orders, voluntary isolation, working from home, and the fear of contracting the virus contributed to a substantial reduction in traffic flow (*42; 43*). In the US, people drove 13% fewer miles in 2020 than in 2019 (*44*). Active travel (walking and bicycling) was increased with the reduction in total trips, followed by the variation in COVID-19 case severity initially (*45*). In Tennessee, total vehicle miles traveled (VMT) have been reduced by 20% in 2020 compared to 2019 (*46*). Reduced traffic flow has produced a noticeable decline in congestion and emissions (*42*). Thus, there should be a decrease in the total number of crashes after the pandemic. However, recent information suggests while there are fewer crashes now than in 2019, the fatality rate has increased (*47*).

The national safety council (NSC) reported an 8% increase in fatalities across the U.S. from 2019 to 2020, with over 4,200 fatalities (*48*). Tennessee experienced a more than 7% increase in fatalities in 2020. The increase in the fatality rate from 2019 to 2020 is the highest estimated year-to-year jump in the US in over 96 years (*49*). This unprecedented situation is raising some serious safety concerns. While COVID-19's impacts on road safety are relatively unknown, some factors brought about by the pandemic may shed some light. For example, the combination of risky drivers and near-empty streets may result in faster driving, which in turn increases the likelihood of fatality during a crash. Researchers and policymakers are concerned that the new traffic pattern during the pandemic could lead to excessive speeding (*42*). Reports suggest that speeding cases in the U.S. have risen by 20%, and the number of speeding tickets has more than doubled (*49; 50*). It is also reported that fewer drivers were wearing a seatbelt on the road during the pandemic (*49*). Moreover, more seriously impaired drivers are found on roads, making the roads more vulnerable to fatal crashes (*42*). The pandemic has increased alcohol and cannabis sales (*51*). It is also reported that stress, anxiety, and depressive traits have been prevalent during the pandemic (*52*), and these have been identified for reckless driving behavior in the past (*53*).

This extraordinary situation is also contributing to more economic damage. Road crashes alone cost about $1 trillion in the loss of life and productivity each year in the US, where each crash fatality costs an average of $1.4 million (*54*). However, this loss in value may not necessarily reflect the true scenarios in the wake of worsened road safety arising from reduced mobility during COVID-19. Hence, there should be an in-depth investigation of road crashes and the overall economic impacts due to COVID-19. This investigation would help us design more effective and safe interventions for the current and forthcoming pandemic waves and similar outbreaks. Furthermore, understanding how COVID-19 has impacted road safety is imperative for the vision zero goals, as we seek to eliminate traffic fatalities and severe injuries while increasing safe mobility. Specifically, this study attempts to examine the factors associated with the increased crash fatalities within the state of Tennessee during COVID-19. The study also makes an effort to analyze how the economic harm in a crash has changed during the pandemic.

3.3 Literature Review

Quantifying the changes in travel and safety during the pandemic can be challenging. Positive impacts of the stay-at-home pandemic order may include less congestion, fewer traffic crashes and fatalities, fewer emissions, and less vehicular energy consumption (*42; 55*). Some previous studies showed a positive association between congestion and injury crashes (*56-58*). The number of crashes increases moderately with the increase of traffic in an uncongested roadway segment (*59*). However, when the critical traffic density is attained, the number of crashes starts to surge rapidly with the traffic (*60; 61*). Therefore, in a congestion-free environment, the number of crashes should decline. During the pandemic, statewide stay-at-home policies led to a 20% decline in road crashes in the US (*44; 62*). Some states experienced a higher reduction in crashes. For example, total crashes decreased by 50% in North Carolina compared to the pre-pandemic era (*63*). Although the number of crashes has decreased, there are contradictory reports about the number of fatalities during the pandemic. In some cases, minor injury crashes decreased, whereas severe and fatal crashes stayed the same (*64*). Moreover, the number of fatalities decreased in some states; Hawaii, Wyoming, Delaware, and Nebraska have experienced a decline of 20%, 13%, 11%, and 9%, respectively. In contrast, fatalities increased in some other states and contributed to the overall 8% increase across the US (*48; 65*).

The association between negative safety effects and reduced congestion from the pre-pandemic era has been studied (*66; 67*). Since congestion is usually localized, specific time analysis may be needed to form a better association. Exposure tends to be a key confounding factor, especially for vulnerable road users' activities. Traffic exposure can be measured in vehicle miles traveled (VMT), average annual daily traffic (AADT), and the number of trips in a certain unit of time in a certain region (*60*). A few earlier studies explored the cross-sectional relationship between VMT and fatal crashes and found both positive and negative associations (*68; 69*) (Table 3-1). For example, Doucette et al. (*70*) found that the crash rate more than quadrupled after accounting for the VMT reductions during COVID-19 in Connecticut. Average speed is another exposure used in literature. A higher speed is associated with more fatal and severe injury crashes (*42; 56*). For instance, California experienced little to no reduction in fatalities with the decline of VMT because of speeding during the pandemic (*56*). A 10-mile-per-hour (mph) higher speed limit increases the chance of fatal crashes between 15% to 60% (*71*). However, this can largely differ across locations over time because traffic flows in urban areas are often limited by congestion rather than speed limits. It is observed that during COVID-19, the speed effect is generally the largest in locations with some pre-existing level of congestion (*42; 56*). This effect can partially offset the reduced traffic flow on fatal crashes. Therefore, regional differences in terms of topography, roadway type, lighting conditions, or weather could be explored as possible input variables on fatal crashes. Empty roads trigger speed-related violations (e.g., speeding, red-light running, failure to comply with stop signs, failure to yield to other drivers or vulnerable road users) (*58*) (Figure 3-1). For example, crashes remain low in New York City due to low traffic volume. Yet, an increase in severe injury and fatality rate hints at higher traffic speeds, which is supported by an increase of the number of speeding tickets in downtown by 108% and in school zones by 72% during the pandemic (*57*). A recent study by Doucette et al. (*72*) on post-stay-at-home periods showed crash rates are slowly starting to return to previous year averages. Overall, measures like speeding, alcohol & drugs, other types of reckless behavior, and trips per person not staying home can be used to analyze the potential behavior of the drivers and their safety consequences. In addition, vehicle factors like the increased use of commercial trucks amid the pandemic could add more insights. All the relevant studies are summarized in Table 3-1.

Literature suggests that time-series analysis models like seasonal autoregressive integrated moving average (ARIMA) have been used to examine potential changes in fatalities or crashes across time (*58; 73*). Linear models have been adopted to explore crash harm resulting from fatalities and crashes (*74*). Generally, count models like Poisson and negative binomial regression are powerful predictive tools that are being applied in crash frequency analysis (*75; 76*). Most of the crash data are over-dispersed, which is a condition suggesting the need for correction to Poisson regression assumptions. In that case, the negative binomial often performs better than Poisson regression in crash frequency analysis (*77*). The zero-inflated negative binomial model has also been used to address the overdispersion problem caused by excessive zero counts (e.g., zero fatalities in a month at a location) (*76*). Besides, geographically weighted regression (GWR) can better capture the inherent spatial autocorrelation and heterogeneity in the crash data (*78*). Therefore, this study will adopt Poisson/negative binomial models (depending on the dispersion of the data) to analyze fatalities and crashes while using the linear model to analyze crash harm. The spatial aspects of crash data will be explored using GWR count and linear models.

The aforementioned studies are selected based on stay-at-home orders, congestion

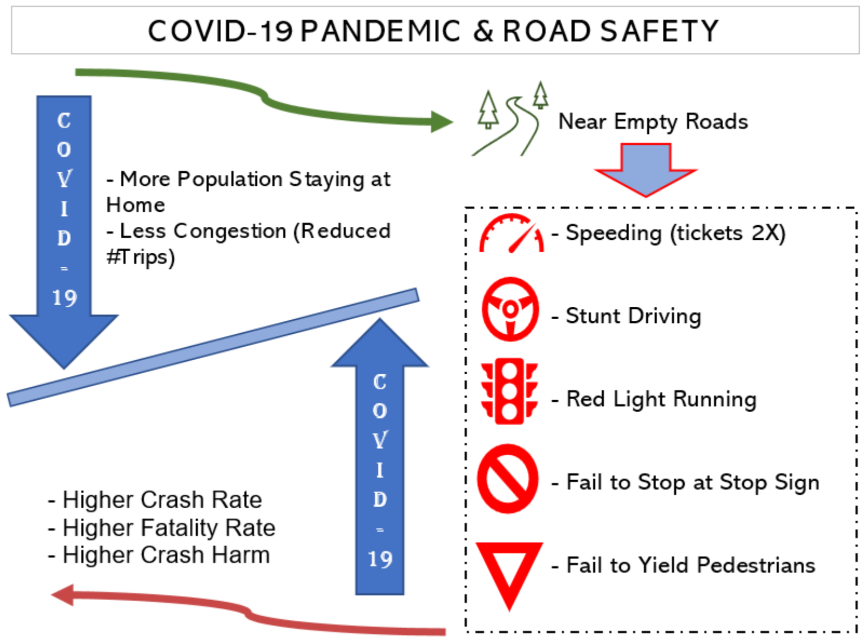


Figure 3-1: Graphical Abstract

Table 3-1: Summary of the Selected Literature

Author	Study Approach			Relationship between road safety and congestion	Study relevance
Name and Year	*Study Period (B = Before, D= During COVID-19 pandemic)*	*Location and Sample size*	*Method*	*Increase safety (+) or reduces safety (-)*	*Relevance to the research topic*
Lin et al., 2020	D	Los Angeles and New York, N= na	Time series data (January to August 2020), Difference-in-difference analysis	-	High
Doucette et al., 2021	B+D	Connecticut, N= na	Time Series Data, (January 1st to April 30th in 2017-2020), Interrupted Time-Series Analysis	-	High
Hughes et al., 2020	B+D	California; N=44 counties	Cross-sectional Data (Mar 2015 to May 2020), Poisson Regression	-/+	Medium
Doucette et al., 2021	B+D	Connecticut; N= na	Time-series data (January 1st-August 31st, 2017, 2018, 2019, 2020), Interrupted Time-Series Analysis	-	High
Inada, 2020	B+D	Japan, N= 121	Monthly time-series data (Jan 2010 to May 2020), ARIMA	-	Medium
Saladie et al., 2020	B+D	Tarragona, Spain; N= 152	Time series (Feb-Apr 2018- 2020), Comparative analysis	+	Medium
Quddus et al., 2010	B	UK, N= 72	Road segments data, Ordered Response Models	-	Medium
Kononov et al., 2008	B	CA, CO, TX of US; 5 years crash data	Freeway crash data, Neural Networks	+	Medium

reduction, and road safety criteria. Research on these topics in the pandemic era is limited, and most of them are descriptive-based analyses, which may be a sign of a rush to publish the papers during the pandemic. As such, the outcome of those studies might influence their premature publication. However, several potential gaps are identified in the existing literature on road safety issues. Existing literature considered factors like VMT, speed, or lockdown dummy in their estimation. However, different types of driving violations during COVID-19 have not been explored in the literature. Also, whether the distance traveled from home during the pandemic has any impact on road safety or not needs to be investigated. In addition, analysis of the crash harm during COVID-19 needs to be explored. As the injury severity is high during the pandemic, it may provide a different level of estimation of the harm. As such, to fill the gaps in the literature, the study sought to investigate the impact of factors contributing to fatalities and economic harm within the state of Tennessee before and during the pandemic by harnessing a unique database integrating crash and COVID-19 travel behavior data.

3.4 Conceptual Framework

The study anticipates that an increase in the number of speeding-related crash cases increases the number of fatalities in crashes. It is believed that this assumption also holds for other types of violations, i.e., alcohol-drugs-involved cases and other reckless driving cases. It is also expected that roadway factors like dark-lighted conditions and interstate crashes are suspected to increase fatalities. Vehicle-specific factors like more trucking activity during COVID-19 make roads vulnerable and can contribute to increasing fatalities in crashes. A higher number of short trips or long trips per population not staying at home is associated with higher exposure and crash risk. Regarding the economic losses, total crash harm could increase during COVID-19 compared to the pre-pandemic scenario. As the fatality rate increases, the crashes' economic loss could increase (Figure 3-2).

3.5 Data: Linking Crashes and Traveler Behavior

County-level monthly data (cross-sectional time series) of the state of Tennessee covering the pre-pandemic and during the pandemic period is used in this study. A unique database is created by integrating two different sources, i.e., the Tennessee Integrated Traffic Analysis Network (TITAN) and the Bureau of Transportation Statistics (BTS) "Trips by Distance" (*79*). These datasets are linked at the county level. To make a consistent timeline, the study considers the same months pre-and-during the pandemic. In this study, ninety-five counties of Tennessee over 18 months (i.e., p*re-pandemic and during-pandemic periods)* constitute a sample of 1,710. The study's pre-pandemic refers to the periods in 2019 before the COVID-19 pandemic. In particular, January 1st to September 30th of 2019 is considered pre-pandemic. On the other hand, during-pandemic periods include the periods during the pandemic in 2020. The study uses a similar timeframe for the during-pandemic periods (January 1st to September 30th). The dates were chosen by following the COVID-19 timeline in the state of Tennessee. Specifically, the Tennessee state health operation center was activated in January 2020 with the declaration of COVID-19 as a public health emergency in the US. Tennessee started the state of emergency in March 2020 and continued through September 2020, which was initiated to encourage social distancing to help mitigate the spread of the virus (*80*). In addition, the use of a similar timeframe for pre-and-during pandemic

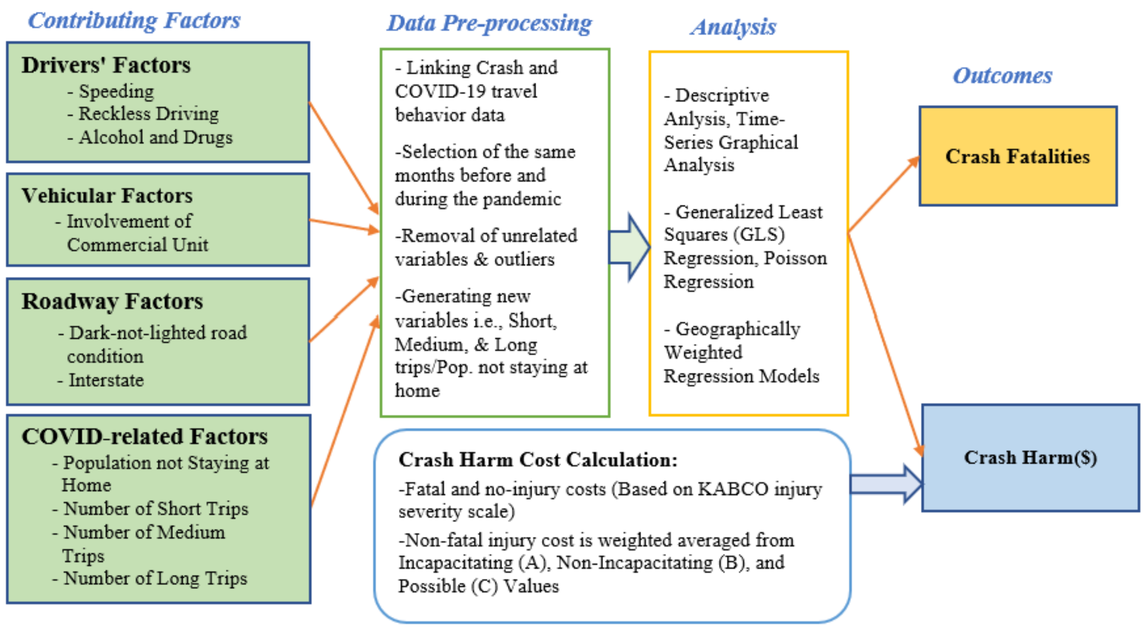


Figure 3-2: Study Framework

periods has already been documented in the literature (*55; 70; 72*). COVID-19-related information and travel information are collected from the BTS. Data on the number of people not staying at home is provided by the Maryland Transportation Institute and Center (MTIC) for the advanced transportation technology laboratory at the University of Maryland (*79*). MTIC collects travel information, such as the number of trips from anonymized "national panel of mobile device" data from multiple sources. The sample of mobile devices is representative of the entire population in a state or county. MTIC does not report data for counties having fewer than 50 devices in the sample on any given day to assure better data quality. MTIC defines trips as movements that last longer than 10 minutes at any location away from home.

The number of crashes and the number of fatalities per month are collected from the TITAN database, which is maintained by the Tennessee Department of Transportation (TDOT). TITAN includes information for all the police-reported crashes in Tennessee. TITAN also provides information regarding the number of injured and non-injured persons. An additional safety measure, crash harm, is adopted to account for the economic value of each injury level and the costs for each injury/property damage. For example, two fatalities in a crash will still be coded as a fatal crash, but crash harm captures this in terms of monetary cost. Comprehensive crash unit cost values (2016 dollars) of the federal highway administration (FHWA) are used to create this unique variable (*81*). The dollar values are in the KABCO injury severity scale, whereas TITAN reports injury severity in three categories: fatal injury, non-fatal injury, & no injury. The study uses the economic values for fatal injury, non-fatal injury, & no injury types severity from the FHWA given values: $11.29 million for each fatal injury, an average value of $0.23 million for each non-fatal injury, and $0.012 million for each non-injured person involved in crashes. Crash harm is calculated at the county level by month and year. The following Equation (1) is applied to calculate the crash harm.

$$H_{mky} = \sum_{i=1}^{3} C_i * N_i \qquad (1)$$

Where,

H_{mky} = Crash harm of county "k" in month "m" and year "y"; (k = 1, 2, 3, ….. , 95; m = 1, 2, ……, 9; y = 2019, 2020)

C_i = cost of each injury severity type i; (i = 1-Fatal injury, 2-Non-fatal injury, 3-No injury)

N_i = Number of persons involved in injury severity type i

For example,

Crash harm of Carter County in March 2020 = $11.295 million * number of fatally injured persons + $0.23 million * number of non-fatal injured persons + $0.012 million * number of non-injured person involved in crashes = $11.295 * 2 + $0.23 * 34 + $0.012 * 185 = $31.61 million.

The dataset used in this study is error-checked through descriptive analysis, and the data is reasonable with no extreme outliers. Table 3-2 reports the descriptive statistics of the data. It is divided into pre- and post-COVID-19 scenarios. The mean values per ten thousand trips (i.e., Mean/Trips) are also generated to compare the pre-and-post COVID-19 values. There are three dependent variables: number of fatalities, number of crashes, and crash harm. As expected, it is found that fatalities have increased by 8.2% during the pandemic compared to the pre-pandemic periods, whereas the number of crashes has decreased (15.3%). Moreover, the total crash harm was lowered by $1.76 billion during COVID-19 in Tennessee. However, if per-trip values are compared, it is found that the crash harm and the number of crashes are higher during COVID-19 than in pre-COVID-19 periods.

Regarding the independent variables, "*No. of Speeding Violations*" represents the number of speeding cases that resulted in crashes. It shows that such cases were higher in 2020 compared to 2019. The frequency of *alcohol and drug-related* cases and *reckless behavior cases* per trip also increased in 2020. Similarly, roadway and vehicular factors per trip are also reported to be increased during COVID-19. These variables are explored graphically in the next section. Three unique COVID-19-specific independent variables are generated from the collected BTS database by combining the information regarding the number of trips in terms of length and population not staying at home, which can reflect how far a person traveled when they did not stay at home during COVID-19. Since these unique variables on "trips per person not staying at home" will be added to the modeling, the crash/fatality per trip is not modeled, i.e., it is not used as a dependent variable. Trips are categorized by length: trips greater than zero to less than or equal to 5 miles are considered short-length, trips greater than 5 miles and less than or equal to 50 miles are considered mid-length, and trips greater than 50 miles are considered long trips. It is observed that the rates of mid-length and long trips were higher during the pandemic, whereas short-length trips showed a decline.

3.6 Exploratory Analysis

Figure 3-3 illustrates the monthly trend of road crashes, fatalities, and total crash harm in Tennessee. The time-series graphs capture several noteworthy facts. All the values plummeted in the beginning phase of the pandemic (Feb'20 to Apr'20). After that, the values begin to catch up and eventually return to the pre-pandemic scenarios. However, the number of fatalities rises higher than in the pre-pandemic era, suggesting a serious concern over road safety during the COVID-19 pandemic.

3.6.1 Drivers' Factors

Drivers generally make different types of violations while driving. Literature suggests that speeding-involved crashes are common during COVID-19. Drug and alcohol use and reckless behaviors are also other top-ranked causes of car crashes. Reckless behaviors include tailgating, failing to stop at red lights or stop signs, braking abruptly, not using turn signals when changing lanes or turning, failing to use headlights at night or in extreme weather, and making illegal turns or lane changes. Figure 3-4 suggests that reckless driving contributed to a 50% increase in fatal crashes in Tennessee, while the total number of crashes decreased by 15%. The increase was spotted during the beginning phase of the pandemic (i.e., March 2020) and in the latter part of 2020. Speeding-related crashes and fatalities have both increased during COVID-19. There was an overall increase of 15 speeding-related fatal crashes in 2020 compared to 2019. The cases remained high for almost all of 2020.

3.6.2 Roadway Factors

Roadway factors include the lighting condition of the roads and the speed limits of interstates, among others. It is difficult to detect a pedestrian walking or an object at night, even on a lighted road, due to a lack of good visual acuity and contrast sensitivity. Crashes and fatalities in the dark (not lighted) conditions are higher than in the day-light conditions, according to the literature (*82*). Roads such as interstates with higher speed limits can be vulnerable to fatal crashes, especially with the presence of reckless drivers. In Figure 3-5, it can be observed that crashes on the

Table 3-2: Descriptive Statistics (N = 1,710) ***(Monthly, Per County)***

	Pre-pandemic (Jan-2019 to Sept-2019)						**During-pandemic (Jan-2020 to Sept-2020)**						**Differences**	
Variable	***Mean***	***SD***	***Min***	***Max***	***Total***	***Mean/ Trips ('10,0 00)***	***Mean***	***SD***	***Min***	***Max***	***Total***	***Mean/ Trips ('10,0 00)***	**"Mean " Diff (2020-2019)**	**"Mean/Trips " Diff (2020-2019)**
No. of Fatalities*	1.01	2.03	0	19	867	0.044	1.06	2.63	0	29	938	0.052	0.05	0.01
Total Crash Harm ($Millions)*	28.92	66.9 5	0.02	554.33	24723.33	0.991	26.86	67.31	0.02	647.7 7	22966. 45	1.032	-2.06	0.04
No. of Speeding Violations	1.37	6.27	0	59	1172	0.012	1.54	6.87	0	56	1313	0.017	0.17	0.01
No. of Reckless Driving Cases	7.65	22.9 7	0	256	6537	0.22	6.52	21.72	0	232	5577	0.23	-1.13	0.01
No. of Alcohol & Drugs Cases	6.47	10.0 5	0	93	5486	0.243	6.48	9.82	0	5	5541	0.329	0.01	0.09
No. of Cases Involving Commercial Units	4.16	8.07	0	59	3558	0.15	4	7.57	0	53	3417	0.171	-0.16	0.02
No. of Cases on Interstate Roadways	18.09	60.9 4	0	459	15469	0.29	14.94	48.81	0	425	12776	0.323	-3.15	0.03
No. of Cases on All Other Roadways	97.36	318. 15	0	3030	164342	1.84	84.64	278.1	0	2797	13954 3	2.054	-12.72	0.21
No. of Day-Lighted Cases	146.04	12.8 9	1	2940	124864	3.31	119	9.99	0	2475	10174 9	3.62	-27.04	0.31
No. of Dark-not-Lighted Cases	17.77	0.89	0	212	15191	0.76	17.13	25.04	0	239	14643	0.87	-0.64	0.11
Short-length trips rate	2.85	0.53	1.62	5.13	2436.44	-	2.37	0.34	1.49	4.05	2022.4	-	-0.48	-
Mid-length trips rate	2.03	0.38	1.29	4.22	1735.35	-	4	0.39	2.95	6.53	1400.8 2	-	1.97	-
Long trips rate	0.15	0.06	0.06	0.47	129.61	-	0.16	0.05	0.04	0.4	138.22	-	0.01	-

* Dependent variables

interstate decreased by 17% in 2020; however, fatal crashes surged by 21% in 2020 compared to 2019.

3.6.3 Vehicular Factors

Figure 3-6 presents the trends of crashes and fatalities involving commercial units in Tennessee during 2019-2020. It is observed that the total number of crashes involving commercial vehicles decreased by 7%, while fatal crashes increased by 11% in 2020 compared to 2019. Fatal crashes were 17 at the start of the pandemic, 24 in August 2020, and 10 in September 2020.

3.7 Modeling

In this study, the number of fatalities and crash harm are modeled as the dependent variables. To show the COVID-19 impacts solely, the first differences are calculated using the monthly county-level data for 2019 and 2020. The first differences are the values found by subtracting 2019 values from 2020, as shown in *Equations (2) and (3)*. These monthly differences account for the variation across counties and months and provide a better measure of the correlates of fatalities and crash harm. The differences can have both positive and negative values. Positive difference values show that 2020 values are higher than 2019, and vice-versa. In the analysis, county and month are indexed to confirm the panel structure of the data.

$$Y = \Delta y_i = y_{i,2020} - y_{i,2019} \quad (2)$$
$$X = \Delta x_i = x_{i,2020} - x_{i,2019} \quad (3)$$

Where, Δy_i is the difference between 2020 and 2019 in the dependent variable i and Δx_i is the difference between 2020 and 2019 in the independent variable i.

3.7.1 Generalized Least Squares Linear Regression Model

The difference in crash harm is a continuous variable that can be modeled using the generalized least squares (GLS) linear regression model. Previously, linear models were applied to estimate the coefficients for total crash harm in work zone crashes (*74*). GLS extends the ordinary least squares estimation by addressing the possible unequal error variances and correlations between different errors in the time-series data (*83*). Therefore, GLS can efficiently estimate regression coefficients, as shown in *Equation (4)*.

$$Y = X\beta + \varepsilon, \quad E(\varepsilon) = 0, \quad Cov\,(\varepsilon) = \omega \quad (4)$$

Where Y is the dependent variable, and X are the independent variables for a set of units, i.e., TN counties over time. β denotes the unknown regression coefficients. ε is the vector of random errors, and ω is the variance-covariance matrix. GLS involves minimizing $(Y - X\beta)'\omega^{-1}(Y - X\beta)$ with respect to β. The resultant estimator b of the regression coefficients β can be expressed in *Equation (5)*.

$$\mathrm{b} = (\mathrm{X}'\omega^{-1}X)^{-1}X'\omega^{-1}Y \quad (5)$$

Equation (6) denotes the estimated covariance matrix V of b.

$$V = \widehat{Cov}(b) = (X'\omega^{-1}X)^{-1} \quad (6)$$

A notable property of GLS is that its estimate of β is unbiased (*E(b) = β*).

Under the assumption of normality distributed random errors, the log-likelihood function (ℓ) can be written as follows in *Equation (7)*:

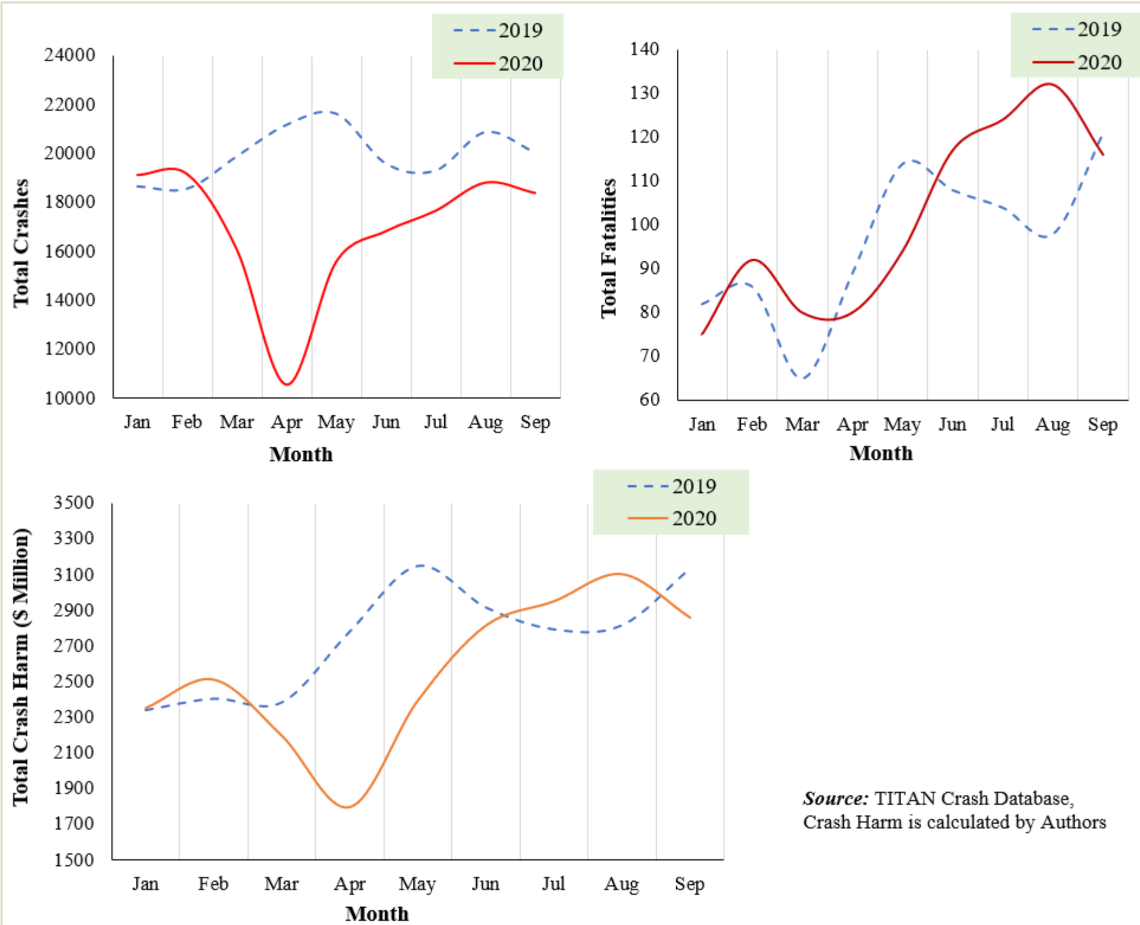
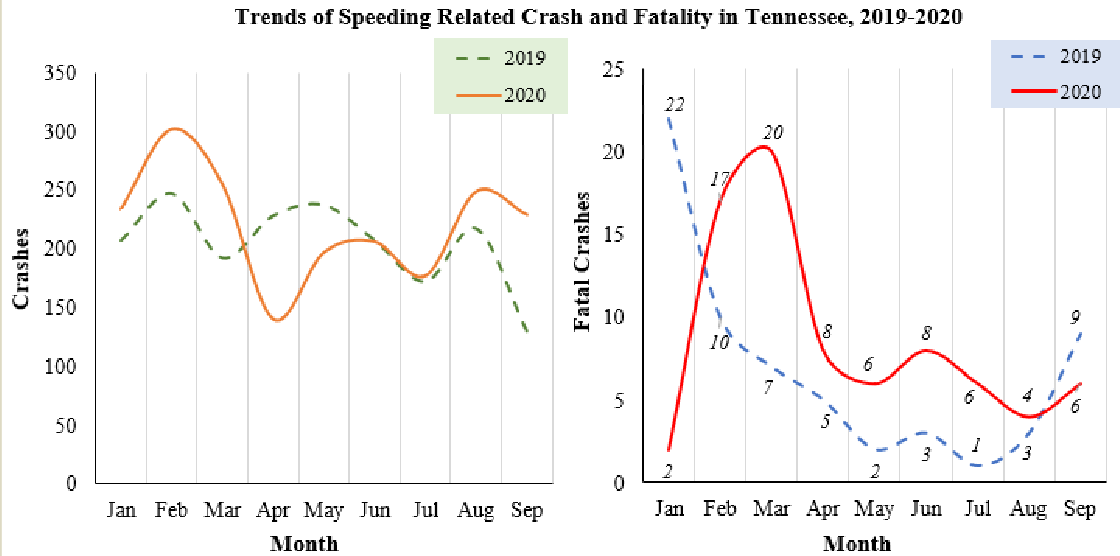
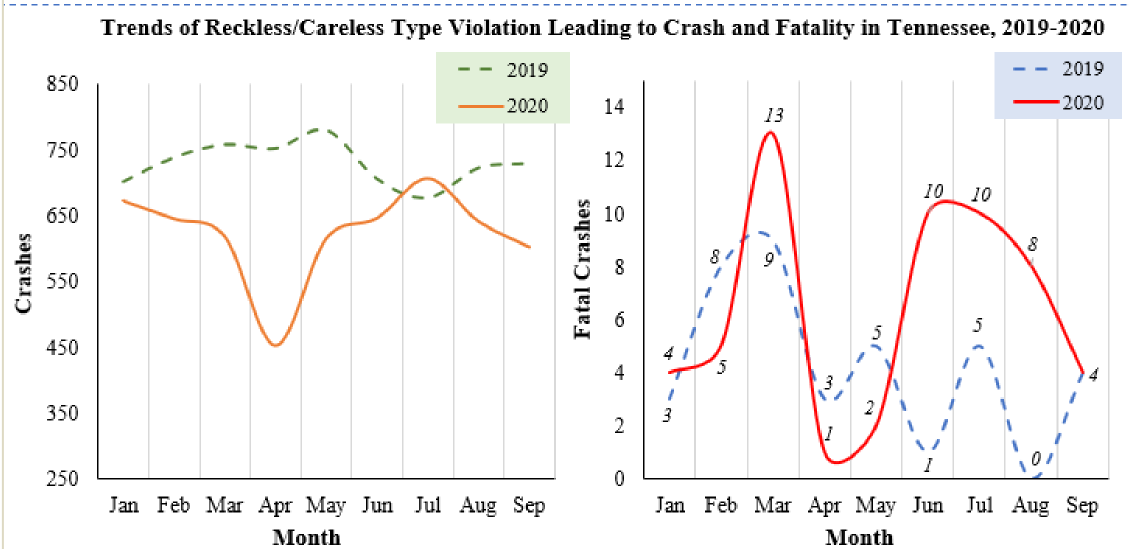


Figure 3-3: Monthly Trends of Crashes, Fatalities, and Crash Harm in Tennessee, 2019 (pre-pandemic) – 2020 (during pandemic)

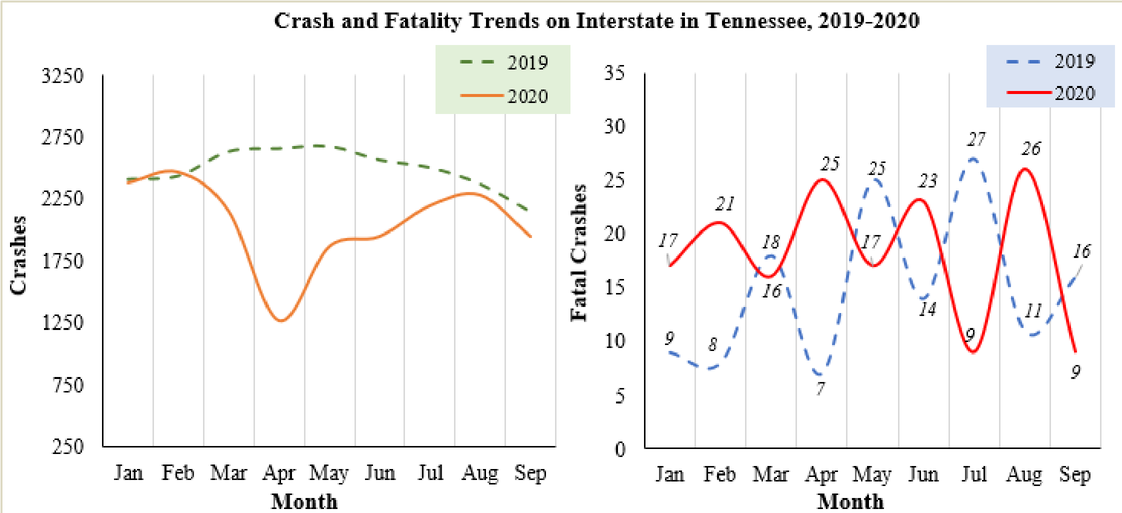


Year	Crashes	Fatal Crashes
2019	1,836	62
2020	1,989 (+8%)	77 (+24%)



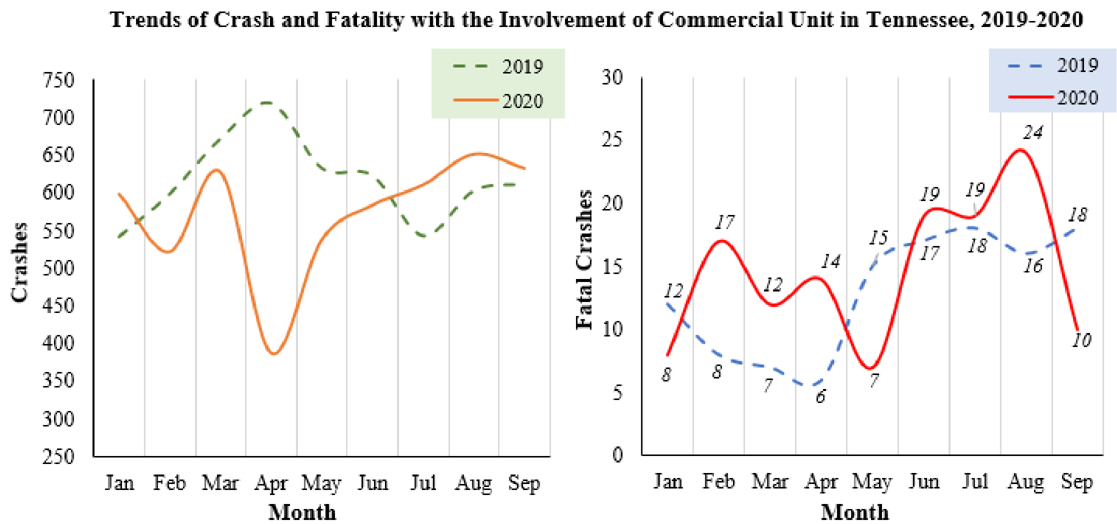
Year	Crashes	Fatal Crashes
2019	6,561	38
2020	5,598 (-15%)	57 (+50%)

Figure 3-4: Speeding and Reckless Driving- related Crashes & Fatalities in Tennessee, 2019 (pre-pandemic) - 2020 (during-pandemic)



Year	Crashes	Fatal Crashes
2019	22,389	135
2020	18,522 (-17%)	163 (+21%)

Figure 3-5: Roadway Factors' Related Crashes and Fatalities in Tennessee, 2019 (pre-pandemic) – 2020 (during-pandemic)



Year	Crashes	Fatal Crashes
2019	5,537	117
2020	5,154 (-7%)	130 (+11%)

Figure 3-6: Involvement of Commercial Vehicle in Crashes & Fatalities in Tennessee, 2019 (pre-pandemic) – 2020 (during-pandemic)

$$\ell = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\ |\omega| - \frac{1}{2}[(Y - X\beta)'\omega^{-1}(Y - X\beta)] \quad (7)$$

3.7.2 Poisson Regression

The difference in the number of fatalities is a count variable. However, the variable contains both positive and negative count values. Since count models cannot handle negative values, a constant value (minimum of the difference in the count outcome variable) is added to the differences. Then, the transformed differences would be greater than or equal to zero. This transformation makes the outcome variables eligible to use the count models (*84; 85*). Count models, e.g., Poisson and negative binomial regression, are powerful predictive tools applied in crash frequency analysis (*75; 76*). The Poisson regression model can be employed to analyze count data when there is no overdispersion in the data.

The Poisson distribution of a random variable Y follows the following probability density function in *Equation (8)* for a given value Y=y:

$$P(Y = y|\varphi) = \frac{e^{-\varphi}\varphi^{y}}{y!} \quad (8)$$

Where φ is the mean rate of occurrence. This rate is determined by a set of k predictors, X = (X_1, X_2, ……., X_k). It can be expressed by *Equation (9)*:

$$\varphi = \exp\,(X\beta) \quad (9)$$

Then, the Poisson regression model for observation *i* can be defined by *Equation (10)* below:

$$P(Y_i = y_i|X_i, \beta) = \frac{e^{-\exp\,(X_i\beta)}\exp\,(X_i\beta)^{y_i}}{y_i!} \quad (10)$$

The likelihood function for a sample size n is given by *Equation (11)* below:

$$L(\beta; y, X) = \prod_{i=1}^{n} \frac{e^{-\exp\,(X_i\beta)}\exp\,(X_i\beta)^{y_i}}{y_i!} \quad (11)$$

Then, the log-likelihood function is generated, as shown in *Equation (12)*.

$$\ell(\beta) = \sum_{i=1}^{n} y_i X_i\beta - \sum_{i=1}^{n} \exp\,(X_i\beta) - \sum_{i=1}^{n} \log\,(y_i!) \quad (12)$$

An overdispersion test on *Equation (13)* can be performed to reflect how much the sample fluctuates around a mean value.

$$\mathrm{Var}(Y) = \mu + \alpha\,\mu^2 \quad (13)$$

Where α reflects the amount of overdispersion, which is non-negative and implies the variance Var(Y) can exceed the mean (μ). When α approaches zero, there is no overdispersion in the data (i.e., expected mean = variance). A likelihood ratio test is performed in STATA (statistical software) to test for the significance of the overdispersion parameter (α). When α is statistically not significant (5% level), the Poisson distribution can appropriately model the data.

3.7.3 Geographically Weighted Regression Models

Geographically Weighted Regression (GWR) models are also adopted in this study to explore further the existence of spatial non-stationarity or heterogeneity in the correlates of the difference in crash fatalities, crashes, and crash harm. Spatial heterogeneity shows different mean and variance values at each location if there exists any (*86*). The GWR model allows the parameters to vary over space; hence, it is believed to be applicable to the current analysis. GWR models were applied in the literature to analyze the spatial heterogeneity of related factors in road crashes (*87; 88*). Geographically weighted Poisson regression models are estimated for analyzing the difference in the number of fatalities. Also, a conventional geographically weighted linear regression is estimated for the differences in crash harm. Fixed Gaussian kernel functions have been used to determine the GWR weights that estimate the geographical changes in local extent.

The equations for the GWR models are given in *Equations (14) and (15). Equation (14)* describes the geographically weighted Poisson regression model, and *Equation (15)* shows the geographically weighted linear regression model.

$$Y_i = \sum_k \beta_k (u_i, v_i) X_{k,i} + \varepsilon_i \quad (14)$$

$$Y_i \sim Poisson[N_i \exp (\sum_k \beta_k (u_i, v_i)\, Y_{k,i})] \quad (15)$$

Here,

Y_i = dependent variable at location *i*;

$X_{k,i}$ = k^{th} independent variable at location *i*;

ε_i = Gaussian error at location *i*;

(u_i, v_i) = x-y coordinate of the i^{th} location;

$\beta_k(u_i, v_i)$ = coefficients that are varying conditionals on the locations

The equation to estimate $\beta_k(u_i, v_i)$ is as follows, i.e., *Equation (16)*:

$$\acute{\beta}\,(i) = (X^T W(i) X)^{-1} X^T W(i) Y \quad (16)$$

Here, W(i) represents a matrix of weights specific to location i such that observations nearer location *i* are given more weight than observations that are located far away from *i*. The form of the matrix W(i) is as follows, i.e., *Equation (17)*:

$$W(i) = \begin{bmatrix} w_{i1} & 0 & 0 \\ 0 & w_{i2} & 0 \\ 0 & \dots & w_{in} \end{bmatrix} \quad (17)$$

Here, w_{in} is the weight given to the observation n for the estimate of the local parameters at *i* location. *Equation (18)* denotes the adopted fixed Gaussian kernel of the GWR model.

$$w_{ij} = \exp\left(\frac{-d_{ij}^2}{\theta^2}\right) \quad (18)$$

Here,

i = regression point index; *j* = locational index;

w_{ij} = the weight value of observation at location j for estimating the coefficient at location *i*;

d_{ij}= Euclidean distance between *i* and *j*;

θ = A fixed bandwidth size defined by a distance metric;

3.8 Results

3.8.1 Results of the Preliminary Models

Results of the Poisson regression models and the GLS linear regression model are presented in Table 3-3. Columns (1) and (2) denote the results for the three dependent variables: difference in the number of fatalities (model 1) and crash harm (model 2). Model significance tests show that the models fit the data well. The pseudo-R^2 value of model (1) is 8%, and the R^2 value of the crash harm model is 44%. The correlation among the independent variables is checked, and the values are less than or equal to 0.5 or -0.5, referring to no multicollinearity issues. The distributions of the dependent variables are shown in Figure 3-7. In model (1), the Poisson distribution is shifted to 7 units to the right after adding the minimum difference of fatalities (which is -7) between 2020 and 2019 (Figure 3-7a). The general relationships between the independent variables and dependent variable (i.e., direction) do not change when a constant is added to the dependent variable, except everything is shifted to the same constant units to the right (*84*). In Table 3-3, the Poisson coefficient sign of the independent variables indicates the direction of their effects on the dependent variable. The generated average marginal effect explains the probability of the association. The coefficient of each variable can be interpreted one by one for all three models.

The first one is the "*Diff. in the No. of Speeding Violations"*, which is positive and significant for the three models. It indicates that a unit increase in the differences in *speeding* violation cases is associated with an increase in the difference in crash harm between 2020 and 2019 by $0.87 million. Also, an increase in the differences in *speeding*-related cases is associated with increased probabilities of crash fatalities by 0.07% in 2020 during COVID-19. The findings are consistent with the earlier assumption. Speeding-related crashes are dangerous and fatal. Speeding makes the vehicles more difficult to control, especially when driving around a curve or encountering a road hazard or other cars. Since speeding exerts the most force upon impact and involves a larger mass or higher acceleration, it causes the most severe injuries and fatalities and eventually generates more economic harm. Less traffic during COVID-19 encouraged drivers to speed, eventually leading to fatalities. "*Diff. in the No. of Reckless Driving Cases"* is also positive and statistically significant in model (2). The difference in reckless driving behaviors is found to be associated with an increase in the probability of crash harm by $0.31 million during COVID-19. Reckless behaviors, e.g., failing to yield to traffic and running stop signs and red lights, make roads more vulnerable to fatal crashes. "*Diff. in the No. of Alcohol & Drugs Cases"* is associated with an increase in the probability of the number of fatalities in crashes by 0.46%. The increase of one alcohol and drug-related crash is associated with an increase in crash harm differences by $7.42 million, indicating much greater damage during COVID-19 as a result of increased *Alcohol and drug*-related cases.

An increase in *"Diff. in the No. of Cases on Interstate Roadways"* is associated with a 0.02% increase in the probabilities of crash fatalities during COVID-19. Whereas on other roads, excluding interstates, the coefficient appears negative for fatalities. Interstates expose drivers to a higher speed limit than highways, which became deadly during COVID-19. An increase in *"Diff. in the No. of Dark-not-Lighted Cases"* is associated with an increase in total crash harm by $0.64 million. Fatalities appear to be increased in dark-not-lighted cases. However, it was not significant in the fatality model. On the other hand, it is found that if the "*Diff. in the No. of Cases Involving Commercial Units"* increases by one, the chance of crash fatalities increases by 0.02%. Besides, crash harm increases by $0.51 million. Long-haul drivers underwent significant changes during the pandemic that might have affected their health and safety. Traveling at higher speeds and risky behaviors could lead to vulnerability to crash fatalities.

The difference in the number of *Short-length trips* per person not staying at home is negative and statistically significant in model (1). Besides, the "*Diff. in Long trips rate"* is significant and positive in the model (2); however, the *"Diff. in Mid-length trips rate"* is not statistically significant. These results suggest that the increase in short trips per population not staying at home was associated with fewer crash fatalities. However, the longer trip lengths are associated with the increase in the probability of the differences in crash fatality, as shown by the increasing likelihood of crash fatalities for the *long trips rate*. Specifically, an increase in the differences in long trips per person not staying at home are associated with an increase in the chance of the differences in crash fatalities by 15.34% during the pandemic. Similarly, crash harm showed an increased association with the increase of long trips per person not staying at home

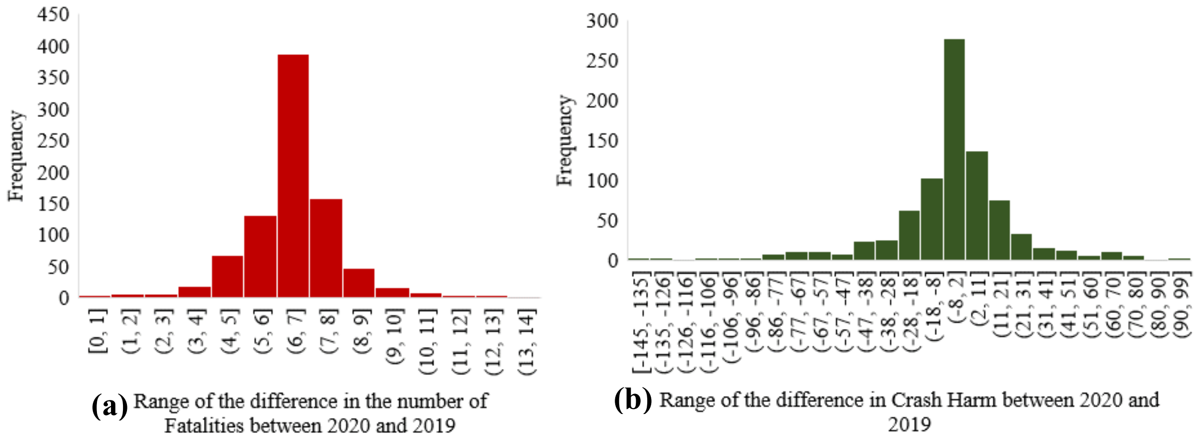


Figure 3-7: Histograms of the differences between 2020 and 2019 in (a) the number of Fatalities and (b) Crash Harm

Table 3-3: Estimation Results of the Preliminary Models

Variables (County Level)	(1) Diff. in the No. of Fatalities (Poisson Regression Model)			(2) Diff. in Crash Harm (Millions) (Generalized Least Squares Linear Regression Model)	
	Coef.	*P-value*	*Marginal Effect*	*Coef.*	*P-value*
Diff. in the No. of Speeding Violations	0.074	0.022**	0.075	0.866	0.001***
Diff. in the No. of Reckless Driving Cases	0.026	0.144	0.027	0.313	0.021**
Diff. in the No. of Alcohol & Drugs Cases	0.465	0.025**	0.468	7.42	0.000***
Diff. in the No. of Cases Involving Commercial Units	0.041	0.164	0.041	0.515	0.015**
Diff. in the No. of Cases on Interstate Roadways	0.021	0.052*	0.022	0.052	0.522
Diff. in the No. of Cases on All Other Roadways	-0.008	0.002***	-0.008	-0.114	0.000***
Diff. in the No. of Day-Lighted Cases	-0.001	0.700	-0.001	0.054	0.001***
Diff. in the No. of Dark-not-Lighted Cases	0.004	0.781	0.004	0.641	0.000***
Diff. in Short-length trips rate	-0.651	0.091*	-0.656	-0.703	0.728
Diff. in Mid-length trips rate	0.500	0.678	0.503	-0.731	0.781
Diff. in Long trips rate	15.237	0.233*	15.347	30.559	0.064*
Constant	13.516	0.000***		-0.977	0.319
Model Fit Statistics					
χ^2	128.62			588.45	
Model Significance Test (Prob>χ^2)	0			0	
Overdispersion (α)	0			na	
Log-likelihood	-1771.25			-3722.45	
AIC	3568.51			7470.71	
BIC	3630.27			7527.72	
Pseudo-R^2/ R^2	0.08			0.44	

*Note: * p<0.1; ** p<.05; *** p<.01*
"na" = not applicable

Table 3-4: Estimation Results of the GWR Global Models

Variables (County Level)	(1) Diff. in the No. of Fatalities (Poisson Regression Model)		(2) Diff in Crash Harm (Linear Model)	
	Estimate	***t (Est/Standard Error)***	***Estimate***	***t (Est/Standard Error)***
Diff. in the No. of Speeding Violations	0.010	1.57*	1.334	4.754***
Diff. in the No. of Reckless Driving Cases	0.000	0.019	0.452	3.108***
Diff. in the No. of Alcohol & Drugs Cases	0.583	2.128***	5.041	2.977***
Diff. in the No. of Cases Involving Commercial Units	0.040	1.318	0.631	2.757***
Diff. in the No. of Cases on Interstate Roadways	0.004	2.096**	0.205	2.349**
Diff. in the No. of Cases on All Other Roadways	-0.005	-1.953*	-0.135	-6.368***
Diff. in the No. of Day-Lighted Cases	0.002	1.710*	0.107	6.676***
Diff. in the No. of Dark-not-Lighted Cases	0.008	0.154	0.26	2.265**
Diff. in Short-length trips rate	-0.409	-0.765	0.78	0.359
Diff. in Mid-length trips rate	-0.065	-0.054	-4.395	-1.549
Diff. in Long trips rate	3.046	0.247	40.664	2.280**
Constant	13.539	9.821***	-1.339	-1.266
AIC	281.74		7590.9	
BIC	355.71		7652.66	
Pseudo-R^2/ R^2	0.085		0.45	

*Note: * p<0.1; ** p<.05; *** p<.01; "na" = not applicable*

Table 3-5: Estimation Results of the GWR Local Models

Variables (County Level)	(1) Diff. in the No. of Fatalities (Poisson Regression Model)				(2) Diff. in Crash Harm (Linear Model)			
	Mean β	Min β	Max β	*Test of Spatial Variability	Mean β	Min β	Max β	*Test of Spatial Variability
Diff. in the No. of Speeding Violations	0.007	-0.034	0.050	-1.580	0.927	-1.126	5.120	-35.686
Diff. in the No. of Reckless Driving Cases	-0.009	-0.015	0.001	0.189	-0.038	-0.468	0.800	3.015
Diff. in the No. of Alcohol & Drugs Cases	0.579	0.440	0.706	0.039	8.786	-1.835	46.291	-86.870
Diff. in the No. of Cases Involving Commercial Units	0.041	0.039	0.042	0.727	0.682	0.152	1.350	8.377
Diff. in the No. of Cases on Interstate Roadways	0.004	-0.004	0.010	-0.093	0.195	-0.239	1.463	-39.380
Diff. in the No. of Cases on All Other Roadways	-0.002	-0.006	0.001	-2.563	0.089	-0.161	0.980	-169.921
Diff. in the No. of Day-Lighted Cases	0.001	-0.001	0.002	-0.509	0.004	-1.160	0.170	-271.109
Diff. in the No. of Dark-not-Lighted Cases	0.004	-0.002	0.011	0.075	0.066	-0.480	0.525	2.008
Diff. in Short-length trips rate	-0.403	-0.485	-0.318	0.725	-0.783	-5.235	6.236	8.189
Diff. in Mid-length trips rate	0.004	-0.198	0.193	0.689	-1.468	-9.177	15.604	5.801
Diff. in Long trips rate	3.105	2.029	4.327	0.308	16.179	-35.52	71.928	5.546
Constant	13.537	13.497	13.580	0.536	-1.387	-4.400	2.016	4.725
AIC	279.05				7058.71			
BIC	338.31				7270.12			
Pseudo-R^2/ R^2	0.10				0.59			

** Geographical variability tests of local coefficients. A negative value suggests spatial variability. "na" = not applicable* during COVID-19. Overall, it is observed that the increase in trip length increases the fatality risk in crashes during COVID-19.

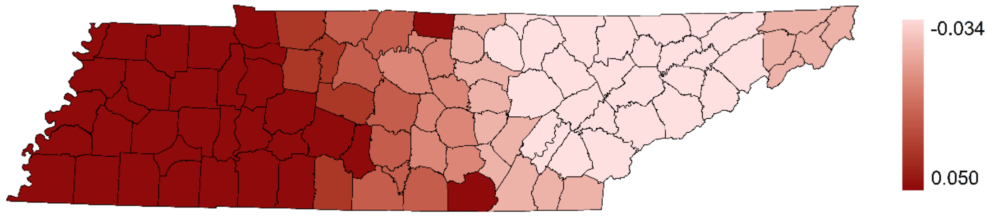
3.8.2 Results of the Spatial Models

Table 3-4 presents the results of the GWR global models with the coefficients and significance level. The t-value >1.96 or < -1.96 indicates that the variables are significant at a 95% confidence level and indicates a p-value of <0.05. The GWR local models' results are similar to the results of the discussed GLS regression and Poisson regression models. Moreover, the pseudo-R^2/R^2 values of the model (1) and (2) have slightly been improved, which are 8.5% and 45%, respectively. The GWR local models' results are illustrated in Table 3-5. The table contains various distribution parameters such as mean, minimum, maximum, and difference of criterion (i.e., a test of spatial variability) for the estimates. These values help to see the distribution of parameters and their range of variation across space. Variables with a negative difference in criterion values indicate the presence of spatial variability in those variables. The local model fits the data better than the global and first difference regression models. The pseudo-R^2/R^2, AIC, and BIC are better than the previously analyzed models (i.e., Poisson, GLS, and GWR global models). The pseudo-R^2 value for model (1) is now 10%. Similarly, the increased R^2 value for model (2) is 59%. In addition, the sign of the estimates in local models is the same as observed in the global models. The range of variation can be explored by looking at each variable's minimum and maximum values. It appears that for most of the variables, the mean values of the coefficients of the local model are closer to their global values. For instance, the range of model (1)'s correlates for the "*Diff. in the No. of Cases Involving Commercial Units*" variable is between 0.039 and 0.042 with a mean of 0.041, which is closer to its global coefficient value of 0.040.

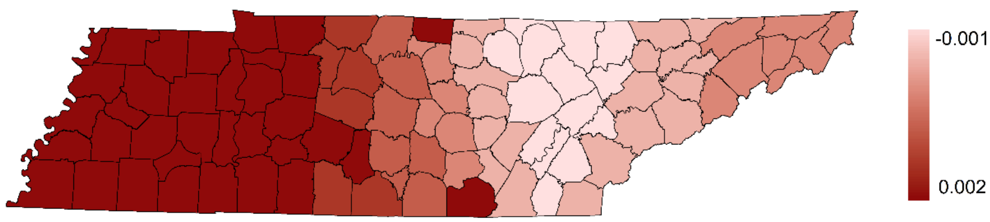
The maps showing the spatial variation of local parameter estimates of Tennessee counties are illustrated in Figure 3-8 and Figure 3-9. The average value of the local parameter estimates is calculated for each county using the Geographic Information System (GIS) software. A darker shade presents the higher values of the coefficients, and lower values are presented by a lighter shade. The maps show that the local parameter estimates vary across the counties of Tennessee. Figure 3-8 and Figure 3-9 show that correlates can be partially stationary in some counties but change across jurisdictions. For example, Figure 3-8(a) and Figure 3-9(a) show the correlates of "*Diff. in the No. of Speeding Violations"* for the differences in fatalities and crash harm that vary across Tennessee. This indicates the impact of speeding, as a positively correlated variable for crash fatalities and crash harm, is higher in West Tennessee. Going from west to east, except for a few counties, the effect of speeding differences decreases significantly. One thing can be presumed that lack of enforcement in the counties with higher estimates could play a role. However, an inverse relationship between speeding and the outcome variables is observed in some counties. One explanation could be that speeding and other violations do not result in fatalities necessarily in some counties, partly because they may not have high-speed roads (e.g., freeways) or differing levels of traffic enforcement (which cannot be captured in these data). Several studies have found similar findings (*89-91*). According to Imprialou et al. (*91*), the increased design standards of some roadways and the longer available distances between vehicles at high-speed conditions (i.e., lower traffic volume) may contribute to the inverse relationship between speeding and crashes in some regions. In Figure 3-9(b), the difference in *Alcohol & Drugs Cases* shows high variability in the south-western regions, e.g., Shelby County, where the difference in crash harm increases with the increase in the differences in *Alcohol and drug*-induced cases between 2020 and 2019. Differences in interstate-related fatalities and economic harm, in Figure 3-8(c) and Figure 3-9(d), are also higher in the western counties of Tennessee. Similarly, all the remaining variables in Figure 3-8 and Figure 3-9 support these assertions. Overall, these spatial variations can be due to deviations in traffic, roadway conditions, socio-economics, and other unobserved factors related to spatial contexts.

3.9 Discussion

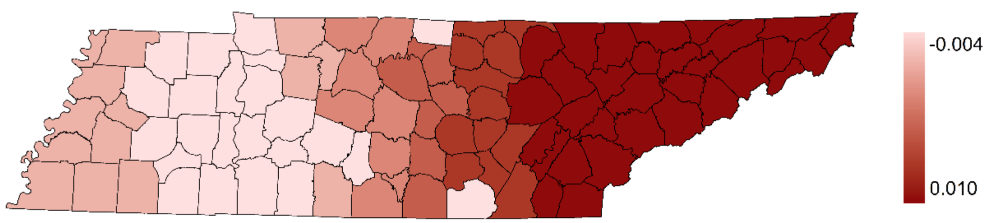
The increase in the number of crash fatalities during the pandemic is associated with the increased differences in the number of violations, including speeding, reckless driving, and alcohol & drugs cases, between 2020 and 2019. With more people staying at home during COVID-19, motorists have opportunities to drive on the near-empty streets. The combination of risky drivers and less congested roads may increase the chance of fatalities in a crash. Tefft et al. (*92*) indicated that risky driving behaviors might be attributable to a small subset of young drivers who have an increased propensity to drive during COVID-19, whereas safer drivers lowered their driving. The finding of this study is aligned with Inada et al. (*58*), and Hughes et al. (*56*), who found the number of fatalities is positively associated with the increased frequency of speed-related violations during COVID-19 in Japan and California, U.S.A., respectively. Speeds were found to increase substantially compared to the forecasted evolution (*93*). Dark-not-lighted roads bring greater danger during COVID-19, which is consistent with Adegbite et al. (*82*), who found that dark-not-lighted roads are responsible for about 31% of intersection crashes and fatalities. Moreover, during COVID-19, crash fatalities happened more on interstates and with the involvement of commercial trucks (*94*). The surge in online delivery during COVID-19 increased commercial trucks' mileage compared to other vehicles (*43*). Fatigue and the urgency of delivering goods and services accompanied by the reckless behaviors of drivers may contribute to their increased involvement in road fatalities during COVID-19. The results further show that more mid-length and long trips per population not staying at home induces fatalities in crashes. This is relatable to Zhang et al. (*95*), who found the average person miles traveled to be positively associated with the person involved in crashes during COVID-19. It may be because traveling longer distances might urge the drivers to speed up and go to the desired places in a short time, given the less congested roads during COVID-19. Total crash harm was reduced during COVID-19 in 2020; however, crash harm per trip went up in 2020 compared to 2019. Fatalities constitute the majority of the crash harm costs. Since the fatalities soared during COVID-19, the economic harm per trip has increased as expected. Moreover, the difference in crash harm is associated with the increased differences in violations, interstate and commercial trucks involved crashes, dark-not-lighted road crashes, and the increased number of long trips per population not staying at home. In addition, GWR models found spatial variation in several parameter estimates, including the differences in speeding violations, alcohol & drugs cases, and the cases on interstate roadways, in Tennessee at the county level. The correlates of these variables are mostly found to be higher in the western regions than the eastern regions of Tennessee, suggesting different levels of enforcement, roadway conditions, other built-in environment, and some other unobserved factors, e.g., cultural diversities (*88*).



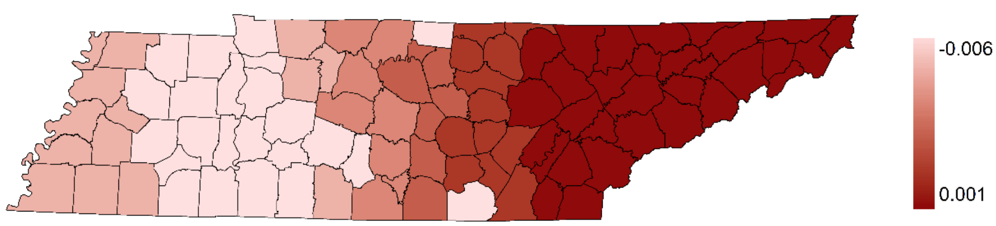
(a) Diff. in the No. of Speeding Violations



(b) Diff. in the No. of Day-Lighted Cases

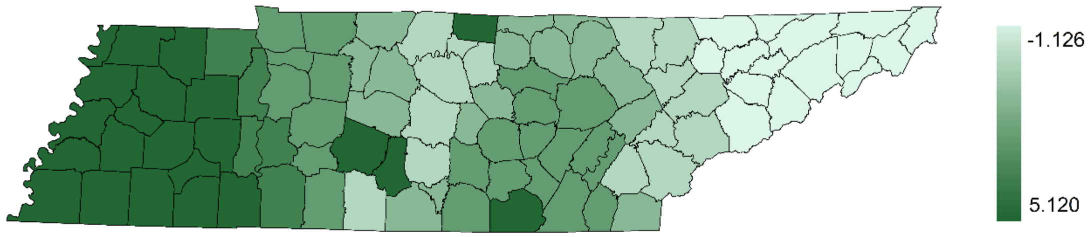


(c) Diff. in the No. of Cases on Interstate Roadways

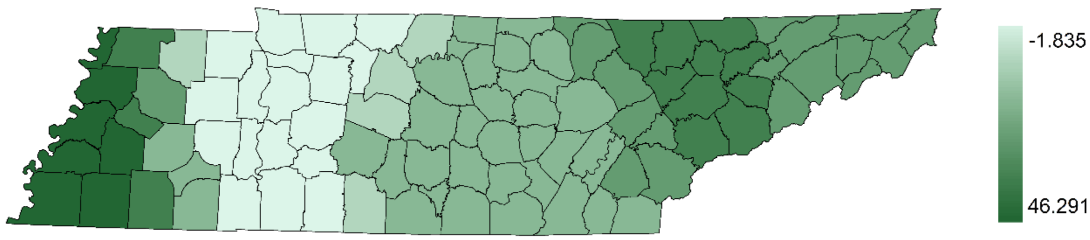


(d) Diff. in the No. of Cases on All Other Roadways

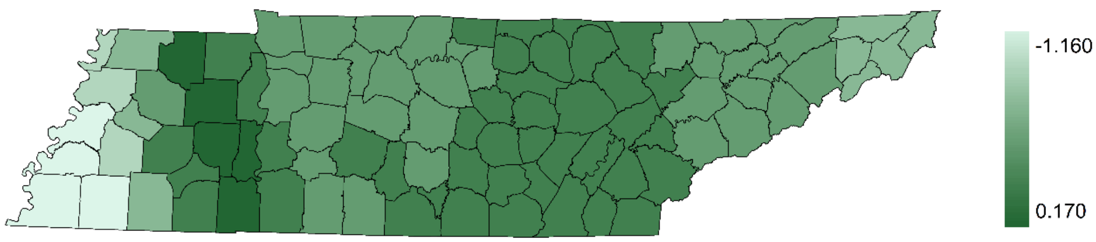
Figure 3-8: Spatial variation of local parameter estimates for the difference in the number of fatalities in Tennessee at the county level



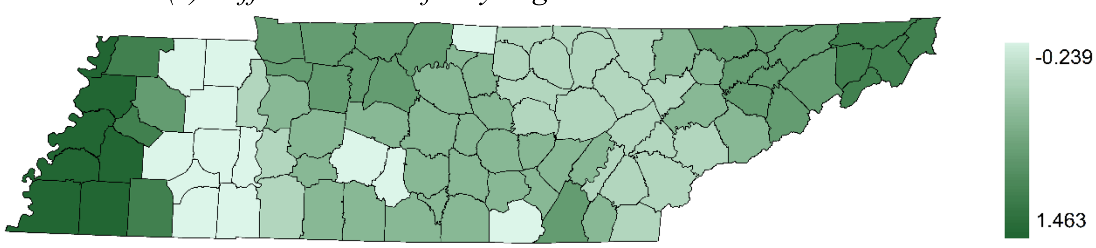
(a) Diff. in the No. of Speeding Violations



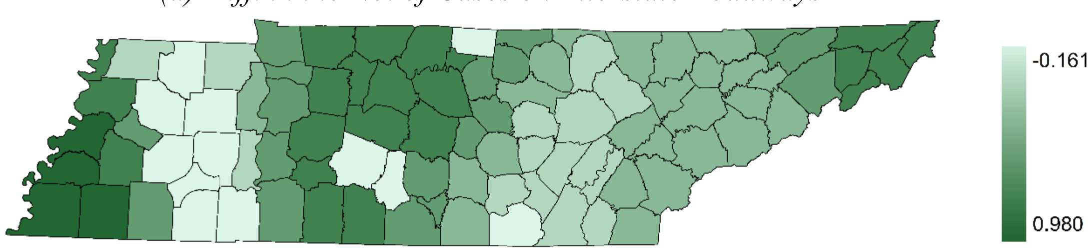
(b) Diff. in the No. of Alcohol & Drugs Cases



(c) Diff. in the No. of Day-Lighted Cases



(d) Diff. in the No. of Cases on Interstate Roadways



(e) Diff. in the No. of Cases on All Other Roadways

Figure 3-9: Illustration of the spatial variation of local parameter estimates for the difference in crash harm across the State of Tennessee

3.10 Limitations

This research is not without limitations. Estimates in the Bureau of Transportation Statistics' "Trips by Distance" database are relatively new and scantly peer-reviewed, which may potentially serve as a source of bias. Also, these data are experimental and may not have the highest quality standards. However, recent reports and studies are starting to cite this source. The inclusion of pre-existing regional characteristics, such as weather and terrain information, could have added more insights into the analysis. Although this study applied a framework for spatial heterogeneity or non-stationarity estimation, the results of the study may not be applicable to other states because of differences in geography, roadway network, and socio-economics.

3.11 Conclusion

The COVID-19 pandemic has impacted the whole world, including the transportation sector. The number of fatalities in crashes has increased in the US despite a significant reduction in traffic flow. The emphasis of this study is to use a comprehensive set of safety measures and assess what happened to road safety in Tennessee during COVID-19 by exploring the contributing factors. The findings are based on a unique dataset linking crash data and COVID-19 travel behavior data. The results show that while fatalities and crash harm per trip increased on roadways, there was still a reduction in total crashes and total monetary harm. Additionally, several models, including generalized least squares linear, Poisson, and geographically weighted regression models using the differences between 2020 and 2019 values, are adopted to rigorously quantify correlates of crash fatalities and crash harm.

The modeling results show that the difference in the number of crash fatalities between 2020 and 2019 is associated with the increased differences in violations, including speeding, reckless driving, and alcohol & drugs cases. Fatal crashes are more likely to occur on interstates and dark-not-lighted roads and involve commercial trucks with the surge of online delivery during COVID-19. Importantly, these similar factors largely contribute to the overall crash harm. In addition, more long trips per person not staying at home during COVID-19 are associated with more crash fatalities and more crash harm at the county level. GWR models show that several correlates of fatalities and crash harm are spatially varied across the counties of Tennessee. Importantly, the parameter estimates for the difference in speeding cases are higher in the western regions of Tennessee and lower in the eastern regions. This variation suggests different levels of traffic enforcement, socio-economics, roadway, traffic, and other built-in environmental factors.

The study findings may help safety practitioners better understand the factors contributing to crashes and fatalities, even during a safety-critical event like the COVID-19 pandemic. Reducing the violations identified in this study may lower the number of crashes, fatalities, and, eventually, the overall crash harm. Regarding traffic enforcement, more effort should be given to preventing risky driving behaviors, including speeding, reckless driving, and night driving, especially during a global emergency like COVID-19 when the traffic volume is lower and these behaviors are more commonplace. Proper countermeasures may help to improve road safety. Speeding-related violations may be reduced through speed camera enforcement, reducing the speed limit in hotspots, placing more warning signs, and using vehicular technology, e.g., intelligent speed adaptation (ISA) (*96; 97*). Furthermore, as suggested in the literature, automated vehicles (AVs) and big data applications have the potential to improve road safety in these aspects

(*98-100*). For example, the Cincinnati crash analysis reduction strategy (CARS) is a big data-oriented approach designed to identify dangerous crash hotspot locations, unravel the persistent crash contributing factors, and provide flexibility to explore strategies to reduce traffic crash harms (*101*).

Future researchers may investigate whether mobility to a specific location or for a certain activity is related to the increase in fatalities or not. The research can be extended with the inclusion of some key spatial variables, e.g., roadway traffic, weather information, terrain, etc., that may effectively reveal the role of the regional built environment (pre-existing characteristics) in the occurrence of crashes and fatalities. In addition, the use of daily time-series data for all the state counties of the US may provide more insights with the consideration of the weekend and weekday aspects. Overall, collective efforts by researchers and public and private sectors are required to gather more related data and develop road safety strategies concerning the new reality of the COVID-19 pandemic.

3.12 Acknowledgments

This study was partially supported by the Collaborative Sciences Center for Road Safety, CSCRS, a U.S. DOT-funded National Transportation Center, under Grant No. 69A3551747113. The authors thank Meredith King for providing proofreading assistance.

CHAPTER 4 INVESTIGATING TRANSPORTATION SAFETY IN DISADVANTAGED COMMUNITIES BY INTEGRATING CRASH AND ENVIRONMENTAL JUSTICE DATA

A version of this chapter was originally published by A. Latif Patwary, Antora Mohsena Haque, I. Mahdinia and Asad J. Khattak in the Journal of Accident Analysis and Prevention:

Patwary, A. L., Haque, A. M., I. Mahdinia, & Khattak, A. J. (2023). *Investigating Transportation Safety in Disadvantaged Communities by Integrating Crash and Environmental Justice Data.* Accident Analysis and Prevention, Vol. 194, 2024, p. 107366.

4.1 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. Census tracts with elevated traffic volume, higher levels of binge drinking, and the absence of mobile phone laws while driving exhibited higher rates of 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. Conversely, 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.

4.2 Introduction

To achieve the Vision Zero goal, the United States Department of Transportation (USDOT) has formulated the 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. 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 (*102*). 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) (*103*). 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 (*104*). 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 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.

4.3 Literature Review

The main focus of safety improvement studies has been analyzing the relationship between crashes and various features like driver behavior (*105-109*), vehicle features (*110*), roadway characteristics (*78*), traffic condition (*111; 112*), weather (*113; 114*), land use (*115*), 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 (*116-120*). Also, exposure measures such as high average daily traffic (ADT) and vehicle miles traveled (VMT) are positively associated with fatal crashes (*121; 122*). 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 (*123-127*). 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 (*123*). The residence characteristics of an at-fault-drivers-based study found that the percentage of people who work from home and commuters who commute for less than 15 minutes is negatively associated with the number of at-fault drivers. An increase in population is positively associated with at-fault drivers (*124*). Whereas studies found that high fatality and severe injuries are seen in lower-income regions in the US (*126*). Similar associations are observed in other countries. For instance, Christie et al. (*128*) found in the UK that the residents of the most deprived regions are five 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 (*126; 129*).

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 (*130; 131*). 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 (*131*). 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 (*132*). 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 1999-2006, the Hispanic population was the highest among all age groups to suffer from alcohol-impaired driving deaths in the US (*133*). 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 (*134*).

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 (*135*). The increased pedestrian activity can lead to more conflicts between vehicles and pedestrians. Additionally, research by Jacobsen and Rutter (*136*) 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 (*137*) 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 (*135*), 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 (*138*). Numerous studies have emphasized the role of sufficient bicycle infrastructure in reducing crash risks (*139-141*). 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 (*142*). 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. (*143*) 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 (*144*) 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 (*145*).

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% (*146*). 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 (*147*).

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 (*148*). 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 (*149*). 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 (*150*). Nonetheless, several studies have incorporated Hurdle models in the crash investigation (*150-153*). For instance, Hosseinpour et al. (*152*) 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. (*150*) 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 (*104*). 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.

4.4 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 (*104*). Fatal crash information is retrieved from the Fatality Analysis Reporting System (FARS) website, and some socio-demographic information is collected from the US Census Bureau's website. A brief overview of the datasets is provided below:

4.4.1 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 (Figure 4-1, Figure 4-2). These six broad themes are termed disadvantage indicators throughout the paper. The short definitions of them are provided below:

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 the CDC Social Vulnerability Index, Census America Community Survey, and FEMA Resilience Analysis & Planning Tool.

2. Environmental Disadvantage

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

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.

4. Health Disadvantage

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

5. Resilience Disadvantage

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

6. Transportation Disadvantage

This theme identifies communities affected by more prolonged and expensive ways of transportation. The data for this theme is collected from the 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 50^{th} percentile (75^{th} percentile for resilience) across more than three of the six themes. The list of those indicators is provided in Table 4-1. For details on these indicators, refer to USDOT's website (*104*).

The Disadvantage indicator-based dataset has information for 72843 census tracts but no tract information for the states 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 4-2 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 Figures 4-1 and 4-2. Environmental DACs are concentrated in the southeastern and western regions of the US, while transportation DACs are distributed throughout the country.

4.4.2 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.

4.4.3 Fatal Crash-based Dataset

The 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.

4.4.4 Data on Other Variables

Several other variables can provide valuable insights into the 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 (*154*). 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 (*155*). 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 (*156*). Alcohol consumption per capita is obtained from the National Institute of Alcohol Abuse and Alcoholism (NIAAA) of the National Institute of Health (NIH) (*157*). Mobile phone use law data by state is gathered from the Insurance Institute of Highway Safety (IIHS) (*158*). Finally, seat belt usage information for 2021 is collected from the National Traffic Safety Administration (*159*). Incorporating these state-level variables can give a more comprehensive understanding of the analysis.

4.5 Methods

4.5.1 Data Processing

The study framework is illustrated in Figure 4-3. In the Disadvantaged indicators dataset, GEO_IDs are developed similarly to the GEO_IDs observed in the 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% of 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.

4.5.2 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

Table 4-1: List of Disadvantage Indicators under Six themes (*160*)

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

Table 4-1 continued

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

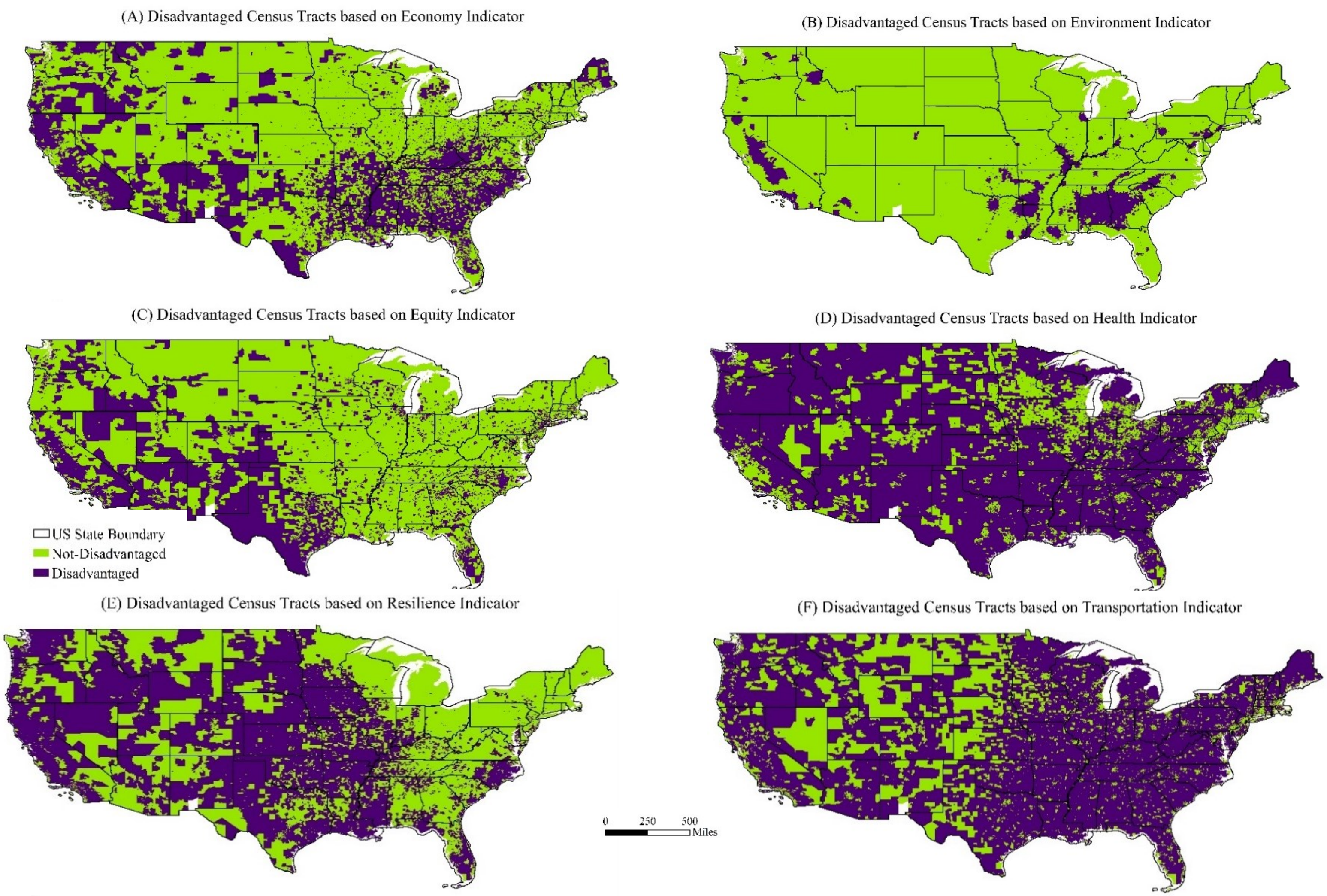


Figure 4-1: Disadvantage Census Tracts based on the Six Disadvantage Indicators

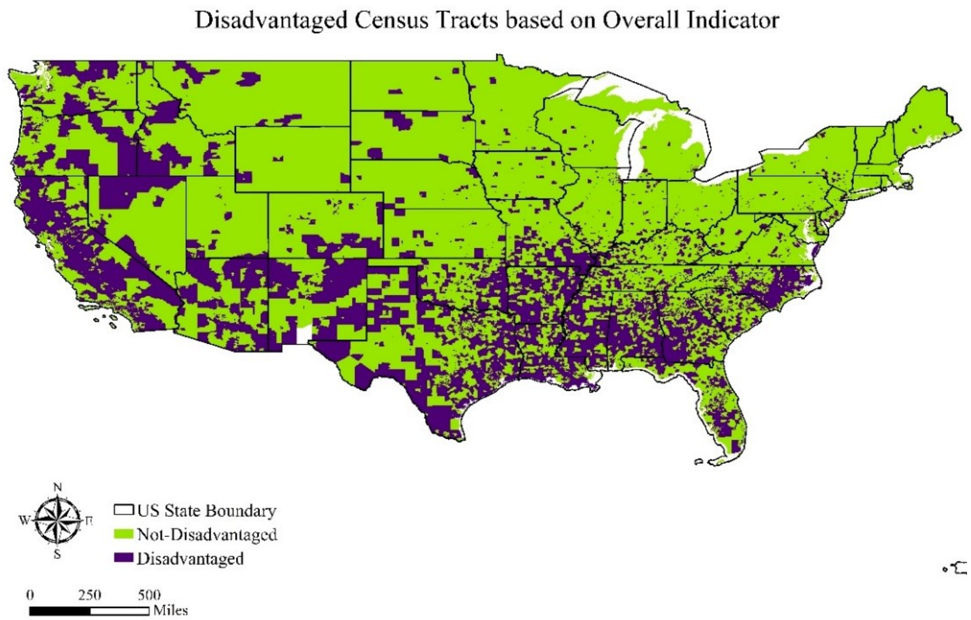


Figure 4-2: Disadvantage Census Tracts based on the Overall Disadvantage Indicator

Table 4-2: Area and Population in Census Tracts

Disadvantage Indicator	Census Tract Status	Census Tracts	Census Tract (%)	Area (Sq. Miles)	Area (%)	Population	Pop. (%)
Economy	Disadvantaged	34,253	47.05	514.42	46.67	145,088,086	44.71
	Not-Disadvantaged	38,543	52.95	587.79	53.33	179,393,640	55.29
Environment	Disadvantaged	36,008	49.46	66.5	6.03	159,076,339	49.02
	Not-Disadvantaged	36,788	50.54	1035.7	93.97	165,405,387	50.98
Equity	Disadvantaged	36,067	49.55	276.29	25.07	171,885,819	52.97
	Not-Disadvantaged	36,729	50.45	825.91	74.93	152,595,907	47.03
Health	Disadvantaged	36,063	49.54	815.22	73.96	145,076,075	44.71
	Not-Disadvantaged	36,733	50.46	286.99	26.04	179,405,651	55.29
Resilience	Disadvantaged	18,173	24.96	497.47	45.13	98,263,387	30.28
	Not-Disadvantaged	54,623	75.04	604.74	54.87	226,218,339	69.72
Transportation	Disadvantaged	35,323	48.52	615.05	55.80	159,111,566	49.04
	Not-Disadvantaged	37,473	51.48	487.16	44.20	165,370,160	50.96
Overall	Disadvantaged	21,938	30.14	241.83	21.94	99,639,881	30.71
	Not-Disadvantaged	50,858	69.86	860.38	78.06	224,841,845	69.29
Total		72,796	100	1102.21	100	324,481,726	100

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 (*149*). 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-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 the 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:

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

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

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

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

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 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 the 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.

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

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

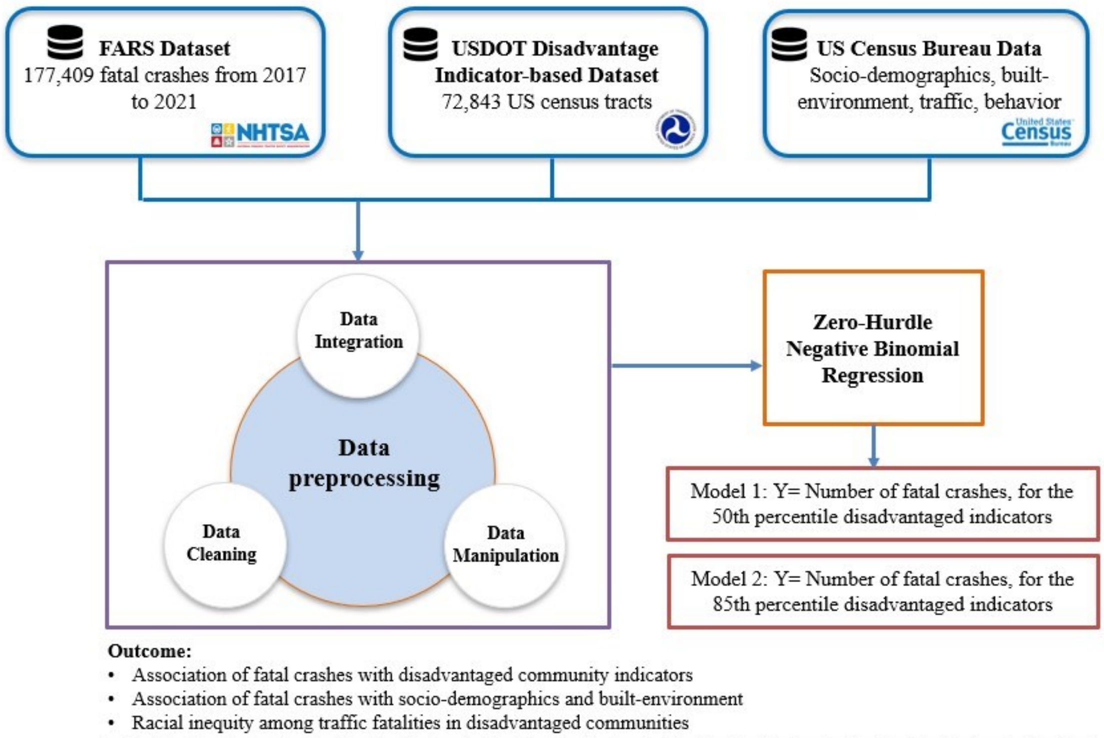


Figure 4-3: Study Framework

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 (*161*). 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.

4.6 Results and Discussion

4.6.1 Descriptive Statistics

Table 4-3 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 some 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

Table 4-3: 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

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 the *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 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.

4.6.2 Hurdle Models on All-Fatal Crashes in the US

Table 4-4 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 or higher than -0.25, as shown in Table 4-5. However, the correlation between median income and economy DAC is slightly lower than -0.5.

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 and 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. Also, the pseudo-R^2 is slightly lower in the 85th percentile model (18%) than the 50th percentile model (19%). Hence, the 50th percentile-based model is selected for the full dataset. The Hurdle model with 50th percentile-based disadvantage indicators is explained, and the 85th percentile-based model is shown for comparison. Figure 4-4 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-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 track 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 (*162*). 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 1-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. (*163*), 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

Table 4-4: Hurdle Models Results

Hurdle Count 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 Models						
Constant	0.4787	9.53	0	0.4787	9.53	0
Population	0.0001	11.15	0	0.0001	11.25	0
Work from home percentage	-0.0404	-26.28	0	-0.0404	-26.28	0
Model Fit Statistics						
AIC	280304			282592		
BIC	280497			282785		
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)	71728			71728		
Non-zero observations	51061			51061		
Zero Observations	20667			20667		

Table 4-5: Correlation coefficients between median income and the disadvantage indicators

	Median Income	Transport	Health	Economy	Equity	Resilience	Environment
Median Income	1.00						
Transport	-0.29	1.00					
Health	-0.40	0.18	1.00				
Economy	-0.56	0.17	0.19	1.00			
Equity	-0.07	-0.06	-0.12	0.24	1.00		
Resilience	0.04	0.03	0.03	-0.05	0.00	1.00	
Environment	0.00	-0.09	-0.12	0.21	0.24	-0.06	1.00

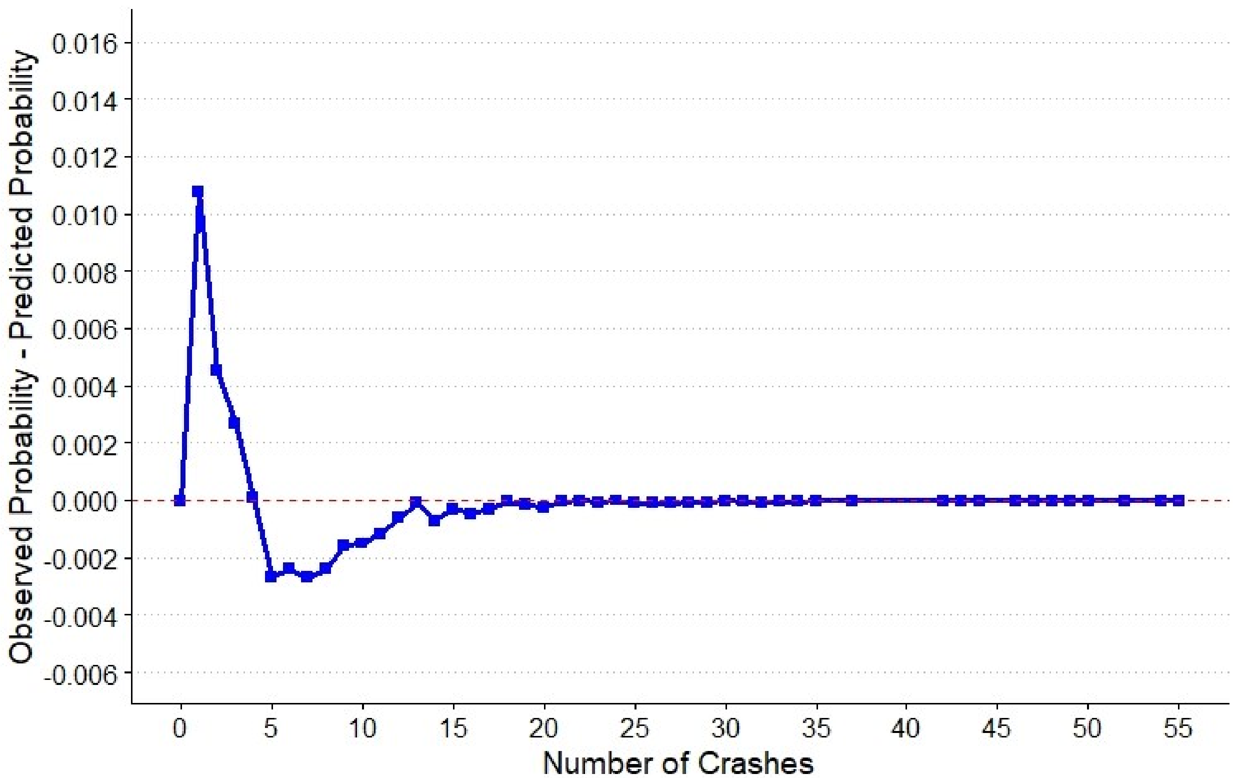


Figure 4-4: Difference between observed and predicted probabilities of crash occurrence

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 associated with 5.15% fewer fatal crashes. This is similar to the findings by Lim and Chi (*164*), 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. (*112*) 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 (*165-167*). According to Murphy et al. (*166*), 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 (*165*). 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. (*168*) 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. (*169*), 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 (*170*) 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* has a positive coefficient. This indicates that if the population of an area increases, the odds of a non-zero fatal crash increases. The work-from-home percentage has a negative coefficient of −0.0404. It implies that an increase in working from home population is associated with lower odds of observing a non-zero fatal crash count. However, although working from home eliminates the need for commuting, it encourages non-work travel, including shopping trips, restaurant trips, and recreational trips (*43*).

4.6.3 Safety in Disadvantaged Communities by Race

Figure 4-5 represents the number of fatalities per 100,000 population per year by race. It shows that in the disadvantaged tracts, the highest fatality per 100,000 per year 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. Figure 4-6 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.

4.7 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 identifies traffic safety risks in disadvantaged communities.

4.8 Conclusions

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

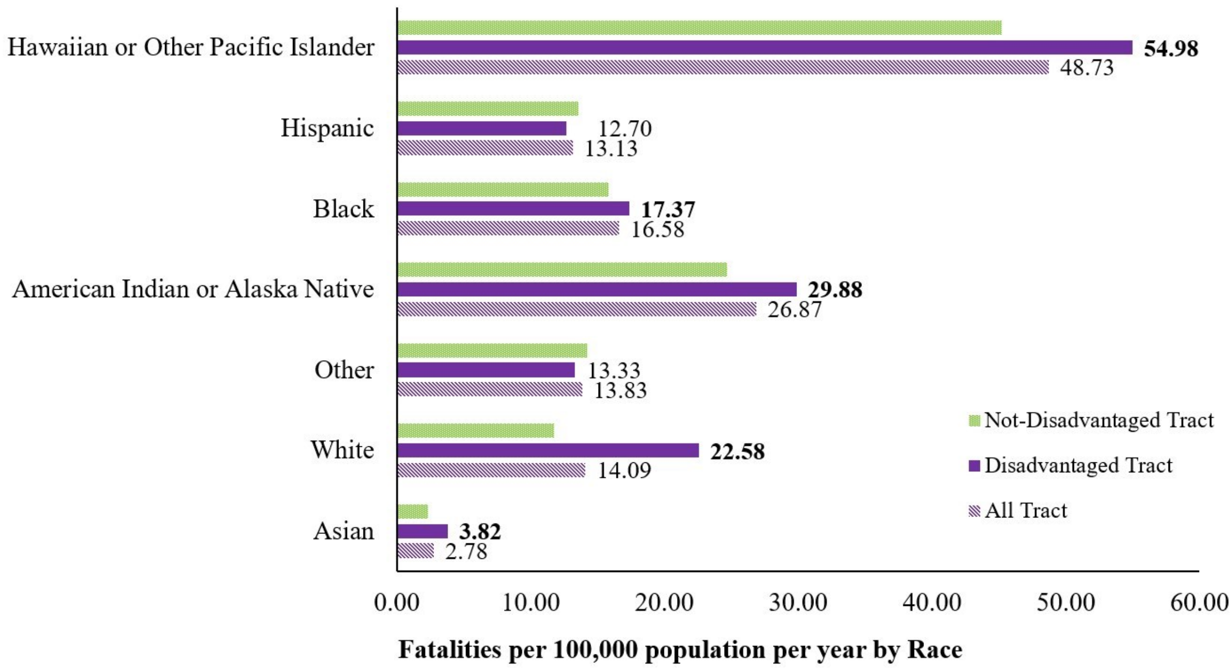


Figure 4-5: Fatalities per 100,000 population per year across different races in the US

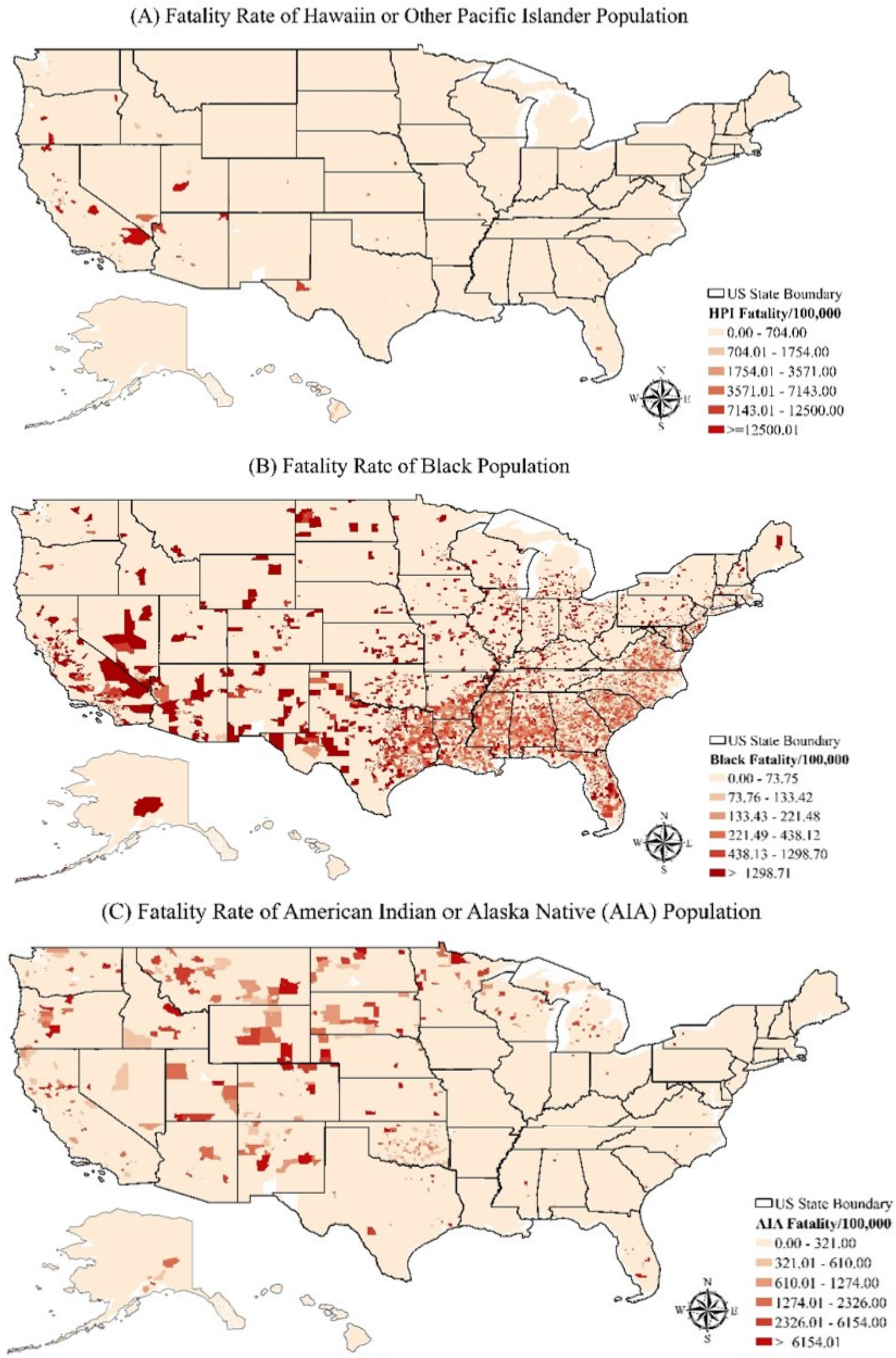


Figure 4-6: Fatality Rate of (a) HPI (b) Black and (c) AIA Population in the US

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.

4.9 Acknowledgments

The authors would like to thank the Collaborative Sciences Center for Road Safety for providing financial support for this article's research, authorship, and/or presentation/publication.

CHAPTER 5 INTERACTION BETWEEN THE EMERGING COMPONENTS OF ONLINE SHOPPING AND IN-PERSON ACTIVITIES: INSIGHTS FROM BEHAVIORAL SURVEY AND JUSTICE40 INITIATIVE DATA

A version of this chapter was originally submitted by A. Latif Patwary and Asad J. Khattak for publication:

Patwary, A. L., & Khattak, A. J. (2023). *Interaction between the Emerging Components of Online Shopping and In-Person Activities: Insights from Behavioral Survey and Justice40 Initiative Data.*

5.1 Abstract

The rise of technological advancements has led to the commonplace practice of online shopping for retail, grocery, and food. However, little research has been conducted on the interplay of these components in disadvantaged communities (DACs) that face issues of marginalization and limited access to digital resources. This study aims to provide a comprehensive understanding of travel behavior changes by analyzing the interconnectedness of the emerging components of online shopping (retail, grocery, and food) and in-person activities in both DACs and non-DACs. A unique household-level database is created by linking the 2021 Puget Sound Household Travel Survey and the US Department of Transportation's Justice40 databases, and a conditional mixed process model is estimated to account for unobserved endogeneity. The findings suggest that households living in DACs are less likely to order online retail goods and groceries than non-DAC households. Additionally, the probability of making more restaurant trips decreases for households living in DACs. The study highlights the digital divide that exists in DACs and the differences in online and in-person shopping activities across socioeconomic levels. Policymakers can address these disparities to promote equity and equal access to goods and services for all. Besides, planners may need to improve the travel demand models by accounting for the emerging components of online shopping and the trip frequencies by purpose in DACs.

5.2 Introduction

The advancement in information and communication technologies (ICTs) makes many activities easily accessible from home that previously needed a fixed location. In the US, the share of people having smartphones has increased by about 70% in 2021 (*171*). A combination of the internet, smartphones, tablets, laptops, smart service providers, innovations, and widely improved wireless communications allows individuals to access goods and amenities anytime from anywhere. Therefore, the partial decoupling of virtual access and physical trips to access the service will likely continue in the future (*1; 2*). Virtual interactions can take different forms, including online shopping, working from home, telemedicine, etc.

The COVID-19 pandemic, with the technological landscape and the types of services provided by connectivity, have significantly changed our travel behavior (*172-175*). In the US, vehicle miles traveled (VMT) has decreased as fewer people were traveling, especially during 2020 at the peak of COVID-19. However, freight activity has shown an upsurge during the pandemic. Demand for freight truck drivers to restock goods at the beginning of the pandemic has been followed by unstable delivery demands because of the slowing and then the resurgence of the economy and the surge in online delivery of goods and services. E-commerce sales raised by 17% in 2019, and sales surged by 37% in the third quarter of 2020 (*5*). On average, online retail sales have increased by 15-30% each year in the past decade (*176*) and are expected to continue to grow in the coming years because of its convenience, much wider choices, and price flexibility over traditional physical stores (*177*).

This surge in online shopping can significantly affect an individual's travel behavior. Numerous prior studies have examined the relationship between online shopping and the in-store shopping behavior of consumers (*10; 16*). It is either found online shopping reduces shopping trips (substitution) or increases (complementary). However, the many ways online and physical shopping interact may go beyond the traditional substitution and complementarity framework. Most studies draw conclusions after accounting for online shopping as a single category. The shopping behavior during COVID-19 shows that online shopping may have several components. For instance, grocery shopping behavior differs from shopping for other personal or household items, which drove many purchases during the pandemic. Online grocery sales were predicted to be worth $200 billion in 2022 (*178*). Purchasing and delivering prepared meals is another component of online shopping that emerged fully in the wake of the pandemic, and it is predicted to grow from 6% in 2018 to 13% in 2025 (*179*).

People who live in an urban area generally have higher accessibility to online and in-store shopping and dining in physical restaurants. However, the effects are relatively unknown if the area is disadvantaged in terms of lacking the required infrastructure, operations, and investments. Meanwhile, the disadvantaged communities (DACs) that are underserved, marginalized, and polluted are receiving 40% of the benefits of federal investments (*180*). It is important to understand how these disadvantaged communities' online shopping components and physical shopping behavior are interconnected. Therefore, this study aims to investigate the relationships between the online and in-person activity engagements for the three shopping components (i.e., food, grocery, and retail) across DACs and non-DACs. It further attempts to explore the correlates of the socio-demographics, locational, and travel attributes on shopping behavior at the household level.

5.3 Literature Review

Virtual shopping behaviors can potentially substitute physical activities that previously required physical travel. This may as well stimulate more physical travel (*4; 181*). Numerous studies have investigated the relationship between online shopping and physical trips (*10; 16; 182*). In addition to the well-known online delivery of durable goods, the emergence of newer components of online shopping (i.e., food and grocery) has received wider attention during the COVID-19 pandemic. This section provides a synthesis of previous studies on the three components of online shopping and their influencing factors, including socio-demographics, spatial, and travel-related factors.

5.3.1 Online/Retail Shopping

Online shopping is of interest to many policymakers, engineers, and planners. Past research mainly analyzed online shopping as a single activity or focused on shopping for durable goods and services that are bought less frequently, e.g., clothing, household items, kitchen utensils, electronics, books, and other special items (*36*). Regarding the decoupling of virtual and fixed activity space, there is a long-ongoing debate on the association between online shopping for durable goods and physical shopping trips. Some studies suggested it as substitutive, and others found it complementary. Sim and Koi (*183*) found that 12% of online shoppers reduced their shopping trips while investigating the travel behavior of 1500 consumers in Singapore. It may be because some people do not need to make physical trips to stores while they can shop online. On the other hand, Cao et al. (*10*) suggested that online shopping increases shopping trips in Minneapolis using survey data from 539 individuals. However, some other studies also found both substitution and complementary effects. For example, Zhou and Wang (*16*) analyzed the bidirectional relationship between online shopping and shopping trips using the 2009 National Household Travel Survey (NHTS) database with a structural equation model. They found that more online shopping encourages more shopping trips, and frequent shopping trips tend to reduce online shopping frequencies. While analyzing the factors affecting this relationship, studies listed some related contributing factors, including income, household size, the number of household vehicles, household race, urban location, and car sharing status, among others (*10; 16; 184*). However, it is important to examine online shopping after dissecting its distinctive components (food and grocery shopping).

5.3.2 Food and Grocery Shopping

The history of food and grocery shopping online dates back to the early 1990s, when Pizza hut began its Pizza delivery services and Peapod (the first company that introduced online grocery shopping) started its online delivery (*178*). Over the years, many platforms, e.g., Uber-Eats and Grub hub, Amazon Whole Food Market, Walmart+, Instacart, etc., have emerged with the advancement of ICT applications. While the literature on prepared meals/food shopping is quite limited, a recent study suggests that a large household is 15% more likely and a higher-income household is 12% more likely to receive prepared meal deliveries. However, these activities did not reduce the total number of personal trips (*185*).

A few studies have analyzed in-store and online grocery shopping behaviors (*186-190*). Research has been conducted on the underlying motives, situations, and other related factors of consumers' online grocery shopping behavior (*187*). Customers usually order groceries online to save time and if any item or specialized items are unavailable at the grocery store (*191*). Income, age, household location, and structure are found to affect the frequency of online grocery shopping (*188*). Households having more members with a greater share of older individuals tend to buy groceries from online stores more, as it is easier to order more supplies and not carry them from a grocery store for such a large household (*192*). Recently, Kim and Wang (*178*) analyzed the factors affecting the components of online shopping, including retail, grocery, and food. They found that the factors affecting different online delivery types differ with the trip mode. However, they ignored the interconnection between different types of delivery and associated physical travel behaviors. For instance, they ignored that more food shopping could result in less grocery shopping online. Dias et al. (*36*) also attempted to jointly analyze the components of online shopping using a multivariate probit model. However, they only used a few household attributes, e.g., household income, household size, and homeownership. Other variables that represent transportation accessibility, employment, emissions, and inequalities in terms of wealth, development, and infrastructure investment, of the community they live in, and their travel attributes are ignored. Besides, their study sample size is small and could not capture the COVID-19 shocks and online food & grocery shopping surge.

Overall, from the above-mentioned studies in this review, it is evident that there remain a few clear gaps in the literature. One major gap is how online shopping and travel behavior changes in a DAC have not been investigated. It is important to explore these since DACs may not access online ordering services due to the digital divide. Moreover, most studies considered online shopping as a single activity without drawing any clear distinction and interaction between different shopping activities. Therefore, while filling the gaps, this study intends to contribute to the literature by analyzing the three components of online shopping (retail, grocery & food) and associated in-person travel for these activities across the DACs and non-DACs. It harnesses the most recent 2021 Puget Sound Household Travel Survey and US DOT's Justice40 databases with a conditional mixed process model that accounts for unobserved endogeneity.

5.4 Conceptual Framework

The interaction between the components of online shopping (retail, food, and grocery) and related travel behavior is not straightforward. Furthermore, disadvantaged communities, in terms of equity, environment, transportation access, health, economy, and resilience, may intervene in those interactions. The study framework in Figure 5-1 is depicted from the relationship found in the literature among the endogenous variables (in-person and online shopping for retail, grocery, and food purposes) (*36; 178*). These relationships are established through hypothesis testing of different combinations of endogenous variables (*36*). The framework is also supported by the recursive nature of different multivariate models (*193*). Importantly, the literature suggests that more online retail shopping (i.e., durable goods shopping) is associated with more in-person grocery shopping and in-person eating food at restaurants (*36*). Ordering prepared food online is associated with fewer physical trips for retail shopping to stores (*36*). The relationship between online and in-person shopping for retail goods can be either complementary or substitutive (*10; 16; 36; 43*). The complementary aspect indicates more in-person shopping is associated with a higher propensity for online shopping and vice versa for substitution. Besides, more in-person grocery shopping frequency is related to lower frequencies of online shopping. Additionally, according to the literature, these endogenous variables are expected to be associated with some exogenous features. Household attributes, e.g., higher income, large households, more household vehicles, and non-white households, are anticipated to be positively associated with online shopping (*16; 36; 178*). Urban location and tech-savviness are also expected to have a positive association. Finally, households living in DACs are generally expected to have a lower frequency of deliveries due to limited infrastructure investments, limited access to technology and other emerging online services.

5.5 Methodology

5.5.1 Data

The data sources of this study are the 2021 Puget Sound Household Travel Survey (*194*) and the US Department of Transportation's Justice40 database (*195*). A combined dataset of households is created after linking these databases. The travel survey collects information on socio-demographics, online delivery frequencies, locational, and travel information at the household and person level of the residents covering the Greater Puget Sound area from April-June of 2021. The survey is a standard web-based online survey (rSurvey) designed to collect complete household travel diary information from invited participants. The survey provides data on grocery, food, & retail delivery frequencies and the trip counts for a pre-assigned travel day.

The survey data comes with household, person, trip, days, and vehicle data files, which are

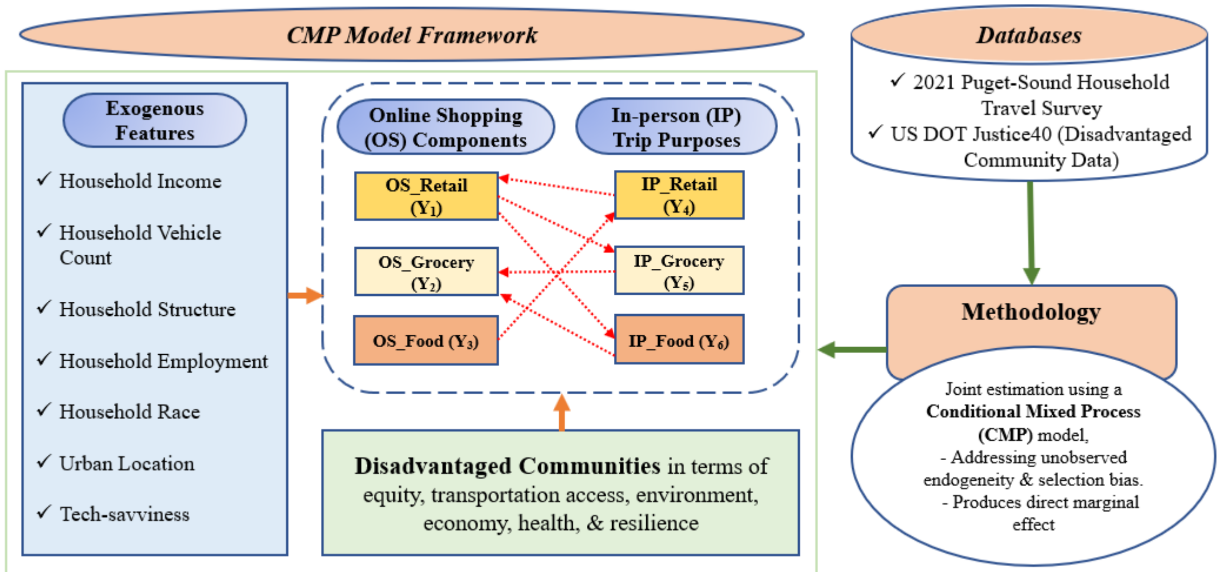


Figure 5-1: Study Framework

merged to produce household-level data. Data is cleaned by removing rows having missing information, e.g., skip logic, prefer not to answer. The final cleaned dataset has a sample size of 2,501 households. Table 5-1 reports the descriptive statistics of the cleaned data. The endogenous variables are food, grocery, retail deliveries on travel day, and the number of physical trips for retail, grocery, and restaurant on travel day. Exogenous variables include household structure, household employment, Household vehicle count, household income, race, Seattle city home, and tech-savviness. Table 5-1 shows that 27% of households receive retail delivery at least once, whereas only 4% receive food and grocery deliveries each. 11% of households take in-person retail trips on travel day at least once. On the other hand, 25% of households took grocery trips, and 13% took restaurant trips at least once on the travel day. 28% of the sample households have an income of less than $50,000, and 41% have an income higher than or equal to $100,000. Moreover, 51% are white population, and only 23% live in Seattle city. 40% of households have multiple workers, and 46% have two or more adults without children. 6% use a car share program. The household's participation in a car-sharing program (such as ZipCar, Car2Go, or GIG) is reflected in their car-sharing status. This status can show a connection with their travel habits and opinions towards new technologies, e.g., newer forms of online shopping. Therefore, it is considered a proxy for the household's tech literacy (i.e., tech-savviness).

The PHTS cleaned dataset is linked with the US DOT's Justice40 disadvantaged community (DAC) database. According to USDOT, DACs are the ones that are affected by high poverty, low wealth, low educational attainment, insufficient local jobs, high inequality, and low homeownership. This definition was developed to provide Justice40-based grant programs to the underserved communities of the US. The GIS-based mapping tool displays DACs at the census tract level for public use. The definition of DAC is based on six indicators that cover six broad themes of transportation disadvantages: economy, environment, equity, health, resilience, and transportation. The indicators are explained briefly below:

- *Economic Disadvantaged:* This indicator detects census tracts that are affected by high poverty, low wealth, low educational attainment, insufficient local jobs, high inequality, and low homeownership.
- *Environmental Disadvantaged:* This indicator detects census tracts that possess an unbalanced pollution burden and below-standard environmental quality.
- *Equity Disadvantaged:* This indicator detects those census tracts where a high percentile of the population possesses a "less than well" English speaking skill.
- *Health Disadvantaged:* This indicator detects the census tracts affected by adverse health outcomes, disability, and environmental exposures.
- *Resilience Disadvantaged:* This indicator detects those census tracts that are vulnerable to climate change hazards.
- *Transportation Access Disadvantaged:* This theme identifies communities affected by more prolonged and expensive ways of transportation.

Table 5-2 illustrates disadvantaged indicators for transportation and economy major categories. This shows a glimpse of how the major categories are calculated. Each major disadvantaged indicator is an aggregated index based on the theme's relevant indicators. A census tract is considered overall disadvantaged if it surpasses the 50th percentile (75th for resilience) across more than three of the six aggregated theme-based indicators. The disadvantaged indicator-based dataset has information for 72,843 census tracts. In the Seattle, Washington study area, 620

Table 5-1: Descriptive Statistics of the Data

Variables	Description	Freq.	%
OS_Retail*	Frequencies of retail delivery for durable goods on travel day		
	1 = No	1304	64%
	2 = 1 time	560	27%
	3 = 2 times	150	7%
	4 = 3+ times	37	2%
OS_Grocery*	Grocery delivery frequencies on travel day		
	1 = No	1946	94%
	2 = 1 time	81	4%
	3 = 2 times	13	1%
	4 = 3+ times	11	1%
OS_Food*	Food delivery frequencies on travel day		
	1 = No	1936	94%
	2 = 1 time	86	4%
	3 = 2 times	20	1%
	4 = 3+ times	9	1%
IP_Retail*	Retail trip counts on travel day		
	1 = No	1656	81%
	2 = 1 Trip	236	11%
	3 = 2 Trips	96	5%
	4 = 3+ Trips	63	3%
IP_Grocery*	Grocery trip counts on travel day		
	1 = No	1353	66%
	2 = 1 Trip	514	25%
	3 = 2 Trips	136	7%
	4 = 3+ Trips	48	2%
IP_Food*	Restaurant trip counts on travel day		
	1 = No	1649	80%
	2 = 1 Trip	269	13%
	3 = 2 Trips	102	5%
	4 = 3+ Trips	31	2%
Household Employment	0 = No workers	493	24%
	1 = Single worker	824	40%
	2 = Multiple workers	734	36%
Household Structure	0 = Single adult	566	27%
	1 = Two or more adults without children	936	46%
	2 = Two or more adults with children	549	27%

Table 5-1 continued

Household Vehicle Count	0 = No Vehicle	171	8%
	1 = Single Vehicle	838	41%
	2 = Multiple Vehicles	1042	51%
Household Income	0 = <$25,000	228	11%
	1 = $25,000-$49,999	356	17%
	2 = $50,000-$74,999	326	16%
	3 = $75,000-$99,999	294	15%
	4 = >=$100,000	847	41%
Household Race	0 = Others	833	41%
	1 = White	1218	59%
Home in Seattle City	0 = No	1587	77%
	1 = Yes	464	23%
Tech-savviness	0 = No	1933	94%
	1 = Yes	118	6%

tract's DAC information is linked at the household level (Figure 5-2). The descriptive statistics for the DAC categories in Table 5-3 suggest that 32% of the households in the sample live in the transportation DACs. 45% live in economic DACs, and 35% in the overall DACs. It is important to analyze the shopping behaviors in the DACs. This can help inform policies and initiatives aimed at promoting more inclusive economic growth, where all communities have access to the goods, services, and opportunities they need to thrive. The following chapters will analyze the cleaned dataset and discuss the results.

5.5.2 Model

The components of online shopping and household trips are defined as categorical ordered variables. The frequency of the deliveries is ordered, including "no delivery," "1 time", "2 times", and "3+ times". The in-person retail shopping trips, grocery shopping trips, and restaurant dine-in trips on travel day have four categories: "no trips" as the lowest category and "3+ times" as the highest category. We designate online food shopping as Y_1, online grocery shopping as Y_2, online retail shopping as Y_3, in-person retail trips as Y_4, grocery trips as Y_5, and restaurant trips as Y_6. The following empirical models can be represented mathematically:

$$Y_{1i} = \alpha_0 + \alpha_1 X_i + \delta Y_{4i} + \mu_{1i} \qquad \text{(i)}$$
$$Y_{2i} = \beta_0 + \beta_1 X_i + \delta' Y_{5i} + \delta'' Y_{6i} + \mu_{2i} \qquad \text{(ii)}$$
$$Y_{3i} = \gamma_0 + \gamma_1 X_i + \mu_{3i} \qquad \text{(iii)}$$
$$Y_{4i} = \tau_0 + \tau_1 X_i + r Y_{3i} + \mu_{4i} \qquad \text{(iv)}$$
$$Y_{5i} = \varphi_0 + \varphi_1 X_i + \omega Y_{1i} + \mu_{4i} \qquad \text{(v)}$$
$$Y_{6i} = \vartheta_0 + \vartheta_1 X_i + \omega' Y_{1i} + \mu_{4i} \qquad \text{(vi)}$$

Where X_i indicates a vector of exogenous variables allied with the household. These variables are hypothesized to have correlations with online food, grocery, retail shopping, and the household trips for the same activities on travel day. μ_1, μ_2, μ_3, μ_4, μ_5 and μ_6 are the related random error terms, and α, β, γ, δ, δ', δ'', r, φ, ω, τ, ϑ and ω' are the parameters to be estimated. The parameters δ, δ', δ'', r, ω, and ω' represent the correlations of 1) retail shopping trips with online retail shopping, 2) grocery shopping trips with online grocery shopping, 3) restaurant trips with online grocery shopping, 4) food delivery with retail shopping trips, 5) online retail shopping with grocery shopping trips, and 6) online retail shopping with restaurant trips.

Equations (i) through (vi) are ordered probit regression models that can be jointly estimated using conventional path analysis or structural equation modeling (SEM). However, unobserved endogeneity can lead to biased and unreliable estimates. Unobserved endogeneity occurs when a variable that affects both the dependent and independent variables is not included in the analysis. This is more difficult to detect and correct than observed endogeneity, which occurs when a measured variable influences both dependent and independent variables. For example, household employment can influence online shopping and household factors, and it can be included in the analysis. In the case of unobserved factors, for instance, the lifestyle of a household may influence shopping behavior. Households with a busy lifestyle may be more likely to shop online because it is more convenient. However, this factor may not be observable, leading to unobserved endogeneity. To address unobserved endogeneity, the conditional mixed process model (CMP) proposed by Roodman (*39*) can be used. The CMP considers the correlations between the error processes of two or more equations and provides a new perspective for adopting structural models with different equations to correct associated unobserved endogeneity issues.

Table 5-2: List of Disadvantaged Indicators under Transportation and Economy Themes (*160*)

Major Disadvantaged Category	**Percentile of Disadvantaged Indicators**
Transportation Access Disadvantaged (Average of the percentiles)	Percentile percentage households with no vehicle available estimate
	Percentile of Mean commute time to work (in minutes) (longer commute times)
	Percentile of Transportation Cost Burden: • Total workers 16 or older in a census tract • Percentages of non-transit households who have 0/1/2/3+ vehicles • Number of non-transit workers • Number of transit users 16 and over • Average weekday vehicle miles traveled per state • Calculated Average Annual Vehicle Miles Traveled • Average Annual Median Earnings • Five Year average price of gas per state • Five Year average gas mileage per state • Calculated average number of cars per household • Calculated average cost of owning a car • Calculated national average annual cost of using transit • Calculated average annual cost of transportation • Annual Travel Time in Minutes • Annual Travel Time in Hours • Travel Time Cost • Calculated average annual cost of transportation as a percent of income
	Percentile of National Walkability Index
Economy Disadvantaged (Average of the percentiles)	Percentile Percentage of persons with no high school diploma (age 25+) estimate
	Percentile Overall Renter Rate: Percent of Occupied Housing Units that are Renter-Occupied
	Percentile Percentage of civilian (age 16+) unemployed estimate
	Percentile per capita income estimate
	Percentile Percentage of persons below poverty estimate
	GINI Index Percentile (calculated)
	Percentile Percent of Household Units with 30 percent or more income towards housing cost

Table 5-3: Descriptive Statistics of the Disadvantaged Indicators

Disadvantaged Indicators	Status	Sample Household Freq.	Freq. (%)	Census Tract Pop. (%)
Transportation Disadvantaged	0 = No	1386	68%	65%
	1= Yes	665	32%	35%
Economic Disadvantaged	0 = No	1126	55%	70%
	1= Yes	925	45%	30%
Equity Disadvantaged	0 = No	624	30%	35%
	1= Yes	1427	70%	65%
Health Disadvantaged	0 = No	1399	68%	74%
	1= Yes	652	32%	26%
Environment Disadvantaged	0 = No	597	29%	35%
	1= Yes	1454	71%	65%
Resilient Disadvantaged	0 = No	80	4%	2%
	1= Yes	1971	96%	98%
Overall Disadvantaged	0 = No	1129	55%	65%
	1= Yes	922	45%	35%

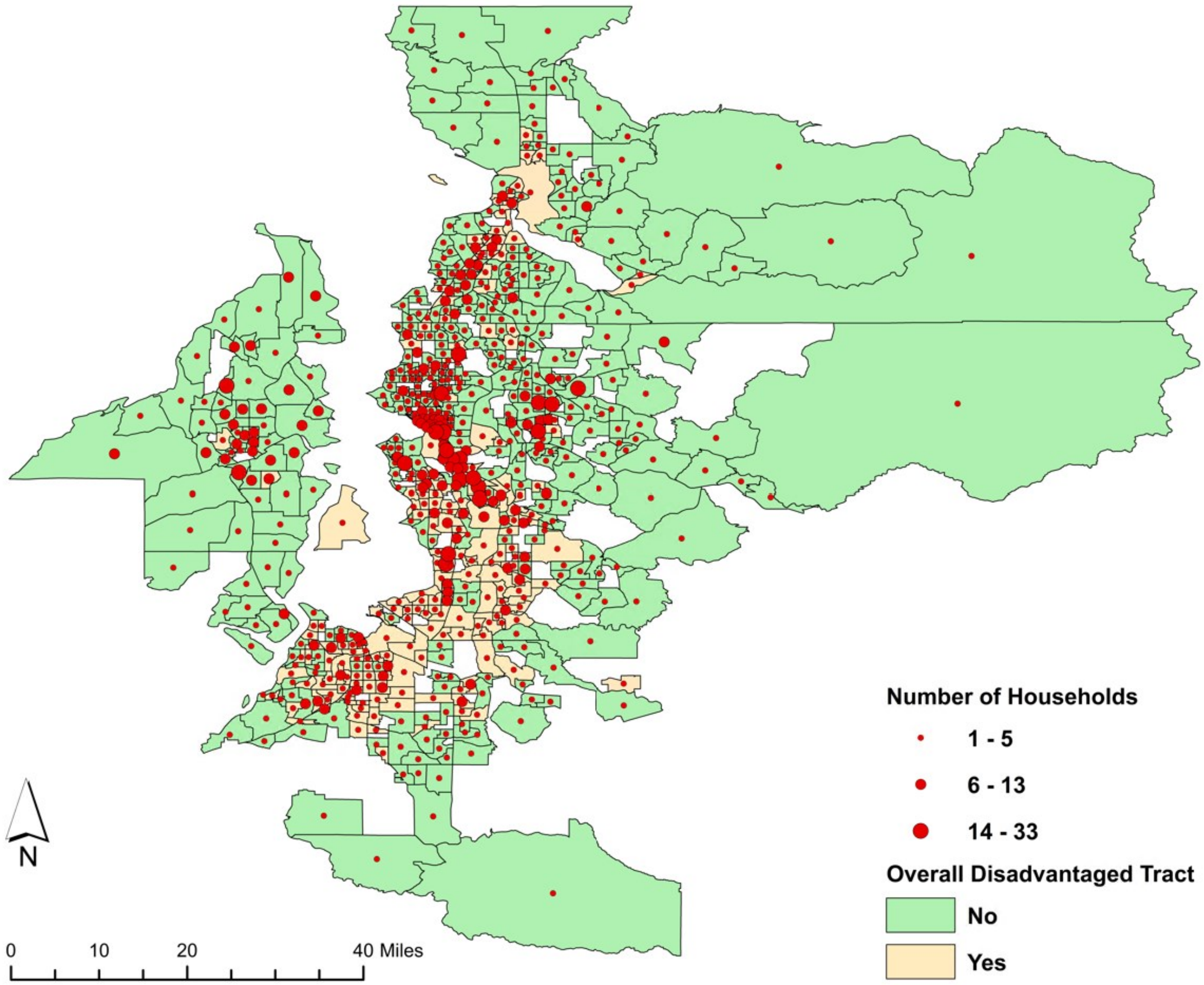


Figure 5-2: Households across the overall disadvantaged and non-disadvantaged communities of the Greater Seattle area in the sample (Generated by the Authors)

In the CMP format, the above-mentioned equations are restructured as follows:

$$y_1^* = \sigma_1 + \mu_1 \quad \text{(vii)}$$
$$y_2^* = \sigma_2 + \mu_2 \quad \text{(viii)}$$
$$y_3^* = \sigma_3 + \mu_3 \quad \text{(ix)}$$
$$y_4^* = \sigma_4 + \mu_4 \quad \text{(x)}$$
$$y_5^* = \sigma_5 + \mu_5 \quad \text{(xi)}$$
$$y_6^* = \sigma_6 + \mu_6 \quad \text{(xii)}$$

Where,

$$\sigma_1 = \alpha_1 X_i + \delta Y_{4i}, \sigma_2 = \beta_1 X_i + \delta' Y_{5i} + \delta'' Y_{6i}, \sigma_3 = \gamma_1 X_i, \sigma_4 = \tau_1 X_i + r Y_{3i}, \sigma_5 = \varphi_1 X_i + \omega Y_{1i}, \sigma_6 = \vartheta_1 X_i + \omega' Y_{1i}$$

$$y = g(y^*) = \left(O_J[c_{J-1} < y_1^*, y_2^*, y_3^*, y_4^*, y_5^* < c_J]\right)' \quad \text{(xiii)}$$

$$\mu = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6)' \sim N(0, \Sigma\ ) \text{ and } \Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} & \rho_{15} & \rho_{16} \\ \rho_{12} & 1 & \rho_{23} & \rho_{24} & \rho_{25} & \rho_{26} \\ \rho_{13} & \rho_{23} & 1 & \rho_{34} & \rho_{35} & \rho_{36} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 & \rho_{45} & \rho_{46} \\ \rho_{15} & \rho_{25} & \rho_{35} & \rho_{45} & 1 & \rho_{56} \\ \rho_{16} & \rho_{26} & \rho_{36} & \rho_{46} & \rho_{56} & 1 \end{bmatrix}$$

Here, $y_1^*, y_2^*, y_3^*, y_4^*, y_5^*$ and y_6^* are latent factors for online food, grocery, retail shopping, and in-person restaurant, grocery, and retail trips, respectively. Equation (xiii) represents a mapping or transformation from the set of latent variables y^* to a set of observed variables y. Where, assume y has *J* outcomes denoted as O_1 ,.., O_J and C_{J-1}, C_J are the cut points for defining the regions into which y* may fall. The terms $\rho_{12}, \rho_{13}, \rho_{23}, \rho_{14}, \rho_{24}, \rho_{34}, \rho_{15}, \rho_{25}, \rho_{45}, \rho_{16}, \rho_{26}, \rho_{36}, \rho_{46}$, and ρ_{56} represent the correlations between the error terms of the online shopping and in-person shopping variables. Suppose that $y_i = (0, y_{i2}, 0)'$ is observed for y_2 and subsequently for other dependent variables, where y_{i2} is the observed component of y_2, while the other two components (i.e., 0,0) of y_2 are unobserved. Now, a likelihood function can be expressed as follows in equation (xiv). The function would involve integrating the unobserved components of y_i. The integral over the latent variables has been approximated using a Monte Carlo integration.

$$L_i(\alpha_1, \beta_1, \gamma_1, \tau_1, \varphi_1, \vartheta_1, \delta, \delta', \delta'', r, \varphi, \omega, \vartheta, \omega', \Sigma;\ y_i|x_i) = \int_{-\infty}^{-\sigma_1}\int_{-\infty}^{-\sigma_2}\int_{-\infty}^{-\sigma_3}\int_{-\infty}^{-\sigma_4}\int_{-\infty}^{-\sigma_5}\int_{-\infty}^{-\sigma_6} \emptyset_j\,\{\mu_3, y_{i2}, y_{i1}, y_{i4}, y_{i5} - \sigma_{i2}, \sigma_{i1}, \sigma_{i4}, \sigma_{i5}\ \mu_6)'; \Sigma\} d\mu_1 d\mu_2 d\mu_3\ d\mu_4 d\mu_5 d\mu_6 \quad \text{(xiv)}$$

5.6 Results and Discussion

The analysis of the CMP model is carried out using "STATA" version 16 statistical software. After estimation, direct marginal effects are derived. The joint estimation results of the CMP model are displayed in Table 5-4 and Table 5-5. The coefficient estimation with significance level and marginal effects are reported. The estimations for the online delivery components, i.e., retail, grocery, and food, are shown in columns (i), (ii), and (iii), respectively. Columns (iv), (v), and (vi) report the estimates for the number of retail, grocery, and restaurant trips on travel day, respectively. The Wald X^2 test of the CMP model shows the model provides a better fit for the study data. This section will first discuss shopping behavior in disadvantaged communities while exploring the components of online shopping. Second, the discussion will be continued to understand the interactions of the online shopping components (i.e., endogenous variables). Finally, the results for the exogenous variables will be discussed.

5.6.1 Shopping Behavior in Disadvantaged Communities

Two DAC indicators, i.e., economic DACs and transportation DACs, are found to be statistically significant, and the associations vary across the endogenous variables. These results are reported in Table 5-4. The relationship between online retail shopping and the economic DAC is statistically significant and negative. The results suggest that households living in an economic DAC are less likely to order online for retail goods than those not living in such DACs. Specifically, compared to the non-DAC households, DAC households are 2% and 1% less likely to order retail goods "1 time" and "2 times", respectively. They also receive fewer grocery deliveries compared to non-DACs. Households in non-DACs are generally expected to order more than DACs because of their increased access to technologies, electronic devices, and emerging online delivery services. In contrast, people in DACs have limited access to technologies, infrastructure, and investments. Moreover, these DACs are affected by high poverty, inequality, and unemployment; hence, the people from these communities are less aligned to order online (*195*). Research shows individuals living in DACs may experience difficulties with online shopping due to a lower rate of internet access (e.g., digital divide) (*196*). Besides, due to lower incomes, they may struggle to afford the expenses associated with online shopping, such as computer or mobile devices and shipping fees (*43; 197; 198*). Furthermore, individuals unfamiliar with the internet or have limited experience with online shopping may struggle with the process and find it confusing or overwhelming, which can discourage them from engaging in online shopping (*199*). In some DACs, individuals may live in areas where delivery services are either unreliable or unavailable, which can decrease the appeal of online shopping (*200*). Additionally, a lack of trust in online shopping is found to demotivate people to order online (*201*). There may be individuals in DACs who have these concerns about online shopping. As a result, they may opt to shop at physical stores where they can physically inspect and try out products before making a purchase, as suggested by Zhu & Wang (*16*) and Patwary & Khattak (*43*).

Similarly, the results suggest that the probability of in-person retail shopping trips decreases in transportation DACs. However, it is not statistically significant. The association between restaurant trips and transportation DAC indicator is statistically significant. The probability of making more restaurant trips on travel day decreases if a household lives in a transportation DAC. Several reasons may contribute to these findings. Households in disadvantaged areas have a lower average income (Figure 5-3), making it challenging to pay for restaurant meals (*36*). Besides, higher food insecurity makes it difficult for residents to pay for meals at restaurants regularly (*202*). The availability of restaurants is usually limited in DACs, making it less convenient for residents to dine out (*202*). Average trip length and duration are higher in a transportation DAC than in a non-DAC (*195*). Therefore, people in DACs tend to make fewer number of trips.

Equity conditions in terms of household income and race may contribute to these findings. Figure 5-3 suggests that the percentage of lower-income households is higher than that of higher-income households in a disadvantaged community. Also, it shows that more white people live in

DACs than the non-white population. The correlates of household income in Table 5-4 suggest that lower-income households are less inclined to receive food, grocery, and retail deliveries (details on household income correlates are discussed in later sections). The impact of COVID-19 on low-wage industries may have hindered the ability of people from DACs to participate in online ordering due to widespread job losses, leading to a reduction in disposable income for a significant number of individuals from these communities. This may indicate the presence of a digital divide. Since physical stores are becoming less popular, lower-income people may find it challenging to access goods and services. Therefore, updating existing policies for lower-income people (e.g., food stamps and paratransit) is important in the wake of a digital divide and digital poverty.

5.6.2 Key Relationships among the Endogenous Variables

The results on the key relationships between the endogenous variables in Table 5-5 suggest that the effect of in-person shopping trips on online retail shopping is negative and statistically significant. Marginal effects indicate that compared to the base of retail shopping trips (i.e., no delivery), a "1-time" retail shopping trip on a travel day is associated with being 3% less likely to receive “1 time” and 1% less likely to receive retail deliveries for higher categories. Consequently, the effects of other higher categories of retail shopping trips on grocery shopping confirm similar assertions, and the effect increases (e.g., 16% less likely for “3+ times”). This is consistent with Patwary and Khattak (*43*), who also found the relationship between online shopping and physical shopping trips is substitutive. Making more physical shopping trips reduces the frequency of online retail shopping. It is probably because people travel to physical stores to compare or experience the actual goods they browsed online. Also, in-person shopping offers the chance for social interaction and the instant gratification of being able to take the purchased items home right away. Also, in-person shopping trips provide greater flexibility in choosing departure time (*203*).

The effect of online retail shopping on grocery shopping trips is positive and statistically significant. The findings show that households having "3+ times" online retail shopping is 25%, 12%, and 11% less inclined to make grocery shopping trips for "1 time", "2-times" and "3+ times" categories, respectively, than the households that do not order online retail goods. It may be because more people are opting for online grocery purchases instead of physical store visits. This shift in consumer behavior is likely due to a combination of convenience, a wider selection of products, and the ability to compare prices easily (*204; 205*). A rapid growth in online grocery shopping has been observed during COVID-19. With this new trend, physical grocery stores may experience a drop-in customer foot traffic and fewer trips to the store.

Moreover, restaurant trips (i.e., dine-in) and grocery shopping are negatively associated and statistically significant. For example, compared to the base category (i.e., no restaurant trips), households that make "3+ times" restaurant trips for dine-in are 7%, 2%, and 2% less likely to order groceries online for "1 time", "2 times" and "3+ times" grocery shopping categories, respectively. Previous studies have observed similar trends (*36; 206*). This is probably because dining out can be an alternative to cooking meals at home, and purchasing groceries out can be an alternative to cooking meals at home and purchasing groceries, leading to a reduction in online grocery shopping (*206*). Individuals who frequently dine at restaurants may feel less of a need to shop for groceries and prepare food at home, which can result in fewer online grocery purchases (*36*). Moreover, dining out offers a change of pace and a social atmosphere that some people may find attractive, further contributing to a preference for restaurant trips over online grocery

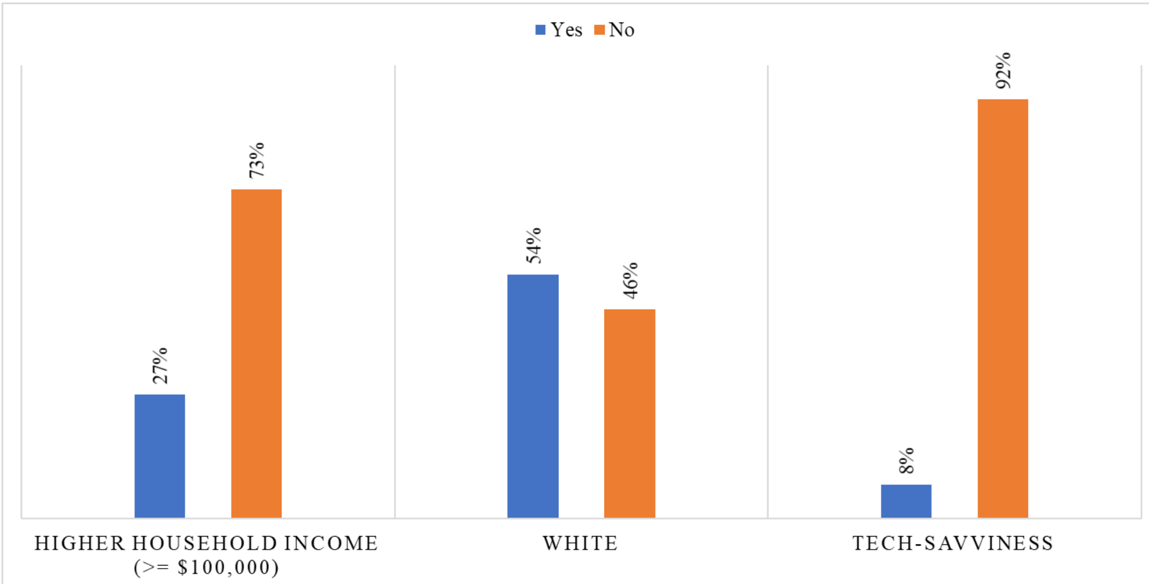


Figure 5-3: The distribution of key exogenous factors in economic DAC

Table 5-4: CMP Joint Estimation Results of the Exogenous Variables (N= 2,051)

Variables	(i) Package Delivery Freq.					(ii) Grocery Delivery Freq.					(iii) Food Delivery Freq.				
	Coef.	Marginal Effect				Coef.	Marginal Effect				Coef.	Marginal Effect			
		No	1 time	2 times	3+ times		No	1 time	2 times	3+ times		No	1 time	2 times	3+ times
Household income (<$25,000)															
$25,000-$49,999															
$50,000-$74,999	0.27	-0.09	0.06	0.03	0.01										
$75,000-$99,999	0.37	-0.13	0.07	0.04	0.01						0.31	-0.05	0.03	0.01	0.001
>=$100,000	0.44	-0.16	0.09	0.05	0.02										
Household Employment (No workers)															
Single worker											0.32	-0.03	0.02	0.01	0.00
Multiple workers						0.27	-0.04	0.02	0.01	0.01	0.42	-0.04	0.03	0.01	0.00
Household vehicle count (no vehicle)															
single vehicle											-0.26	0.03	-0.02	-0.01	0.00
multiple vehicles	0.21	-0.07	0.04	0.02	0.01						-0.37	0.05	-0.03	-0.01	-0.01
Household Str. (Single adult)															
Two or more adults without children	0.16	-0.06	0.03	0.02	0.01	0.34	-0.04	0.03	0.01	0.01	-0.25	0.02	-0.01	-0.01	0.00
Two or more adults with children	0.35	-0.13	0.07	0.04	0.02	0.54	-0.07	0.04	0.01	0.01	0.33	-0.04	0.03	0.01	0.00
Household Race (Others), White											-0.24	0.03	-0.02	-0.01	0.00
Seattle Home (No), Yes						0.34	-0.05	0.03	0.01	0.01					
Tech savviness (No), Yes	0.50	-0.19	0.08	0.07	0.04	0.47	-0.08	0.05	0.02	0.02	0.65	-0.10	0.06	0.03	0.01
Economy DAC (No), Yes	-0.08	0.03	-0.02	-0.01	0.00	-0.04	0.02	-0.01	-0.01	0.00					
Transport DAC (Base: No), Yes															

Table 5-4 Continued

Variables	(iv) Retail Shopping Trips					(v) Grocery Shopping Trips					(vi) Restaurant Trips				
	Coef.	Marginal Effect				Coef.	Marginal Effect				Coef.	Marginal Effect			
		No	1 time	2 times	3+ times		No	1 time	2 times	3+ times		No	1 time	2 times	3+ times
Household income (<$25,000)															
$25,000-$49,999															
$50,000-$74,999															
$75,000-$99,999															
>=$100,000															
Household Employment (No workers)															
Single worker	-0.36	0.11	-0.05	-0.03	-0.03	-0.33	0.11	-0.04	-0.04	-0.03					
Multiple workers	-0.59	0.17	-0.07	-0.05	-0.05	-0.36	0.13	-0.05	-0.04	-0.04					
Household vehicle count (no vehicle)															
single vehicle											0.22	-0.06	0.03	0.02	0.01
multiple vehicles	0.30	-0.08	0.04	0.02	0.02						0.26	-0.07	0.04	0.02	0.01
Household Structure (Single adult)															
Two or more adults without children	0.28	-0.07	0.03	0.02	0.02	0.31	-0.11	0.04	0.03	0.04					
Two or more adults with children	0.23	-0.06	0.03	0.02	0.01	0.45	-0.15	0.05	0.04	0.06					
Household Race (Others), White											0.21	-0.06	0.03	0.02	0.01
Seattle Home (No), Yes											0.12	-0.04	0.02	0.01	0.01
Tech-savviness (No), Yes															
Economy DAC (No), Yes															
Transport DAC (No), Yes	-0.03	0.007	-0.003	-0.002	-0.002						-0.16	0.041	-0.02	-0.01	-0.01
Model Fit Statistics															
Number of observations			2,594												

Table 5-4 continued

Wald chi2 (39)	327.4
Model Significance Test	0
Log Likelihood	-7274
AIC	14666
BIC	15011

**** p<.01, ** p<.05, * p<.1*

Table 5-5: CMP Joint Estimation Results of the Endogenous Relationships (N= 2,051)

Variables	(i) Package Delivery Freq.					(ii) Grocery Delivery Freq.					(iii) Food Delivery Freq.				
	Coef.	Marginal Effect				Coef.	Marginal Effect				Coef.	Marginal Effect			
		No	1 time	2 times	3+ times		No	1 time	2 times	3+ times		No	1 time	2 times	3+ times
Package Delivery Freq (No Delivery)															
1 time															
2 times															
3+ times															
Grocery Delivery Freq (No Delivery)															
1 time															
2 times															
3+ times															
Food Delivery Freq (No Delivery)															
1 time															
2 times															
3+ times															
Retail Shopping Trips (No Trips)															
1 time	-0.15	0.05	-0.03	-0.02	-0.01										
2 times	-0.23	0.08	-0.05	-0.03	-0.01										

Table 5-5 continued

3+ times	-0.84	0.25	-0.16	-0.06	-0.02						
Grocery Shopping Trips (No Trips)											
1 time						0.01	-0.001	0.001	0.001	0.001	
2 times						-0.03	0.003	-0.00	-0.001	-0.001	
3+ times						0.05	-0.01	0.004	0.001	0.001	
Restaurant Trips (No Trips)											
1 time						-0.88	0.08	-0.05	-0.013	-0.015	
2 times						-1.32	0.09	-0.06	-0.014	-0.016	
3+ times						-1.67	0.1	-0.07	-0.015	-0.016	
Correlation Terms											
Online Grocery Shopping	0.31										
Food Delivery	0.36					0.58					
Retail Trips	0.36					0.13					0.32
Grocery	0.81					0.18					0.15
Restaurant Trips	-0.062					0.63					-0.19

Table 5-5 continued

Variables	(iv) Retail Shopping Trips					(v) Grocery Shopping Trips					(vi) Restaurant Trips				
	Coef.	Marginal Effect				Coef.	Marginal Effect				Coef.	Marginal Effect			
		No	1 time	2 times	3+ times		No	1 time	2 times	3+ times		No	1 time	2 times	3+ times
Package Delivery Freq (No Delivery)															
1 time						-0.81	0.29	-0.12	-0.08	-0.09		0.18	-0.05	0.03	0.01
2 times						-1.50	0.43	-0.21	-0.11	-0.11		0.26	-0.08	0.04	0.03
3+ times						-2.07	0.48	-0.25	-0.12	-0.11		0.49	-0.15	0.07	0.05
Grocery Delivery Freq (Base: No Delivery)															
1 time															
2 times															
3+ times															
Food Delivery Freq (Base: No Delivery)															
1 time	-0.30	0.07	-0.04	-0.02	-0.02	-0.30									
2 times	-0.87	0.16	-0.09	-0.04	-0.03	-0.87									
3+ times	-1.33	0.19	-0.11	-0.05	-0.03	-1.33									
Retail Shopping Trips (Base: No Trips)															
1 time															
2 times															
3+ times															
Grocery Shopping Trips (Base: No Trips)															
1 time															
2 times															
3+ times															

Table 5-5 continued

Restaurant Trips (Base: No Trips)		
1 time		
2 times		
3+ times		
Correlation Terms		
Online Grocery Shopping		
Food Delivery		
Retail Trips		
Grocery	0.14	
Restaurant Trips	0.22	-0.07

shopping. However, food delivery and retail shopping trips are inversely related. For example, compared to the households that do not buy food online, households that buy food online "3+ times" are 11%, 5%, and 3% less interested in making "1 time", "2 times", and "3+ times", respectively. It may be the households that are comfortable ordering food online may want to go outside at least fewer times, making fewer in-person trips.

5.6.3 Correlations of Exogenous Variables

The discussion is shifted to the exogenous factors that influence households to shop in-person and online for groceries, food, and retail goods on travel day (Table 5-4). It has been observed that *household income* is a significant exogenous factor and is positively correlated with online and in-person retail shopping. Compared to the lower income households (<$25,000), the higher income households (>=$100,000) are 9%, 5%, and 2% more likely to buy retail products online for "1 time", "2 time", and "3+ times" delivery frequencies, respectively, on travel day. Besides, higher household income is associated with more prepared meal deliveries on the travel day. These results are as we anticipated earlier and are consistent with the previous studies (*10; 34; 36; 182*). For example, Dias et al. (*36*) found a positive correlation between a higher-income household and online shopping frequency, including retail and food deliveries. Farag et al. (*34*) and Cao et al. (*10*) also found similar results while investigating the effects of income on online shopping frequencies. Research during COVID-19 also showed a positive association between household income and online shopping (*32; 43*). Higher-income households have the income or flexibility to shop online more often. Lower-income people may not be able to transition to the virtual means of purchasing compared to higher-income households. The findings from *household employment* also support this. The results suggest that multiple-worker households order more groceries and prepared meals than single-worker households. In contrast, they are less inclined to buy retail and groceries from physical stores.

The *household structure* is found to be a significant factor. Compared to single adult households, the probability of buying retail products online for households with two or more adults with children increases by 7%, 4%, and 2% for the "1 time", "2 time", and "3+ times" delivery frequencies. Consequently, an increase in household size is associated with more grocery shopping and retail shopping. It is consistent with our earlier assumptions. Larger households are expected to consume more across all components of online shopping; therefore, they buy online more than smaller households. This is in line with the prior studies (*36; 207*). Moreover, the *household race* is statistically significant. White households are less likely to opt for online ordering of prepared meals compared to non-white households and are more inclined towards traveling for in-person dining experiences. This is similar to Kim and Wang (*178*), who suggested white people receive more durable goods and grocery deliveries.

Household vehicle count is also found to be an important exogenous factor that can affect the components of online shopping. Multiple-vehicle households are less inclined to order prepared meals online. The more vehicles owned by a household, the less likely they are to order prepared meals and groceries. Having more cars in a household generally led to a reduced number of online deliveries and more in-person trips. More cars make households flexible enough to travel to stores when necessary. Besides, when a household lives in an urban area like *Seattle city*, they are 3% more likely to receive grocery shopping (i.e., "1-time" category), and at the same time, 3% more likely to make restaurant trips than the ones who live outside the Seattle city. Those who live in an urban city have more accessibility to physical stores and more online delivery coverage.

Therefore, they have an increased propensity to receive online deliveries. Faster internet and early access to newer technologies in urban areas may further encourage buying online more than in rural areas (*34*).

Moreover, the estimates of *tech-savviness* support these findings. Tech-savvy households are found to be associated with buying more online compared to non-tech-savvy households. For example, tech enthusiast households are 8%, 5%, and 6% more likely to order online for retail, groceries, and prepared meals, respectively. This is similar to the findings of Patwary and Khattak (*43*), who found that individuals with daily internet use are highly associated with online shopping. Menon (2018) also found that customers' tech-savviness positively influences online shopping behavior. Usually, tech-savvy people have a seamless and efficient experience while shopping online due to their familiarity with the technology used and ability to navigate websites and apps easily (*208*). Their knowledge of online security measures and the ability to recognize potential scams also make them feel more secure in making online purchases (*208*). Additionally, these tech-savvy individuals often have higher expectations for technology in online shopping, such as personalization and advanced search options, and are more likely to choose retailers that provide these features.

5.7 Limitations

Some of the important limitations of this research should be recognized. Although the sample size is larger than the prior similar studies, this is still a small data sample from a regional household travel survey. Tract-level estimations covering different states or the whole US can provide different estimations, as travel behavior could vary across regions. Furthermore, the 2021 Puget Sound Household Travel survey is based on the one-day delivery and travel diary. However, covering longer periods, e.g., one-week delivery and travel diary, would provide more reliable estimates. Other limitations of survey research, such as errors in the coding of the survey, the potential for non-response bias, and non-coverage of certain populations and areas, are recognized while noting that the survey was conducted professionally.

5.8 Conclusions

This study found online shopping is different in disadvantaged communities. The COVID-19 pandemic, with the ever-changing ICT landscape, has brought significant changes in individual travel and purchase behavior. Online delivery of goods and services surged, and new forms of online shopping, i.e., retail, grocery, and food, have emerged. However, little research has been conducted on the interaction of these components in DACs that face issues of marginalization and limited access to digital resources. This study analyzes the interaction among the three components of online shopping and in-person travel for these activities and explores the correlates of socio-demographics, spatial, and technological attributes on online shopping and in-person travel. The study uses a unique dataset that combines the 2021 Puget Sound Household Travel Survey data and US DOT's Justice40 database and employs a conditional mixed process model that accounts for related unobserved endogeneity.

The overall results show that in-person retail shopping is generally correlated with a reduction in online retail shopping frequency, while an increase in restaurant visits is similarly linked to a decrease in the frequency of online grocery shopping. Households living in DACs are less likely to buy retail goods and groceries online compared to those in non-DACs. Furthermore, the chance of going to restaurants more often is lower for households located in DACs. The findings also highlight the disparities between DACs and non-DACs in regard to their in-person and online shopping habits, influenced by technology access and income inequality. Importantly, the findings suggest that lower-income households in DACs are less inclined to order online services than higher-income households. Also, tech-savvy households are found to be associated with purchasing more online compared to non-tech-savvy households. These results may reflect the wake of a digital divide and digital poverty originating from income, racial disparity, and lack of required infrastructure and policies.

The findings of this study are based on the greater Seattle area of Washington state. Therefore, these findings should not be generalized to the context of other US states. Nonetheless, the results are informative and have important policy implications. Policies that make regular activities at hand (i.e., digital) and connect households of all socioeconomic strata should be undertaken so that access to all goods and services for lower-income people would be easier and help improve the quality of life. In addition, planners and engineers must account for the shopping behavior across DACs in the travel demand model. Currently, they do not fully account for online shopping, let alone its different components. Since DACs are overburdened by congestion and pollution, an updated four-step model with the inclusion of the newer components of online shopping may provide benefits in terms of future trip prediction and associated policies. They can account for trip frequencies by purpose in DACs. For instance, if lower online shopping in DACs increases the need for trips to physical stores, it may change the distribution of trip purposes in travel demand models. Additionally, planners also need to adjust assumptions about trip length and time of day in travel demand models. If fewer people in DACs shop online, it may increase the need for longer trips or trips during peak hours, which could affect the travel demand estimates in these models. Since decisions on online and in-person activities are taken as a cohesive lifestyle package, they should be reflected in transport and related policies.

This work reflects current societal priorities, i.e., a focus on equity and disadvantaged communities. Several extensions for future work can be anticipated. First, efforts should be made to collect data on home deliveries in a large-scale survey covering all the states or at least several regions to better understand the interactions between retail, food, and grocery deliveries. Second, since e-commerce is growing rapidly, it is important to regularly conduct interactions between online and in-person activity with the availability of newer data. Third, future research may also improve the modeling results by collecting one-week delivery and travel diary data rather than depending on the one-day diary data. Finally, efforts need to be made to analyze shopping behavior at the census tract level across all the DACs in the US.

5.9 Acknowledgments

This project was partially supported by the Educational Component of the Collaborative Sciences Center for Road Safety (CSCRS), a US Department of Transportation National University Transportation Center.

CHAPTER 6 EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR DECARBONIZATION: ALTERNATIVE FUEL VEHICLE ADOPTION IN DISADVANTAGED COMMUNITIES

A version of this chapter authored by A. Latif Patwary and Asad J. Khattak is currently in the final review stages for potential acceptance in the International Journal of Sustainable Transportation:

Patwary, A. L., & Khattak, A. J. (2023). *Explainable Artificial Intelligence for Decarbonization: Alternative Fuel Vehicle Adoption in Disadvantaged Communities.*

6.1 Abstract

This paper explores the adoption of alternative fuel vehicles (AFVs), leading to decarbonization, in disadvantaged communities (DACs) by applying statistical and explainable artificial intelligence (XAI) techniques to understand the factors associated with AFV adoption in these communities. The study harnesses a unique and comprehensive database of surveys and public databases for the Puget Sound region in the US. The XAI techniques, specifically the Extreme Gradient Boosting (XGBoost) algorithm with Shapely Additive Explanations (SHAP), provide interpretable and understandable explanations of factors associated with AFV adoption in DACs. The study findings provide an understanding of the social and economic factors and challenges of DACs. The results suggest several key factors, especially a lack of access to charging infrastructure, consumer attitudes, and income, play a substantial role in adopting AFVs. As expected, AFV adoption in DACs (12.96%) is lower than non-DACs (15.30%). More public charging stations strongly correlate with AFV adoption in DACs. Tech-oriented households in DACs are more likely to adopt AFVs compared with non-DACs. The findings also point to the significant effects of home charging facilities while adopting AFVs in DACs. The XAI results emphasize the importance of social and economic factors in AFV adoption programs and provide insights into decision-making in DACs. This research contributes to the literature on AFV adoption and suggests opportunities for improvements in DACs transitioning to AFVs. The study findings can be used to assess the planning-level impacts of refueling or charging infrastructure in DACs while enabling DACs to benefit from infrastructure investments.

6.2 Introduction

The US transportation sector is one of the major energy-consumptive sectors and is also responsible for substantial environmental emissions. The US Energy Information Administration (EIA) states that the transportation sector consumes about 37% of all energy consumption (*209*). The energy sources vary widely, including 90% of the transportation sector's energy consumption being provided by petroleum, 4% by natural gas, 5% from renewable energy, and less than 1% from electric power (*209*). Since the majority comes from fossil fuels, the transport sector produces the highest greenhouse gas (GHG) emissions, which is about 27%, based on the US Environmental Protection Agency (EPA)'s 2020 estimates (*210*). To mitigate risks to the environment and public health, decarbonization is necessary. Therefore, there is an increasing focus on developing and promoting alternative technologies that rely less on fossil fuels and help decarbonize by reducing GHG emissions. Alternative fuel vehicles (AFVs) are vehicles that use full or partial alternatives to fossil fuels. AFVs are generally fueled by hydrogen, propane, biofuel, flex fuel, battery electric, and natural gas. With technological advances, hybrid electric vehicles (HEVs), plug-in hybrid electric vehicles (PHEVs), and battery electric vehicles (BEVs) are currently being developed and diffused to the automotive market to achieve better fuel economy and reduced emissions.

Therefore, AFVs are generally considered to be a solution to reducing the GHG effect. The total environment of the car fleet will be changed with the replacement of fossil fuel vehicles by AFVs in the future.

The rising investments of auto manufacturers and government-sanctioned subsidies are boosting AFV sales. Thus, the AFV market is predicted to be about $1.3 trillion by 2030 (*211*). However, considerable effort is still needed to promote AFV adoption across the communities. Disadvantaged communities are not highly representative of AFV sales and adoption. For the first time, the federal government has declared that 40% of the overall benefits of certain federal investments will be for the disadvantaged communities (DACs) that are marginalized, underserved, and polluted (*103*). The major notable investments include clean energy, clean transit, climate change, sustainable housing, legacy pollution, and infrastructure. This initiative is known as Justice40 or Environmental Justice (EJ). It covers several programs, including charging and fueling infrastructure grants, a carbon reduction program, a low or no emissions vehicle program, a reconnecting communities pilot program, and a thriving communities program, among others. To benefit from these programs, cities in the US will need to undertake a qualitative and quantitative assessment to understand the community demographics, challenges, and needs. Therefore, knowing how AFVs can be diffused to DACs is crucial considering the EJ investments.

Understanding the consumers' adoption of AFV across disadvantaged and non-disadvantaged communities is complex and pivotal for the long-term success of a sustainable future transportation system. Identifying the underlying factors and barriers to AFV adoption in DACs is important for effectively formulating policies. Existing studies have explored various correlates of AFV adoption, including infrastructure barriers, technology proficiency, home ownership, and household income (*212-214*). The most important distinguishing factor separating DACs and non-DACs in terms of AFV adoption can be household income since research shows an incommensurately lower number of AFV sales in low-income communities (*215; 216*). Besides, insufficient infrastructure can be another major operational barrier that hinders the adoption of AFVs (*217*). However, to the best of our knowledge, the inclusion of elements representing disadvantaged communities in terms of equity, environment, economy, health, resilience, and transportation access does not exist. In this regard, this study mainly investigates 1) are there differences in AFV adoption between DACs and non-DACs? 2) If so, what are the key factors/reasons for differences in adoption?

6.3 Literature Review

Rogers' Diffusion of Innovations theory describes innovation as an idea or practice recognized as new by an entity or another component of adoption (*218*). AFVs can be such innovations perceived by traditional fleet operators and diffused to market for consumers' adoption where conventional fossil fuel vehicles dominate. Numerous studies investigate the correlates of adopting AFVs and related components. Researchers generally have considered AFV adoption because it involves the purchase and uses through behavioral responses. This section presents a synthesis of previous studies on AFV adoption, focusing on the influencing factors.

Scholars studied consumers' AFV adoption behavior from individual, household, and commercial fleet perspectives. The overall factors that affect the adoption of AFV can be categorized as socio-demographics, locational, situational, and travel-specific characteristics. Household income is found to be one of the major influencing factors for AFV adoption and sales.

Higher-income household positively affects AFV sales (*219; 220*). However, there are studies that found different or ambiguous relations between AFV adoption and household income (*221-223*). For example, Sierzchula et al. (*224*) argued that household income is ineffective while analyzing electric vehicle (EV) sales. The effect of income may be weak due to the price differences between AFVs and conventional vehicles (CVs) since the target group for AFVs is much wider than conventional fuel vehicles (*225*). Household size is also found relevant in analyzing AFV adoption and sales. Jia et al. (*214*) predicted the increase of BEV and PHEV counts by 197% and 149%, respectively, with an increase of one standard deviation in household size. A household's number of family members will likely affect vehicle type because of seating requirements and vehicular space usage (*212*). Moreover, the number of cars a household owns is also found to be significant in previous studies. Early adopters of EVs in Norway showed that consumers usually buy EVs as an addition to their multi-car households (*226*). Less driving mileage or limited charging stations may make them alternative household vehicles. In contrast, Hidrue et al. (*227*) found no substantial relationship between household vehicle size and EV preferences. In addition, consumers' participation in a car-sharing program is analyzed for AFV adoption (*212; 228*). Car sharing status indicates the regular participation of a household in a car share program, which offers an alternative to traditional car ownership, allowing people to reserve and use vehicles as needed, often through a smartphone app or online platform. This can influence an individual's travel behavior and attitude toward new technologies. Notably, previous studies have found a positive association between participating in a car-sharing program and the likelihood of AFV adoption (*212*).

Insufficient infrastructure is a major operational barrier that hinders the adoption of AFVs (*217; 229; 230*). Studies found positive associations between the number of charging stations and AFV adoption (*231*). However, the current refueling/charging infrastructure is inadequate to encourage the diffusion of AFVs. The refueling/charging structure is unevenly distributed in all areas. For example, refueling stations are common in large organizations' workplaces, and those organizations' employees also regard home charging facilities as highly important. However, these are not necessarily true for less developed locations, e.g., suburban, rural, and disadvantaged areas.

The utility of AFVs typically increases since it encourages more people in the neighborhood to purchase and use them (Mau et al., 2008). In contrast, household location in an urban or city area usually represents the impact of households' surroundings on AFV adoption and sales, known as a neighbor effect. The neighbor effect has been analyzed in numerous studies (*232; 233*). Axsen and Kurani (*234*) revealed that most households rank social interaction as the highest priority in an AFV assessment project. Also, Zhu and Liu (*222*) found the evidence of the "neighbor effect" in Florida to be higher in surrounding urban and rural areas while analyzing the adoption of hybrid electric vehicles. Research on lower-income and disadvantaged regions for AFV adoption has described consumer anxieties, disproportionate distribution of infrastructure, land use mixing, unbalanced land use, and environmental concerns (*216; 235; 236*). However, research on DACs regarding investments in clean energy, infrastructure, climate change, and sustainable transportation is still limited. Also, there is a need for a more comprehensive understanding of the barriers that are specific to DACs in terms of US Department of Transportation's (USDOT) disadvantaged indicators, i.e., transportation access, equity, environment, health, resilience, and economy. This study attempts to fill these gaps. Investigation of these aspects using explainable artificial intelligence (XAI) techniques provides new insights into AFV adoption in underserved areas and develop various policy incentives.

6.4 Conceptual Framework

As explored in the literature review section, we expect that high household income to positively correlate with AFV adoption (*219; 220*). Higher income enables households to purchase a vehicle. The currently high initial cost of AFV makes it difficult for lower income households to purchase such vehicles. Therefore, higher income has a positive association with AFV purchase. We also anticipate that household size and the number of household vehicles should weigh into the relationship. A household's number of family members can affect vehicle type because of greater seating requirements and vehicular space usage (*212*). A multi-car family is also expected to have a higher probability of adopting AFV than a single-car household. Research shows consumers generally buy EVs as an addition to their multi-car households (*226*). Moreover, living in a city or urban location should provide more incentives (e.g., a higher concentration of charging stations and neighbor effect) to adopt AFV than households in rural areas (*222; 234*). Involvement of a household in a car-sharing program exposes them to newer technologies; hence, they are more interested in adopting AFVs (*212; 228*). More charging/refueling stations should encourage more people to adopt cleaner technologies. Disadvantaged communities should be a barrier to expanding AFV adoption and market diffusion because of consumer anxieties, uneven infrastructure distribution, mixed land use, unbalanced development, and environmental concerns (*216; 235; 236*). Different types of disadvantaged communities could have distinct effects on AFV adoption. Equity-disadvantaged communities should see an increase in AFV adoption since high education and high-income disparities are critical in these areas. The overall study framework is illustrated in Figure 6-1.

6.5 Methodology

6.5.1 Data

The data for this study are collected from three different sources, including the 2021 Puget Sound Household Travel Survey (PHTS) (*194*), the US Department of Energy's alternative fuels data center (AFDC) (*237*), and US Department of Transportation's (USDOT) Environmental Justice databases (*160*). A unique and comprehensive dataset of households at the tract level is created after linking these databases. The dataset contains 2,202 households of the Puget Sound region (i.e., Greater Seattle: King, Kitsap, Pierce, and Snohomish counties) during April-June of 2021. PHTS provides socio-demographics, travel, and vehicle information at the household level. It is a standard web-based online survey designed to collect complete household travel diary information from invited participants. On the other hand, in the USDOT's EJ database, disadvantaged communities (DACs) are designated based on 22 indicators. These indicators are grouped into six major categories of disadvantages:

- Transportation Access Disadvantaged: This category identifies communities or places that are transport disadvantaged in terms of access. Residents in these tracts spend more time and cover long distances to travel where they need to go.
- Health Disadvantaged: It identifies tracts that are vulnerable to harmful health effects, disability, and environmental exposures.
- Environmental Disadvantaged: It recognizes tracts that have extremely high levels of certain air pollutants and lead-based pollution.

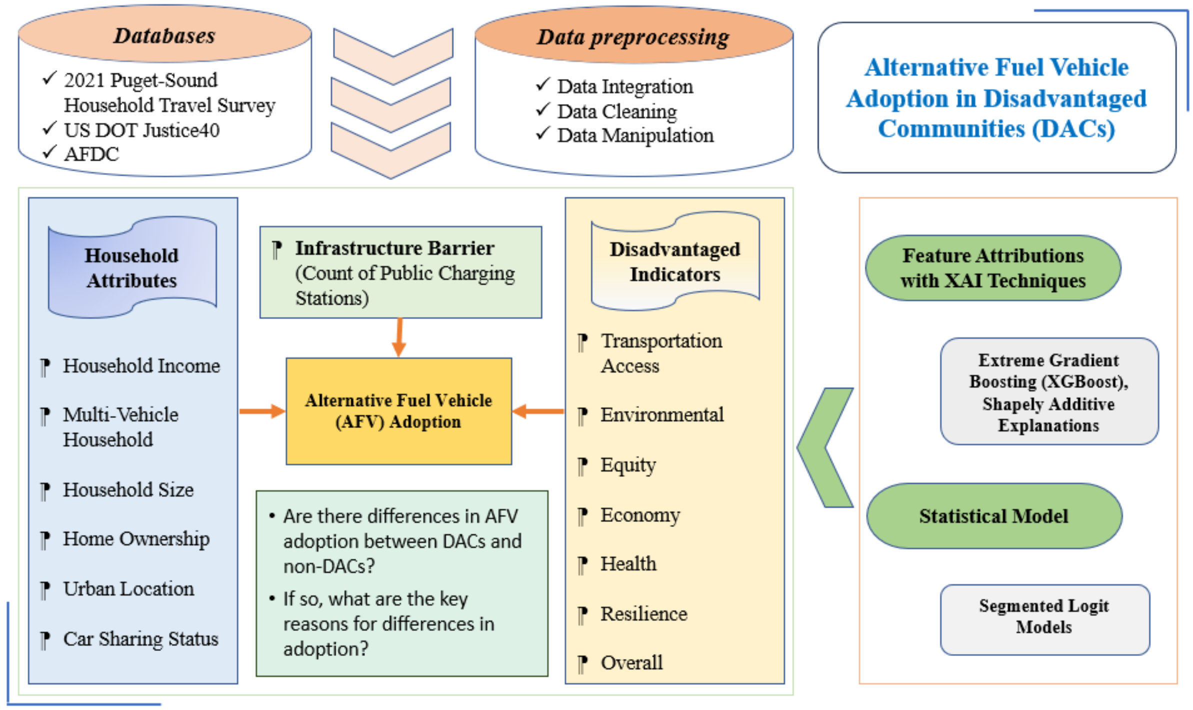


Figure 6-1: Study Framework

- Economic Disadvantaged: It characterizes communities with high poverty, low wealth, high unemployment, high illiteracy rate, low homeownership, and higher inequality.
- Resilience Disadvantaged: It identifies areas more prone to natural hazards triggered by climate change.
- Equity Disadvantaged: It identifies tracts with a high percentage of the population, age 5+, who speak English "less than well." Government agencies use this data to report racial equity and social and economic status (*238*).

Table 6-1 shows disadvantaged indicators for transportation and environmental major categories. This provides a glimpse of how the major categories are calculated. The percentile value of each indicator is calculated for the census tract. The average percentile for each tract is estimated within each major category. A tract in each category (except resilience) is assigned 1 (i.e., disadvantaged) if it is in the 50^{th} percentile (i.e., percentile ranking average values >= 50%) and 0 otherwise. However, the 75^{th} percentile value (i.e., >= 75%) is used for designating "resilience disadvantaged." This category focuses on identifying communities that are particularly vulnerable to hazards caused by climate change, a priority of the US administration. The USDOT used a higher threshold of 75% to pinpoint communities with relatively high vulnerability. A census tract is considered resilience disadvantaged if it ranks in the top 75% of the average scores in this specific category. For other categories, a threshold of 50% is applied, which is more aligned with identifying communities that exhibit above-median levels of disadvantage within their respective categories. The scores of all six categories (1 or 0) for each tract are then summed up, which range from zero (0) to six (6). A census tract is considered overall disadvantaged when it has a score of 4 or higher. It can be interpreted as ranking in the top 50% of the averages in each category (75% for the resilience category), where higher scores indicate a higher level of disadvantage. The disadvantaged indicator-based dataset has information for 72,843 census tracts of the US. In the Seattle, Washington study area, 648 tracts' DAC information is linked at the household level.

The response variable in this study is AFV, which is a binary variable generated by the authors from the fuel type variable in the vehicle data file of PHTS. Hybrid-electric, electric, flex-fuel, biofuel, natural gas, and hydrogen fuel are considered AFVs. The households who own AFV are assigned 1; otherwise, zero. The descriptive statistics in Table 6-2 (green colored) show that 14% of the households own AFV, which is slightly above the national market share of AFV sales (i.e., 12%) (*211*). HEV, PHEVs, and BEVs constitute the most AFVs in the study sample. Specifically, 65% of all household vehicles are HEVs and PHEVs, 23% are BEVs, and 12% belong to other AFVs, including hydrogen fuel, biofuel, flex fuel, natural gas, etc. All these types of AFVs are considered together in this study since they all provide cleaner technologies over conventional fossil fuel vehicles, which help lessen the growing transportation emissions.

Table 6-2 (orange colored) represents a two-way frequency distribution of the categorical variables versus the overall disadvantaged category. As expected, the adoption of AFV in DACs is lower than that in non-DACs. Specifically, AFV adoptions in DACs and non-DACs are (121*100)/(813+121)= 12.96% and (194*100)/(1074+194)= 15.30%, respectively. PHTS has several independent variables, including household income, car share program, household size, household vehicle count, and home in Seattle city. Descriptive statistics suggest that more low-income households (income <$50,000) live in DACs. However, it has been observed that with the increase in household income levels, fewer high-income households are found to be in DACs compared to non-DACs. For example, 33% of people earning over $100,000 live in the DACs,

Table 6-1: List of Disadvantaged Indicators under Transportation and Environment Themes (*160*)

Major Disadvantaged Category	**Percentile of Disadvantaged Indicators**
Transportation Access Disadvantaged (Average of the percentiles)	Percentile percentage households with no vehicle available estimate
	Percentile of Mean commute time to work (in minutes) (longer commute times)
	Percentile of Transportation Cost Burden: • Total workers 16 or older in a census tract • Percentages of non-transit households who have 0/1/2/3+ vehicles • Number of non-transit workers • Number of transit users 16 and over • Average weekday vehicle miles traveled per state • Calculated Average Annual Vehicle Miles Traveled • Average Annual Median Earnings • Five Year average price of gas per state • Five Year average gas mileage per state • Calculated average number of cars per household • Calculated average cost of owning a car • Calculated national average annual cost of using transit • Calculated average annual cost of transportation • Annual Travel Time in Minutes • Annual Travel Time in Hours • Travel Time Cost • Calculated average annual cost of transportation as a percent of income
	Percentile of National Walkability Index
Environmental Disadvantaged (Average of the percentiles)	Percentile for % pre-1960 housing (lead paint indicator)
	Percentile for Diesel particulate matter level in air
	Percentile for Air toxics cancer risk
	Percentile for Air toxics respiratory hazard index
	Percentile for Ozone level in air
	Percentile for PM2.5 level in air

and 47% live in non-DACs. 45% of people who have more than one vehicle live in DACs. In the sample, homeownership in DACs (50%) is lower compared with non-DACs (63%). Home ownership refers to the residence tenure status. If the household owns the residence, the variable is coded as 1; otherwise, 0. One of the significant advantages of owning a home in the US is the increased accessibility and control over the living space. Homeowners can make modifications and investments tailored to their preferences and needs. This autonomy extends to technological advancements, such as home charging facility installation. The availability and utilization of these facilities in owned homes can positively impact the adoption and usage of AFVs. Car sharing status is similar in DACs and non-DACs. Car sharing status indicates the regular participation of the household in a car share program (e.g., ZipCar, Car2Go, GIG). It can correlate with travel behavior and attitude towards new technologies, e.g., AFVs. Therefore, it is considered a surrogate measure of the tech-savviness of the household. The household size in the data varies from a minimum of one to a maximum of seven members. AFDC provides infrastructure data, i.e., the number of public charging station data at the census tract level. Overall, charging stations vary in all census tracts from a minimum of zero stations to 18 stations in a tract. However, the average number of charging stations in DACs (1.13) is expectedly lower than in non-DACs (1.32).

Figure 6-2 illustrates the distribution of households across the DACs and non-DACs of the Greater Seattle area in the sample while considering the overall disadvantaged category. The census tracts shown in Figure 6-2 reflect where the respondent lived when they completed the survey, not necessarily where they lived when they first adopted their AFV. Notably, Seattle is the second-best tech city in the US. The availability of household travel data (PHTS) gives us a unique opportunity to explore the AFV adoption scenarios in the DAC and non-DACs (i.e., census tracts) of the Greater Seattle area. This will help us understand the future of AFVs, the infrastructure needed, and the policy to implement. Notably, the 2021 Puget Sound Household Travel Survey was professionally administered and used an address-based sampling technique to select and invite households to participate. This involves drawing a random sample of addresses from all the residential addresses in each of the four counties of the Puget Sound Region. A comparative table with the distributions of the sample survey obtained and the representative population for key variables is shown in Table 6-3. Household size distribution is quite similar across the sample and the representative population. Lower-income households (<50,000) are slightly overrepresented (~1% difference), and higher-income households are slightly under-represented in the study sample. Multi-vehicle households are also slightly underrepresented in the sample. Given the sophisticated sample bias correction process involved in the PHTS, this study refrains from introducing additional statistical adjustments to the sample to avoid further complexity. Nonetheless, this limitation of the study sample is acknowledged, and we will approach the interpretation of the modeling results with caution.

6.6 Model

6.6.1 XGBoost Model Specifications

In this study, we adopted both XGBoost and binary logit models to assess the consistency of results between the two approaches and explore the XGBoost model's predictive power in the context of our research. Various machine learning (ML) techniques, such as random forests, gradient boosting, K-means clustering, and ensemble tree learning, have been used to analyze survey data related to new mobility options and alternative fuel vehicles due to their capacity for accurate

Table 6-2: Descriptive Statistics (N = 2,202)

Variables	Description	Freq.	%	Overall Disadvantaged Community? *			
				No		Yes	
				Freq.	%	Freq.	%
AFV Adoption	No	1887	86%	1074		813	
	Yes	315	14%	194	15.30%	121	12.96%
Household Income	<50,000	599	27%	255	20.11%	344	36.83%
	>=50,000 - <75,000	361	16%	186	14.67%	175	18.74%
	>=75,000 - <100,000	341	16%	230	18.14%	111	11.88%
	>=100,000	901	41%	597	47.08%	304	32.55%
Multi-vehicle	No	1036	47%	521		515	
	Yes	1166	53%	747	58.91%	419	44.86%
Home in Seattle City	No	1786	81%	1076		710	
	Yes	416	19%	192	15.14%	224	23.98%
Home Own	No	938	43%	470		468	
	Yes	1264	57%	798	62.93%	466	49.89%
Car Share Program	No	2112	96%	1217		895	
	Yes	90	4%	51	4.02%	39	4.18%
Transportation Disadvantaged	No	1478	67%	983		495	
	Yes	724	33%	285	22.48%	439	47.00%
Environmental Disadvantaged	No	680	31%	625		55	
	Yes	1522	69%	643	50.71%	879	94.11%
Equity Disadvantaged	No	691	31%	666		25	
	Yes	1511	69%	602	47.48%	909	97.32%
Economy Disadvantaged	No	1281	58%	1124		157	
	Yes	921	42%	144	11.36%	777	83.19%
Health Disadvantaged	No	1521	69%	1037		484	
	Yes	681	31%	231	18.22%	450	48.18%
Resilience Disadvantaged	No	83	4%	83		0	
	Yes	2119	96%	1185	93.45%	934	100.00%
Overall Disadvantaged	No	1268	58%	NA		NA	
	Yes	934	42%	NA	NA	NA	NA

*Note: For binary variables, the percentage is shown for the "Yes" category only

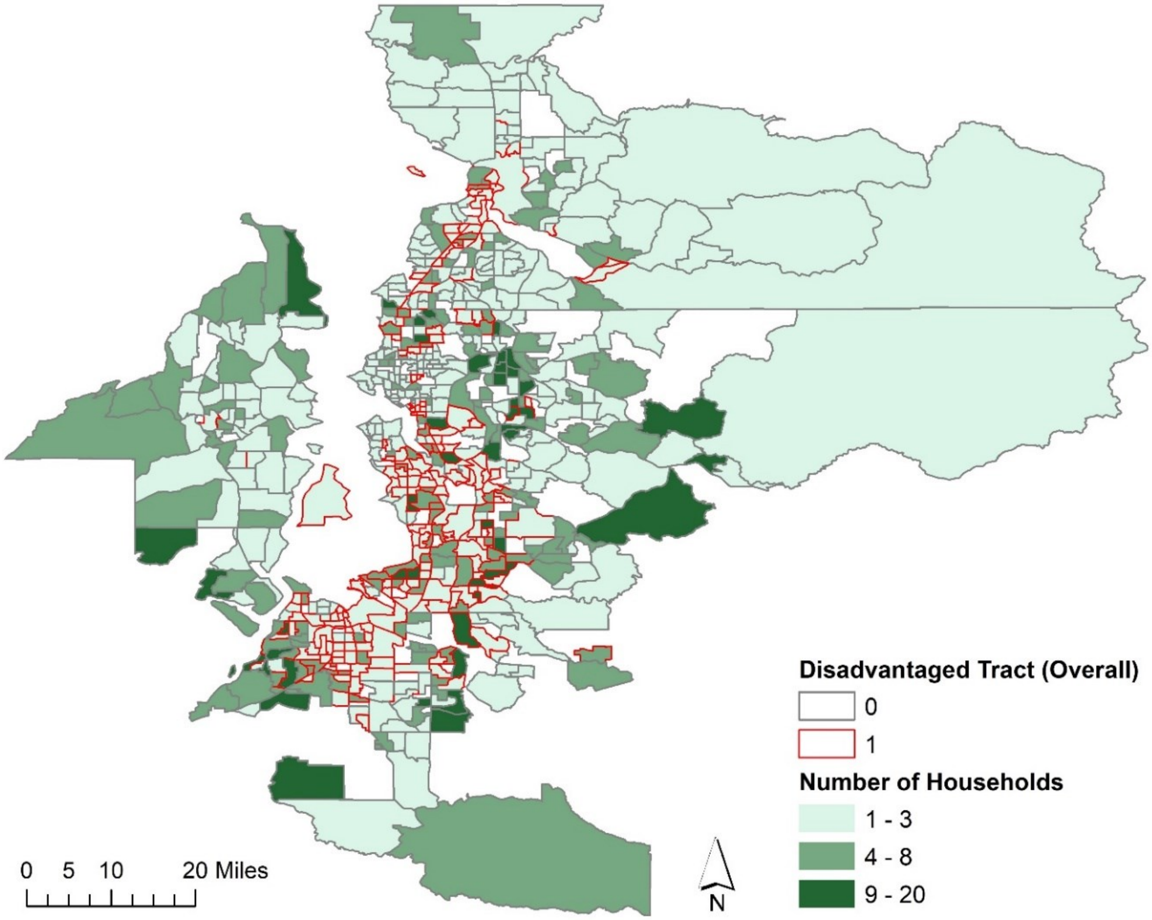


Figure 6-2: The distribution of households across the overall disadvantaged communities of the Greater Seattle area in the sample (Generated by the Authors)

Table 6-3: Distributions of demographics in the survey sample and representative population

Variable	Description	Sample Percentage	Representative Population Percentage
Household Size	1 Person	26%	27%
	2 Person	39%	36%
	3 Person	16%	16%
	4 Person	12%	14%
	5+ Person	7%	7%
Household Income	<50,000	27%	26%
	>=50,000 - <75,000	16%	14%
	>=75,000 - <100,000	16%	15%
	>=100,000	41%	45%
Multi-vehicle	No	47%	42%
	Yes	53%	58%

prediction. For instance, Lee et al. (*239*) estimated gradient boosting ML models to explore user preferences regarding automated vehicles. Similarly, de Rubens (2019) estimated the K-means clustering model for market segmentation analysis of electric vehicle adopters. One noteworthy ML technique in this domain is Extreme Gradient Boosting (XGBoost), a relatively recent addition to the ML toolkit. We have chosen XGBoost as the machine learning technique due to its specific advantages in our research context. XGBoost is a powerful algorithm known for its capacity to handle both categorical and continuous variables, accommodate non-linear relationships, and deliver high predictive accuracy (*240; 241*). Furthermore, XGBoost is recognized for its explainable artificial intelligence (XAI) capabilities, making it suitable for extracting insights from complex survey data. It offers computational advantages, including speed and scalability, crucial for efficiently analyzing survey datasets. The interpretability of XGBoost models is a paramount advantage, allowing us to elucidate and communicate the underlying drivers of AFV adoption. This is in line with the demands of our study, which seeks to uncover nuanced patterns (e.g., feature importance, supervised clustering) for adopting alternative fuel vehicles. Researchers have used XGBoost to find important features in different transportation research. For example, in a previous study by Meng et al. (*242*), XGBoost was utilized to predict the frequency and duration of traffic accidents by leveraging important features from multiple data sources, such as road geometric design, historical accident data, and traffic and weather data. Importantly, researchers found that XGBoost performs significantly better than other approaches (*243*). For example, Ullah et al. (*244*) demonstrated the superior performance of XGBoost compared to various machine learning techniques, such as random forest, categorical boosting, and light gradient boosting machines, for predicting the charging time of electric vehicles. Therefore, our choice of XGBoost is well-justified based on its computational advantages, interpretability, and empirical success in transportation research, which position it as a robust tool for extracting valuable insights from our survey data.

The primary goal of the XGBoost algorithm is to find a function $\oint(X_i)$ that best analyzes the dependent variable y_i from the explanatory variables X_i. It can be described with the equation below:

$$\hat{y}_i = \oint(X_i) = \sum_{k=1}^{K} f_k(X_i), f_k \in F \qquad \text{(i)}$$

Where K denotes the number of iterations. $\oint(X_i)$ represents an ensemble model consisting of base learners $f_k(X_i)$. F is the tree space. The learning objective of XGBoost is like the traditional decision tree model, which is to select splits that optimize the training loss. However, XGBoost can improve gradient tree boosting by regularizing the learning objective. Let K be the additive functions and n is the number of instances. The goal is to minimize the following Equation (ii):

$$\mathcal{L}_k = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \qquad \text{(ii)}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{i=1}^{T} \omega_{ik}^2 \qquad \text{(iii)}$$

Where l is the loss function, and Ω penalizes the complexity of the model. The complexity of f_k will be censured for high values of T & high variations of weights ω_i for T leaves. Thus, XGBoost can learn better and improve tree structure than the traditional tree model. Equation (i)-(iii) are solved for optimal ω_j^*.

After having an optimal model represented by $\oint(X_i)$ with hyperparameter tuning, its prediction results can be analyzed. Notably, the primary interest is to recognize the contribution of each feature to the response prediction by the model. Shapley Additive Explanations (SHAP) is an established and consistent attribution method that can deliver a robust explanation of the XGBoost results. The sum of all the feature importance is:

$$\oint(X_i) \approx S(z_i) = \varphi_0 + \sum_{j=1}^{m} \varphi_{ij} z_{ij} \quad \text{(iv)}$$

$S(z_i)$ is the explanation model intended to explain the output $\oint(X_i)$ inherently. $z_{ij} \in [0,1]^m$ equal to 1 when a variable j is observed, otherwise 0; φ_j is the variable importance that can be attributed to variable j; φ_0 is the output of the base model with no features. According to Lundberg & Lee (*245*), φ_{ij} can be exhibited as follows:

$$\varphi_{ij} = \sum_{r \subseteq X_i} \frac{(m-|r|)!(|s|-1)!}{m!} [\oint(r) - \oint(r \backslash j)] \quad \text{(v)}$$

Here, r denotes all the distinct subsets of X_i, & $1 \leq |r| \leq m$ is their order. The above equation also represents the overall impact of variable j on the model by the weighted sum of the marginal effect $[\oint(r) - \oint(r \backslash j)]$. Overall, SHAP enables the model's prediction to be explained reliably. The method is executed using the SHAP Package of Python programming language.

6.6.2 Binary Logistic Model Specifications

The binary logistic regression models are adopted to compare and validate the feature importance and relationships determined by the XGBoost model and analyze the DACs and non-DACs-specific covariates of AFV adoption (i.e., segmentation). Previous studies often used discrete choice models to analyze vehicle type, including binary logit models (*246-249*). This study analyzes the AFV adoption of an individual household, which is a binary outcome variable. Therefore, the estimation of the binary logit model is adopted in this study. The following Equation (vi) illustrates the modeling approach of the binary logistic regression:

$$P(Y_i = whether\ an\ individual\ household\ owns\ AFV) = \frac{\exp(\alpha + \beta_i X_i + \varepsilon)}{1 + \exp(\alpha + \beta_i X_i + \varepsilon)} \quad \text{(vi)}$$

Where, Y_i is the AFV adoption function determining the AFV adoption by household *i* for the Seattle household data. $Y_i = 1$ indicates a household owning AFV, 0 otherwise. X_i are the explanatory variables; β_i is a vector of estimable coefficients corresponding to X_i; α is the intercept term; and ε is the error term. The coefficients of the binary logit model don't have a sense of magnitude. Therefore, marginal effects are calculated to identify the magnitudes of the estimated coefficients. The following marginal effect equation (vii) of the explanatory variable X_i shows the prediction function's partial derivative, the instantaneous rate of change.

$$\frac{\partial P}{\partial X_i} = \frac{\beta_j e^{-(\alpha + \beta_i X_i + \varepsilon)}}{[1 + e^{-(\alpha + \beta_i X_i + \varepsilon)}]^2} \quad \text{(vii)}$$

6.6.3 Model Segmentation

Given an uneven distribution of AFV users between transportation-disadvantaged and non-disadvantaged communities, we use segmented models. The segmented methodology followed in this study is based on the studies conducted by Misra and Atkins (*250*) and Nitesh and Cherry (*251*). In this approach, two separate binary models are estimated for disadvantaged and non-disadvantaged communities. Then, a pooled binary logit model is estimated by joining both groups.

A likelihood ratio (LR) test is performed to understand whether the segmented models are the same as the pooled model or not. The test can be described as follows:

$$-2[LL(pooled\ model) - LL(disadvantaged\ only\ model) - LL\ (not\_disadvantaged\ \ only\ model) \sim \chi^2{}_{df}$$

where,

$degree\ of\ freedom\ (df) = K_1 + K_2 - K$, where $K_1$and K_2 is a degree of freedom of segmented models (disadvantaged and non-disadvantaged models) and K is a degree of freedom for the pooled model.

6.7 Results and Discussion

6.7.1 XGBoost Feature Importance

We initially used all the input variables in the XGBoost model to identify the most critical variables that can predict AFV adoption in DACs. The model's prediction accuracy is 91.52%. Furthermore, the model accurately predicts 84% out of 86% non-AFV adoption households and 8% out of 14% AFV adoption households. The XGBoost model's performance is shown graphically by a receiver operating characteristic (ROC) curve in Figure 6-3. The plot shows the true positive rate against the false positive rate at various classification thresholds. The curve is closer to the top-left corner, indicating a higher ability to correctly classify positive instances while minimizing false positives. The area under the ROC curve is 0.93, which suggests that the model can distinguish well between the positive and negative classes. Therefore, the model exhibits a high level of discrimination and performs well in predicting AFV adoption. Figure 6-4 illustrates the estimated SHAP feature importance plots from the XGBoost model. These plots indicate that there is no apparent overfitting issue in the model. The plot shows a subset of features with significantly higher importance than others, which is consistent with domain knowledge. The model has successfully captured the relevant patterns and relationships in the data. These plots can be explained both globally and locally. The global explanation uses the average magnitude of SHAP (i.e., in log-odds units). It shows the model output while a feature is deleted from the model.

Figure 6-4 (right) indicates that household income is the most important feature, which is followed by other independent variables, including household size, home ownership, public charging stations, multi-vehicle households, home in Seattle city, and car share program. Among the disadvantaged variables, the environment is the most important, followed by equity, transportation access, environmental, transportation access, overall, economy, and resilience disadvantages. Nonetheless, local interpretations of the variables need to be analyzed to uncover the direction in which the variables are important.

The local interpretation plot is presented in Figure 6-4 (left), in which every dot indicates each sample's location, and each color represents the value of the respective feature for that location. Higher household income is positively correlated with AFV adoption. Correspondingly, public charging stations, multi-vehicle households, home ownership, large households, home in Seattle city, and using a car share program are positively associated with AFV adoption. Importantly, equity and overall disadvantaged categories are positively related to the adoption of AFV, whereas environmental, transport, and economic DACs are negatively associated with AFV adoption. Additionally, it is evident that the resilience category is not a particularly important feature in predicting AFV adoption.

To compare and validate the feature importance determined by XGBoost, a binary logistic

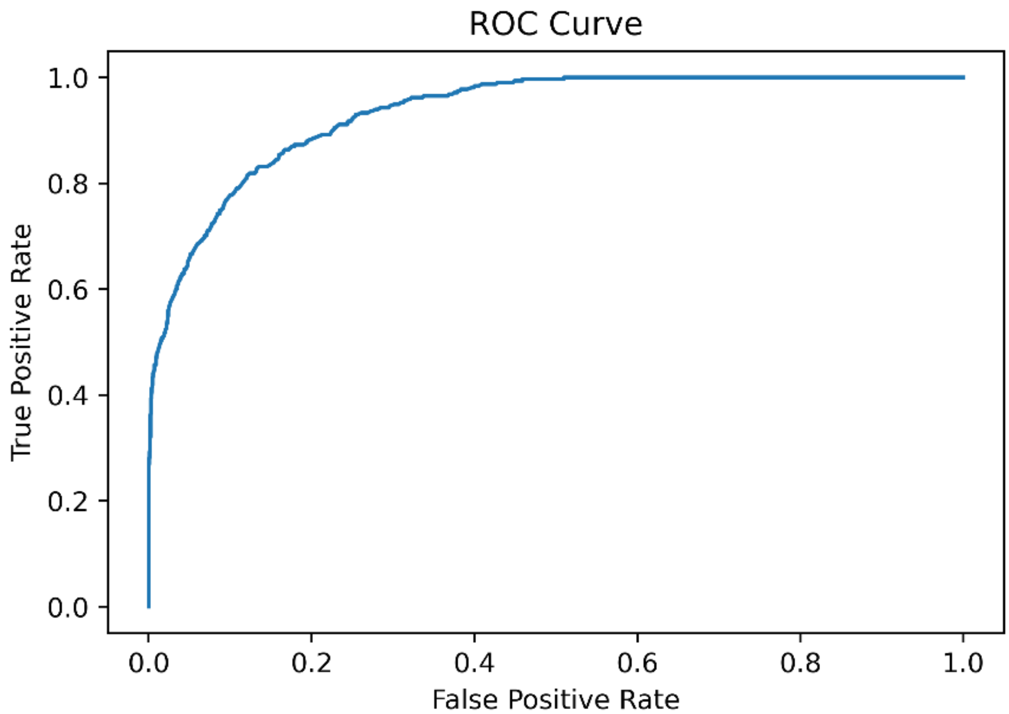


Figure 6-3: ROC curve of the XGBoost model

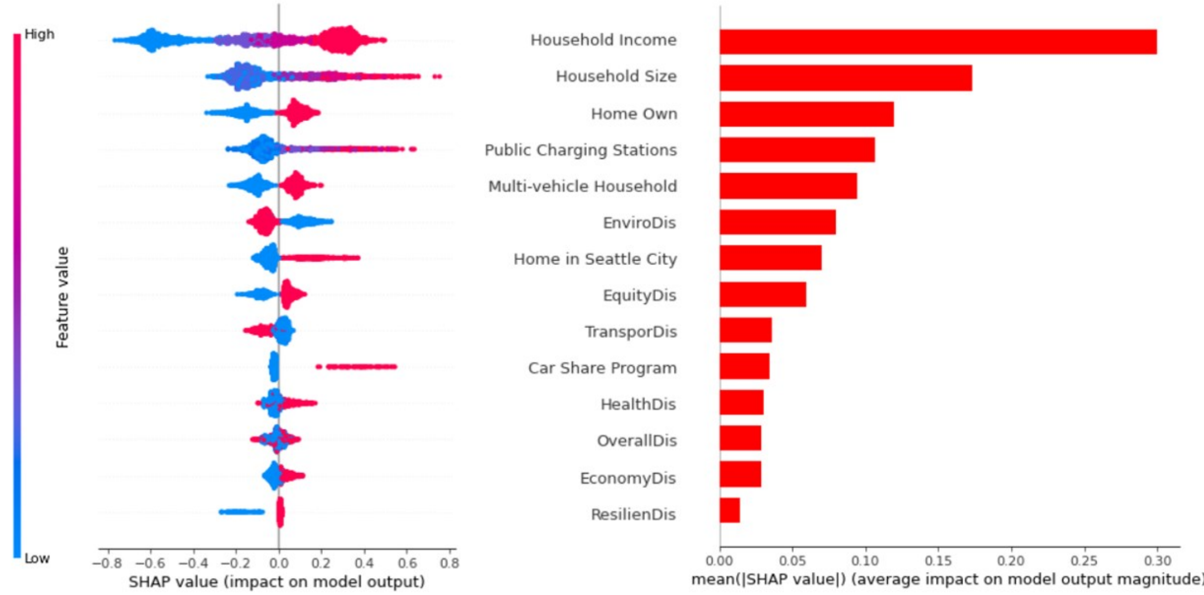


Figure 6-4: Local (left) and Global feature importance plots by SHAP

model is also employed by incorporating all variables and subsequently computing the standardized coefficients for each variable. The standardized scores display the feature importance plot. The plot is shown in Figure 6-5, exhibiting household income as the most important feature to predict AFV adoption, followed by household size, home in Seattle city, and public charging stations. Environment and Equity disadvantages are the top DAC indicators. This reveals a significant consistency in the importance of most variables between the binary logit model and the XGBoost feature importance. This consistency implies that despite their different methodologies, both approaches identify similar features as influential in predicting AFV adoption. The analysis provides robust evidence and increases the reliability of the model's feature selection process.

6.7.2 Supervised Clustering

XGBoost has an interesting application that enables supervised clustering – where clustering on the feature attributions can be performed instead of using unsupervised clustering methods. Supervised clustering is immune to a challenging problem of unsupervised learning: the determination of feature weights. The unsupervised clustering method cluster features with different units, whereas the supervised cluster uses feature attributions to manually convert all the independent variables into values with the same units as the response variable of the model. It indicates that a unit change in any variable is comparable to a unit change in other variables (i.e., feature attribution). Fluctuations in the variables only affect the clustering when those clusters correlate with the response variable. The XGBoost prediction was clustered using hierarchical agglomerative clustering (similar to a dendrogram plot to join the samples). Figure 6-6 demonstrates supervised clustering with SHAP feature attributions of our study data. The X-axis refers to all the 2,202 observations in the data, and the Y-axis presents the log odds of AFV adoption. Red feature attributions push the score higher; in other words, they push households to adopt AFV. At the same time, the blue attributions feature pushes the score to be lower and discourages adopting AFV. A few such noticeable subgroups with related features are annotated in Figure 6-6. This proves the power of supervised clustering to identify the group that shares the common features related to AFV adoption. For example, high-income households, large households, and equity-disadvantaged communities are more likely to increase AFV sales. In contrast, few public charging stations and environmentally disadvantaged communities discourage people from adopting AFV. Similarly, low-income, few public charging stations and single-vehicle and small households in transport DACs are less likely to increase the adoption of AFV in the future. On the other hand, high-income households living in Seattle city are more inclined to buy AFV. Likewise, not overall and not environmental DACs are best suited for AFV adoption. Overall, the supervised clustering and the XGBoost feature attributions plots show that household income, public charging stations, multi-vehicle households, home ownership, household size, home in Seattle city, and car share program are the most important predictors for AFV adoption. Moreover, the XGBoost feature extraction shows equity and environmental DACs are the most critical disadvantaged indicators in the sample. These will be considered in the modeling to understand AFV adoption in the communities with these disadvantages.

6.7.3 Results of the Segmented Binary Logistic Models

The important predictors found in the XGBoost feature attributions are modeled using a binary logistic regression model with segmentations to analyze AFV adoption. Prior to model estimation,

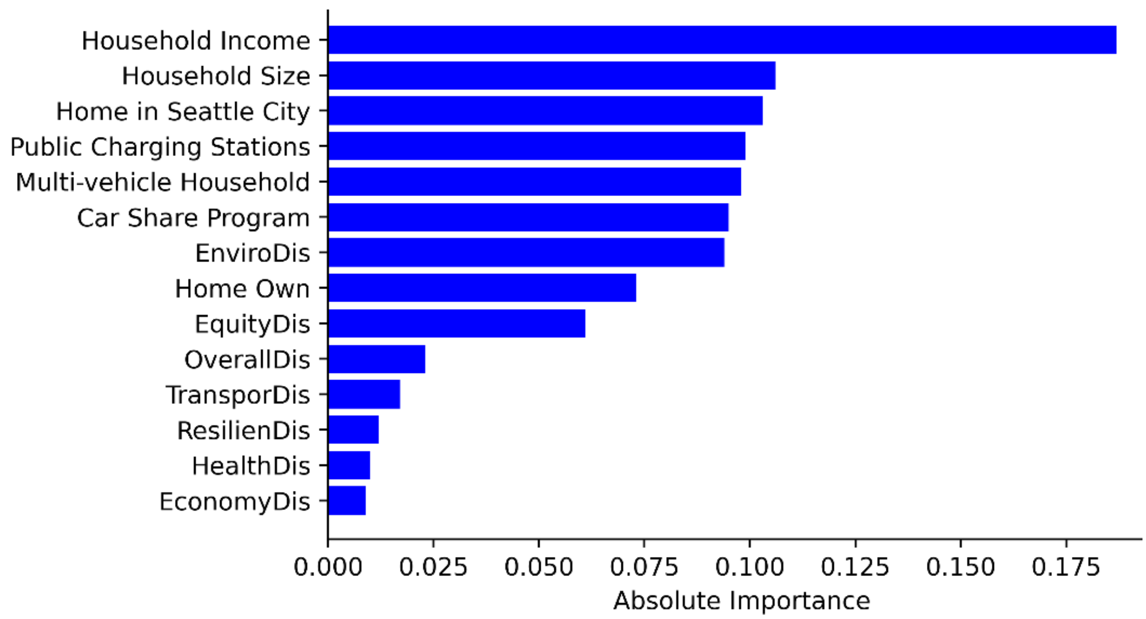


Figure 6-5: Feature importance plot from binary logit standardized coefficients

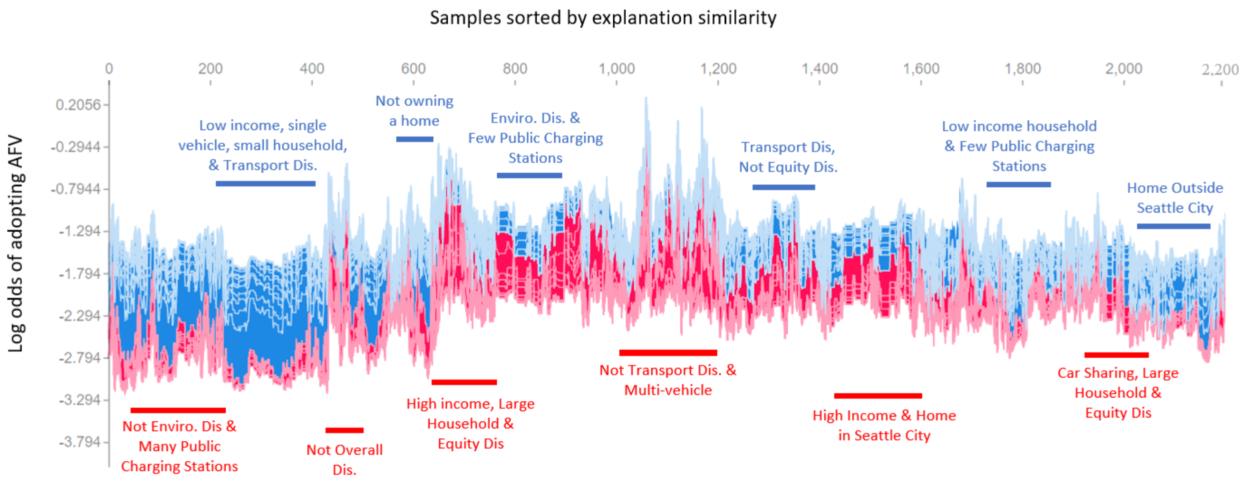


Figure 6-6: Supervised clustering of the study data with SHAP feature attributions

correlations between household income and the DAC indicators are examined. The results in Table 6-4 show no high correlations exist in the variables. Furthermore, the variance inflation factor (VIF) is utilized to investigate the multicollinearity problem in the data. The higher the value of the calculated VIF, the higher the potential for collinearity between the related independent variables used in the model. Generally, VIF values less than 3 suggest the absence of multicollinearity in the model (*252*). VIF can be expressed in Equation (viii) below. Where, R_j^2 is the multiple correlation coefficients for the independent variables. If R_j^2 equals zero, VIF equals 1, and no correlations exist between independent variables. The estimated VIFs for the study data show no signs of multicollinearity between the variables.

$$VIF_j = \frac{1}{1-R_j^2} \qquad \text{(viii)}$$

Table 6-5 presents the results for the impacts of the underlying contributing attributes for the AFV adoption. The first model is the pooled model, and the remaining two are disadvantaged-only and non-disadvantaged-only models. By using the LR test, we test the null hypothesis that the pooled model is identical to the segmented models. The corresponding test has a p-value close to zero, which confirms that the null hypothesis can be rejected at 99% confidence and concludes that the segmented models can sufficiently explain the data separately. All three binary logistic models with the coefficients, significance level, and related marginal effects, are reported in Table 6-5. The direction of all estimated coefficients is similar to the XGBoost prediction.

The number of public charging stations is found to be statistically significant and positively associated with the AFV adoption intent in all three models. More public charging/refueling stations show a strong correlation with AFV adoption in DACs. Specifically, in the disadvantaged-only model, the increase in the number of public charging stations by one unit is associated with the increase of AFV adoption by 2%. This is consistent with the study conducted by White et al. (*231*). The impact becomes larger with the increase of public charging stations in DACs compared with non-DACs. The greater density of charging stations comes with several potential benefits. For example, it reduces the lower-range anxiety and mobility restriction of AFVs. More charging stations provide greater flexibility for AFV usage through a better functional range. This can also be related to the social pressure or neighbor effect people may feel about buying AFVs. More charging stations encourage more AFV adoption and influence other people in the neighborhood to purchase AFVs.

Moreover, predicted margins are estimated to explore the differential effects of infrastructure on AFV adoption across the DACs and non-DACs. Figure 6-7 shows the predictive margins of public charging stations across DACs and non-DACs. Predicted margins are nothing but the predicted probabilities for different values of an independent variable while holding other variables at their mean (*253*). For example, in Figure 6-7, if all households in the sample live in disadvantaged tracts having four public charging stations, they have a 20% chance of AFV adoption. The effect increases for both DACs and non-DACs with the increase in the number of charging stations. However, DACs have higher impacts on AFV adoption with an increase in public charging stations than non-DACs. The higher chances in DACs may be related to the segmented effects of income and tech-savviness of the households discussed below.

The results suggest that the coefficient for *household income* categories is positive and statistically significant. In the pooled model, if a household earns an income of ">=$50,000-<$75,000", it is 6% more likely to adopt alternative fuel vehicles than a low-income household

Table 6-4: Correlations between household income and the DAC indicators

	Income	Transport DAC	Health DAC	Economy DAC	Equity DAC	Resilience DAC	Environment DAC
Income	1.00						
Transport DAC	-0.01	1.00					
Health DAC	-0.08	0.04	1.00				
Economy DAC	-0.21	-0.03	0.16	1.00			
Equity DAC	-0.10	-0.11	-0.11	0.38	1.00		
Resilience DAC	0.01	-0.11	-0.09	0.13	0.35	1.00	
Environment DAC	-0.03	-0.06	-0.05	0.25	0.24	0.28	1.00

Table 6-5: Segmented Model Results

	Pooled Model		Disadvantaged-Only		Non-Disadvantaged Only	
	Coef.	**Marginal Effect**	**Coef.**	**Marginal Effect**	**Coef.**	**Marginal Effect**
Constant	-3.585***	-	-3.64***	-	-3.77***	-
Household Size	0.170***	0.025	0.15	0.016	0.19***	0.02
Multi-vehicle (Base: No), Yes	0.376**	0.042	0.43	0.044	0.32*	0.04
Household Income (Base: <$50,000),						
>=$50,000-<$75,000	0.680***	0.06	1.07***	0.11	0.41	0.03
>=$75,000- <$100,000	0.889***	0.085	0.35	0.03	1.20***	0.12
>= $100,000	1.022***	0.102	0.84**	0.08	1.16***	0.11
Home in Seattle City (Base: No), Yes	0.523***	0.067	0.3	0.034	0.53***	0.07
Home Ownership (Base: No), Yes	0.283**	0.032	0.44*	0.044	0.23	0.03
Car Share Program (Base: No), Yes	0.923***	0.136	2.08***	0.359	0.51*	0.07
Public Charging Stations	0.071***	0.01	0.15**	0.02	0.05**	0.01
Equity Disadvantaged (Base: No), Yes	0.266*	0.03	0.45	0.046	0.24	0.03
Environment Disadvantaged (Base: No), Yes	-0.369**	-0.044	-1.42***	-0.162	-0.14	-0.02
Model Fit Statistics						
N	2,202		724		1,478	
Log-likelihood	-839.79		-258.75		-567.84	
Chi (12)	111.67		60.26		93.9	
Model Significance Test	0		0		0	
AIC	1,705.58		543.5		1161.68	
BIC	1,779.65		603.1		1230.56	
LR test for segmentation	Chi2 (12) = 26.405, Prob>chi2 = 0.009					

**** p<0.01, ** p<0.05, * p<0.10*

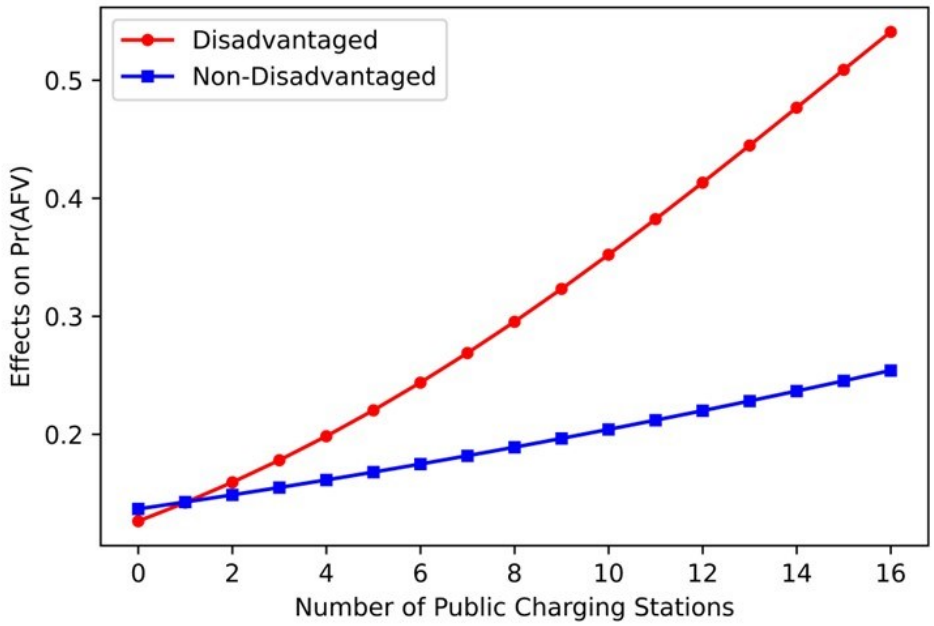


Figure 6-7: Predictive margins of public charging stations across DACs and non-DACs

(<\$50,000). This effect increases with the increase in household income, which is as we expected. For example, households having income ">\$100,000" are 10.2% more inclined to buy AFVs. This is also aligned with the previous studies (*212; 213; 225*). Several reasons may play a role. First, due to the high purchase cost of AFV, it is not surprising that higher household income is an important influencing factor for adopting AFV. Li et al. (*213*) found a significant and positive correlation between alternative fuel vehicle type and income. Second, higher-educated people are generally believed to have higher incomes and are usually more enthusiastic about welcoming and adopting new technologies and sustainable vehicles (*224*).

The segmented models show some variations in *household income*. For example, in the disadvantaged-only model, households having an income of ">=50,000 - <\$75,000" are found to be 11% more likely to adopt AFVs than households having an income of "<\$50,000". Whereas the same category for the non-disadvantaged-only model is found to be statistically not significant. The chance of buying AFVs for the ">\$100,000" category in DACs is 8%, which is lower than the ">=\$50,000 - <\$75,000" category for DACs. Although this appears to be counter-intuitive, this is consistent with Lee et al. (*254*), who suggested this change might result from saturation among high-income AFV owners and diffusion to the other segment of income strata. The increased lower cost and affordability of newer AFVs could incentivize middle-income households to purchase more AFVs (*216*).

Moreover, according to Bas et al. (*255*), having a pro-AFV attitude in middle-income groups may also contribute to adopting AFVs. This is further supported by the findings from the *car share program* variable (i.e., tech savviness), which is significant in our analysis. In the pooled model, the probability of adopting AFV for households that use a car share program is 13.6% higher than the ones that don't use a car share program in a big city. It may be because car sharing may change individual travel behavior, which can impact the frequency at which people make trips, levels of private car ownership with various vehicle choices, and the eagerness to use newer technologies (*228*). Car sharing status represents tech-savvy people who show interest in using different technologies in daily activities. Therefore, they would be more attracted to the benefits of AFVs over conventional vehicles. For example, Khattak and Khattak (*256*) found that people interested in technologies are more likely to purchase AFVs than conventional fossil fuel vehicles. Segmented models show there are significant differences between DACs and non-DACs in AFV adoption by tech-enthusiast people. In DACs, people interested in technologies are 36% more likely to adopt AFVs, whereas the adoption is 7% in non-DACs.

The analysis reveals that *home ownership* has a positive and statistically significant impact on AFV adoption in both the pooled and disadvantaged-only models. The findings suggest that households in DACs who own a home are 4.4% more likely to adopt AFVs. This could be associated with the fact that owning a home offers better opportunities for overnight home charging facilities. Besides, *home in Seattle city* is significant in the pooled and non-disadvantaged only models. The results indicate that the households living in Seattle city (i.e., urban areas) are 6.7% more interested in adopting AFV than those living outside Seattle (i.e., rural or suburban areas). It may be because of the high exposure of AFV in urban areas with favorable built-in infrastructure. Moreover, most urban and city areas are segregated regarding socioeconomic factors and ethnic characteristics (*257*). Therefore, people living near each other are most likely to be similar, which may also be the indictment for AFV ownership. The more people adopt AFV, the more desirable it becomes for others in the neighborhood (*222*).

Household size is statistically significant and positive. It indicates that an increase in the household size by one is 2.5% more likely to increase AFV adoption. This is consistent with Plotz et al. (*258*) and Jia et al. (*214*). The former found that multimember families are more likely and enthusiasts to adopt AFV, specifically electric vehicles. The latter found a higher correlation between average household size and the number of alternative fuel vehicles in the U.S. *Multi-vehicle* household is also significant and positive, which indicates that the household vehicle size of more than one induces the households to buy AFV. It is consistent with Jia et al. (*214*), who stated that the number of vehicles in the household tends to affect the future purchase of a vehicle. Households with more vehicles tend to buy newer vehicles with newer technologies. The findings of Li et al. (*213*) are also in line with the result, which stated the percentage of AFV households with more than one vehicle is much greater than that of households with conventional vehicles.

The important disadvantaged categories are also explored in all three models. The *equity disadvantaged* category is statistically significant in the pooled model. Also, the *environment disadvantaged* is significant and negative in the pooled and disadvantaged-only models. The coefficient of equity disadvantaged is positive, indicating that households living in an equity disadvantaged community are 3% more likely to use alternative fuel vehicles. In an equity disadvantaged area, there may be disparities in terms of income, race, infrastructure investments, technology innovations, and deployments. High-income households and high local populations (e.g., more white population and fewer people of color or vice-versa) are privileged in such areas in terms of offer incentives, such as tax credits, rebates, or free charging, by government and private entities for purchasing or using AFVs (*216; 259*). In contrast, if the households live in environmental and transportation DACs, they are 4.4% less inclined to buy AFV than the ones who do not live in such disadvantaged communities. Transportation and Environment DACs are generally more congested and polluted than non-DACs. Average trip length and duration are higher in a transportation DAC than in a non-transportation DAC (*160*). The level of pollutants, i.e., diesel particulate matter, ozone, and PM2.5 in the air, are high in an environmentally disadvantaged community (*160*). Air-toxic respiratory hazards and air-toxic cancer risks are especially elevated in those areas. Although it is expected that more people in the DACs should adopt AFVs to improve the overburdened by pollution, a high percentage of higher-educated and high-income households are generally discouraged from living in such areas. Therefore, most people living in these areas may not be able to afford to buy AFVs.

6.8 Policy Implications

The findings of this paper have important policy implications. This study can help policymakers, city officials, or other stakeholders recognize the issues of current AFV sales and adoption from the perspective of related correlates. The findings deepen our understanding of disadvantaged communities' social and economic challenges. This is vital for overcoming environmental justice and equity issues. Investments in DACs charging/refueling infrastructure can have a higher impact on AFV adoption, especially if they consider the insights from XAI in the decision-making process and emphasize the importance of considering social and economic factors in AFV adoption programs. Homeownership offers the distinct advantage of having personal home charging facilities for AFVs, a key factor in the convenience and feasibility of owning such vehicles. However, the substantial income needed to purchase a home places this benefit out of reach for middle and lower-income categories. To bridge this gap and encourage the adoption of AFVs among these communities, the installation of charging stations in the apartment complexes might be as a viable solution. This approach not only provides necessary infrastructure for AFV owners but also ensures that the benefits of this technology are more equitably distributed, enabling broader access regardless of homeownership status.

Furthermore, as we know, the relatively high cost of purchasing EVs is a critical barrier in DACs. However, with new incentives, the cost of EVs is decreasing. Specifically, the US federal government provides a tax credit of up to $7,500 for new EVs purchased in or after 2023 (*260*), and used EVs are now eligible for tax credits of up to $4,000 (*261*). Additionally, auto manufacturers can articulate customized sales and incentives, e.g., financial incentive policies and other policy developments, to expand AFV adoption to underserved communities. For example, in 2021, Chevrolet provided a cashback incentive of $7,000 to purchase a new Bolt EV (*262*). Furthermore, manufacturers can identify the most attractive features of AFVs to customers, such as technological and environmental innovation, performance, and operating costs comparison between AFVs and conventional fossil fuel vehicles. Federal investment initiatives and state and public utility agencies may use the AFV information in this study to determine the required infrastructure, e.g., the number and size of the charging/ refueling stations, to install AFVs in underserved communities. By encouraging AFV adoption in DACs, policymakers can promote decarbonization and improve public health, equity, and environmental justice.

6.9 Limitations

This study is not without some limitations. Our theoretical framework may have overlooked other important variables. For example, the inclusion of "average commute time" and "household's yearly mileage" could have added more insights into the analysis. However, these variables are not primarily part of the PHTS databases. Besides, a variety of AFV types are lumped together in this analysis due to not having enough data on all types of AFVs. Moreover, different policies, e.g., VMT-based tax or fuel tax, that can impact AFV adoption could not be explored.

6.10 Conclusions

This study aims to understand the adoption of alternative fuel vehicles in disadvantaged communities. By combining statistical and explainable artificial intelligence (XAI) techniques, the study explores the complex behavioral decision-making processes involved in AFV adoption and the role of various factors in this process. The study utilizes a unique and comprehensive database that links the 2021 Puget Sound Household Travel Survey, the US Department of Energy's AFV data, and the US Department of Transportation's Environmental Justice databases. The data statistic indicates that AFV adoption in DACs (12.96%) is expectedly lower than non-DACs (15.30%). The use of XAI, i.e., XGBoost with SHAP technique, has provided valuable insights and emphasized the importance of considering social and economic factors in AFV adoption programs. The XAI results suggest that lack of access to charging infrastructure, consumer attitudes, and income correlate with AFV adoption in DACs. The modeling results show that more public charging/refueling stations strongly correlate with AFV adoption in DACs. Tech-oriented households are more likely to adopt AFVs in the DACs. The results further point to the importance of home charging facilities while adopting AFVs in DACs. Moreover, AFV adoption is strongly correlated with household income in DACs. Results further indicate that households living in an equity DAC are more interested in buying AFVs, reflecting income, race, and investment disparity. This also highlights the importance of more investments and policy incentives to be taken in the future to increase AFV adoption in DACs.

The findings of this study are based on data from the greater Seattle area of the state of Washington. Hence, they should not be generalized to the context of other cities or states. Although the results are not generalizable to other areas, they will be helpful for similar research in other cities/regions. Moreover, the findings of this study provide important policy implications by offering insights into the factors influencing AFV sales and adoption and highlighting social and economic challenges in disadvantaged communities. A comprehensive discussion of these implications can be found in the *Policy Implications* section of the paper. Additionally, the findings provide a basis for future research on AFV adoption in disadvantaged communities and highlight the need for continued investment. Future research should analyze the AFV adoption in all the DACs of the US. Future studies may consider analyzing the adoption of different types of AFVs in DACs. Also, the relationship between home charging facilities and various types of homeownership, such as apartments, single-family homes, and townhouses, needs to be investigated in the future. Furthermore, it will be interesting to investigate AFV adoption in highly disadvantaged communities, e.g., above the 85th percentile. Finally, AFVs are being widely adopted by public transportation agencies. These vehicles (e.g., buses or service vehicles) are often different from conventional vehicles in design, fuel, and incentives, among others. Thus, examining such AFVs in more detail is important to get a complete picture of how AFVs are reshaping transportation.

6.11 Acknowledgments

This project was partially supported by the Collaborative Sciences Center for Road Safety (CSCRS), a US Department of Transportation National University Transportation Center.

CHAPTER 7 CONCLUSIONS

The COVID-19 pandemic has been an unprecedented systemwide shock that has had long-lasting and significant impacts on various aspects of our daily lives, including transportation. One notable impact has been a shift in travel behavior, with many people choosing to work from home and shop online, resulting in reduced commuting and non-essential trips. This shift has led to a reduction in miles driven and a positive impact on air pollution and carbon emissions. However, despite the decrease in the number of vehicles on the road, there has been a significant increase in crash fatalities during the pandemic, which has disproportionately affected disadvantaged communities. This raises concerns about the equity implications of pandemic-related travel behavior and road safety changes. This dissertation aims to investigate the changes in the transportation system during COVID-19 and explore their future implications, including travel behavior, technology adoption behavior, and road safety aspects in disadvantaged communities. The study uses advanced statistical and artificial intelligence techniques and comprehensive databases to address methodological issues such as spatial heterogeneity and unobserved endogeneity.

The study results suggest that during-pandemic online shopping was expectedly associated with lower in-person shopping trips. People who worked from home were associated with making more shopping trips. WFH went up from 12% to 61% during COVID-19, admittedly an unusual situation. The relationships among online shopping, physical shopping trips, and WFH, found in pre-pandemic data, are similar but differ in magnitude from the during-pandemic periods. Regarding safety, data statistics show that while crash fatalities increased by 8.2%, total crashes decreased by 15.3%, and the total harm cost was lower by about $1.76 billion during COVID-19 (2020) compared with pre-COVID-19 conditions (2019). The results indicate that compared to the pre-pandemic periods, fatal crashes that occurred during the pandemic are associated with more speeding & reckless behaviors and varied across jurisdictions. Fatal crashes are more likely to happen on interstates and dark-not-lighted roads and involve commercial trucks.

Furthermore, results show that DACs experienced heightened adversity than non-DACs regarding road safety during COVID-19. Transportation, health, and resilience disadvantaged tracts are found to be associated with more fatal crashes than non-DACs (an increase of 8% to 57%). Fatalities vary across different races in DACs. Regarding the travel behavior in DACs during COVID-19, households living in DACs are less likely to order online retail goods and groceries than non-DAC households. The probability of making more restaurant trips decreases if a household lives in a DAC. The findings also highlight the differences between the DACs' and non-DACs' in-person and online shopping activities regarding technology access, income, and equity. Finally, the analysis of DAC's decarbonization role through alternative fuel vehicle adoption (AFV) adoption using the Puget-Sound household travel survey conducted during COVID-19 suggests a lack of access to charging infrastructure, consumer attitudes, home ownership, and income play a substantial role in lower AFV adoption rate in DACs.

The results of this research have significant policy implications. First, planners can improve travel demand models by explicitly incorporating WFH as an alternative to commuting to work in trip generation and time-of-day models. Second, the findings in DAC's online shopping highlight the presence of a digital divide and digital poverty, which policymakers must address by promoting equal access to goods and services for households across all socioeconomic levels. The development of digital infrastructure is essential for eliminating the digital divide. This includes measures such as ensuring the availability and affordability of high-speed internet in all regions, including rural and underserved areas, and the creation of digital services and devices that are accessible to individuals with disabilities or limited technological proficiency. Third, the finding on AFV adoption in DACs can help evaluate the planning-level impacts of refueling or charging infrastructure in DACs, allowing these communities to benefit from infrastructure investments and pave the way to decarbonization. Importantly, the findings can inform policy and planning decisions aimed at promoting safer and more equitable traffic safety programs in DACs in the post-pandemic world. Specifically, the study reveals the importance of informing policymaking to strengthen digital traffic law enforcement through appropriate countermeasures, such as intelligent speed adaptation, digital warning signs, and dynamic speed limits in crash hotspots. Finally, the study emphasizes the risks associated with DACs and can aid in designing and implementing traffic safety interventions, e.g., digital countermeasures, to address road safety risks in DACs. Overall, the findings of this dissertation underscore the importance of better preparedness and planning for disadvantaged communities to be equipped to handle future systemic shocks.

REFERENCES

[1] Kwan, M.-P. Mobile communications, social networks, and urban travel: Hypertext as a new metaphor for conceptualizing spatial interaction. *The Professional Geographer,* Vol. 59, No. 4, 2007, pp. 434-446.
[2] Schwanen, T., M. Dijst, and M.-P. Kwan. ICTs and the decoupling of everyday activities, space and time: Introduction. *Tijdschrift voor economische en sociale geografie,* Vol. 99, No. 5, 2008, pp. 519-527.
[3] Kwan, M.-P. Beyond space (as we knew it): Toward temporally integrated geographies of segregation, health, and accessibility: Space–time integration in geography and GIScience. *Annals of the Association of American Geographers,* Vol. 103, No. 5, 2013, pp. 1078-1086.
[4] Dholakia, N., J. J. Xiao, R. R. Dholakia, and N. Mundorf. The impact of retail e-commerce on transportation: a conceptual framework. *Research Institute for Telecommunications and Information Marketing*, 2000.
[5] *QUARTERLY RETAIL E-COMMERCE SALES.* US Census Bureau. https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf. Accessed December 2, 2020.
[6] *2017 American Community Survey (ACS).* US Census Bureau. https://www.census.gov/programs-surveys/acs/news/data-releases/2017/release.html. Accessed December 2, 2020.
[7] BRENAN, M. *COVID-19 and Remote Work: An Update.* GALLUP. https://news.gallup.com/poll/321800/covid-remote-work-update.aspx. Accessed July 16, 2021.
[8] Drucker, J., and A. J. Khattak. Propensity to work from home: Modeling results from the 1995 Nationwide Personal Transportation Survey. *Transportation Research Record,* Vol. 1706, No. 1, 2000, pp. 108-117.
[9] Graaff, T. d. On the substitution and complimentarity between telework and travel: A review and application.In, 2004.
[10] Cao, X. J., Z. Xu, and F. Douma. The interactions between e-shopping and traditional in-store shopping: an application of structural equations model. *Transportation,* Vol. 39, No. 5, 2012, pp. 957-974.
[11] Mokhtarian, P. If telecommunication is such a good substitute for travel, why does congestion continue to get worse? *Transportation Letters,* Vol. 1, No. 1, 2009, pp. 1-17.
[12] Patwary, A. L., T. E. Yu, B. C. English, D. W. Hughes, and S.-H. Cho. Estimating the Rebound Effect of the U.S. Road Freight Transport. *Transportation Research Record,* Vol. 2675, No. 6, 2021, pp. 165-174.
[13] Patwary, A. L., E. Yu, B. English, D. Hughes, and S.-H. Cho. Efficiency Studies of the US Transportation Sector. 2020.
[14] Dong, H., C. Cirillo, and M. Diana. Activity involvement and time spent on computers for leisure: an econometric analysis on the American Time Use Survey dataset. *Transportation,* Vol. 45, No. 2, 2018, pp. 429-449.
[15] *National Household Travel Survey.* Federal Highway Administration, US Department of Transportation, Washington, DC. https://nhts.ornl.gov/. Accessed December 2, 2020.
[16] Zhou, Y., and X. C. Wang. Explore the relationship between online shopping and shopping trips: An analysis with the 2009 NHTS data. *Transportation Research Part A: Policy and Practice,* Vol. 70, 2014, pp. 1-9.

[17] Wilson, R., K. J. Krizek, and S. L. Handy. Trends in out-of-home and at-home activities: evidence from repeat cross-sectional surveys. *Transportation Research Record,* Vol. 2014, No. 1, 2007, pp. 76-84.
[18] Ferrell, C. E. Home-based teleshopping and shopping travel: Where do people find the time? *Transportation Research Record,* Vol. 1926, No. 1, 2005, pp. 212-223.
[19] Corpuz, G., and J. Peachman. Measuring the impacts of internet usage on travel behaviour in the Sydney Household Travel Survey.In *Proceedings of the 26th Australian Transport Research Forum Conference*, New Zealand, 2003.
[20] Loo, B. P. Y., and B. Wang. Factors associated with home-based e-working and e-shopping in Nanjing, China. *Transportation (Dordrecht),* Vol. 45, No. 2, 2017, pp. 365-384.
[21] Weltevreden, J. W., and T. V. Rietbergen. E-shopping versus city centre shopping: The role of perceived city centre attractiveness. *Tijdschrift voor economische en sociale geografie,* Vol. 98, No. 1, 2007, pp. 68-85.
[22] Tonn, B. E., and A. Hemrick. Impacts of the use of e-mail and the Internet on personal trip-making behavior. *Social Science Computer Review,* Vol. 22, No. 2, 2004, pp. 270-280.
[23] Gillespie, A. Substituting Electronic Communications for Physical Travel? The Case of "Teleworking". *Journal of Intelligent Transportation Systems,* Vol. 6, No. 1, 2000, pp. 13-24.
[24] Choo, S., P. L. Mokhtarian, and I. Salomon. Does telecommuting reduce vehicle-miles traveled? An aggregate time series analysis for the US. *Transportation,* Vol. 32, No. 1, 2005, pp. 37-64.
[25] Visser, E. J., and M. Lanzendorf. Mobility and accessibility effects of B2C e-commerce: a literature review. *Tijdschrift voor economische en sociale geografie,* Vol. 95, No. 2, 2004, pp. 189-205.
[26] Kim, S.-N., S. Choo, and P. L. Mokhtarian. Home-based telecommuting and intra-household interactions in work and non-work travel: A seemingly unrelated censored regression approach. *Transportation Research Part A: Policy and Practice,* Vol. 80, 2015, pp. 197-214.
[27] Mokhtarian, P. L. A synthetic approach to estimating the impacts of telecommuting on travel. *Urban studies,* Vol. 35, No. 2, 1998, pp. 215-241.
[28] Rahman, M., K. Murthy Gurumurthy, and K. M. Kockelman. Impact of Flextime on Departure Time Choice for Home-Based Commuting Trips in Austin, Texas. *Transportation Research Record,* Vol. 2676, No. 1, 2022, pp. 446-459.
[29] de Abreu e Silva, J., and P. C. Melo. Home telework, travel behavior, and land-use patterns. *Journal of Transport and Land Use,* Vol. 11, No. 1, 2018, pp. 419-441.
[30] Beck, M. J., and D. A. Hensher. Working from home in Australia in 2020: Positives, negatives and the potential for future benefits to transport and society. *Transportation Research Part A: Policy and Practice,* Vol. 158, 2022, pp. 271-284.
[31] Tahlyan, D., M. Said, H. Mahmassani, A. Stathopoulos, J. Walker, and S. Shaheen. For whom did telework not work during the Pandemic? understanding the factors impacting telework satisfaction in the US using a multiple indicator multiple cause (MIMIC) model. *Transportation Research Part A: Policy and Practice,* Vol. 155, 2022, pp. 387-402.
[32] Wang, X. C., W. Kim, J. Holguín-Veras, and J. Schmid. Adoption of delivery services in light of the COVID pandemic: Who and how long? *Transportation Research Part A: Policy and Practice,* Vol. 154, 2021, pp. 270-286.

[33] Tokey, A. I. Spatial association of mobility and COVID-19 infection rate in the USA: A county-level study using mobile phone location data. *Journal of Transport & Health,* Vol. 22, 2021, p. 101135.
[34] Farag, S., J. Weltevreden, T. Van Rietbergen, M. Dijst, and F. van Oort. E-shopping in the Netherlands: does geography matter? *Environment and Planning B: Planning and Design,* Vol. 33, No. 1, 2006, pp. 59-74.
[35] Cao, X. E-shopping, spatial attributes, and personal travel: a review of empirical studies. *Transportation Research Record,* Vol. 2135, No. 1, 2009, pp. 160-169.
[36] Dias, F. F., P. S. Lavieri, S. Sharda, S. Khoeini, C. R. Bhat, R. M. Pendyala, A. R. Pinjari, G. Ramadurai, and K. K. Srinivasan. A comparison of online and in-person activity engagement: The case of shopping and eating meals. *Transportation Research Part C: Emerging Technologies,* Vol. 114, 2020, pp. 643-656.
[37] *Traffic deaths rose 8% in 2020, even as Americans drove fewer miles during pandemic*. https://www.usatoday.com/story/money/cars/2021/03/05/pandemic-travel-traffic-deaths-up-8-2020-despite-driving-less/4590942001/. Accessed May 2nd, 2021.
[38] Alhassan, H., B. M. Abu, and P. K. Nkegbe. Access to credit, farm productivity and market participation in Ghana: a conditional mixed process approach. *Margin: The Journal of Applied Economic Research,* Vol. 14, No. 2, 2020, pp. 226-246.
[39] Roodman, D. Fitting fully observed recursive mixed-process models with cmp. *The Stata Journal,* Vol. 11, No. 2, 2011, pp. 159-206.
[40] Guess, C. D., L. Boyd, K. Perniciaro, and M. T. Tuason. The Politics of COVID-19: Differences Between Blue, Purple, and Red States in COVID-19 Cases, Deaths, and Regulations. 2022.
[41] Shi, K., R. Shao, J. De Vos, and F. Witlox. Do e-shopping attitudes mediate the effect of the built environment on online shopping frequency of e-shoppers? *International Journal of Sustainable Transportation*, 2021, pp. 1-11.
[42] Vingilis, E., D. Beirness, P. Boase, P. Byrne, J. Johnson, B. Jonah, R. E. Mann, M. J. Rapoport, J. Seeley, and C. M. Wickens. Coronavirus disease 2019: What could be the effects on Road safety? *Accident Analysis & Prevention,* Vol. 144, 2020, p. 105687.
[43] Patwary, A. L., and A. J. Khattak. Interaction Between Information and Communication Technologies and Travel Behavior: Using Behavioral Data to Explore Correlates of the COVID-19 Pandemic. *Transportation Research Record*, 2022.
[44] Brodeur, A., N. Cook, and T. Wright. On the effects of COVID-19 safer-at-home policies on social distancing, car crashes and pollution. *Journal of environmental economics and management,* Vol. 106, 2021, p. 102427.
[45] Zhang, J., Y. Hayashi, and L. D. Frank. COVID-19 and transport: findings from a world-wide expert survey. *Transport Policy,* Vol. 103, 2021, pp. 68-85.
[46] FHWA. *2020 Traffic Volume Trends*. https://www.fhwa.dot.gov/policyinformation/travel_monitoring/20martvt/#:~:text=The%20seasonally%20adjusted%20vehicle%20miles,miles)%20compared%20with%20February%202020. Accessed November 27, 2021.
[47] Walker, A. *This Spring, We All Drove Much Less. Yet Traffic Deaths Went Up. Why?* https://www.curbed.com/2020/10/covid-traffic-deaths-up-speeding.html. Accessed November 1, 2021.

[48] NSC. *Preliminary Semiannual Estimates*. https://injuryfacts.nsc.org/motor-vehicle/overview/preliminary-estimates/. Accessed May 1, 2021.
[49] Palazzo, J. *Historic rise in traffic deaths due to pandemic*. https://www.wkrn.com/news/local-news/historic-rise-in-traffic-deaths-due-to-pandemic/. Accessed November 20, 2021.
[50] Harris, A. *Traffic deaths rise in Tennessee during COVID-19 pandemic*. https://www.wvlt.tv/2020/09/23/traffic-deaths-rise-in-tennessee-during-covid-19-pandemic/. Accessed November 27, 2021.
[51] CCSA. *COVID-19 and Increased Alcohol Consumption: NANOS Poll Summary Report*, Canadian Centre on Substance Use and Addiction. https://www.ccsa.ca/covid-19-and-increased-alcohol-consumption-nanos-poll-summary-report. Accessed July 20, 2020.
[52] Liu, N., F. Zhang, C. Wei, Y. Jia, Z. Shang, L. Sun, L. Wu, Z. Sun, Y. Zhou, and Y. Wang. Prevalence and predictors of PTSS during COVID-19 outbreak in China hardest-hit areas: Gender differences matter. *Psychiatry research,* Vol. 287, 2020, p. 112921.
[53] Wickens, C. M., R. G. Smart, and R. E. Mann. The impact of depression on driver performance. *International Journal of Mental Health and Addiction,* Vol. 12, No. 4, 2014, pp. 524-537.
[54] Blincoe, L., T. R. Miller, E. Zaloshnja, and B. A. Lawrence. The economic and societal impact of motor vehicle crashes, 2010.In, 2015.
[55] Saladié, Ò., E. Bustamante, and A. Gutiérrez. COVID-19 lockdown and reduction of traffic accidents in Tarragona province, Spain. *Transportation Research Interdisciplinary Perspectives,* Vol. 8, 2020, p. 100218.
[56] Hughes, J. E., D. Kaffine, and L. Kaffine. Decline in traffic congestion increased accident severity in the wake of COVID-19. 2020.
[57] Gao, J., J. Wang, Z. Bian, S. D. Bernardes, Y. Chen, A. Bhattacharyya, S. S. M. Thambiran, K. Ozbay, and S. Iyer. The Effects of the COVID-19 Pandemic on Transportation Systems in New York City and Seattle, USA. *arXiv preprint arXiv:2010.01170*, 2020.
[58] Inada, H., L. Ashraf, and S. Campbell. COVID-19 lockdown and fatal motor vehicle collisions due to speed-related traffic violations in Japan: a time-series study. *Injury prevention,* Vol. 27, No. 1, 2021, pp. 98-100.
[59] Khattak, A., A. Mohammadnazar, A. L. Patwary, N. Ahmad, A. Haque, and I. Mahdinia. A Localized Safety Performance Functions Approach Accounting for "Within" Tennessee Variations on Freeways & Interchanges.In, 2022.
[60] Kononov, J., B. Bailey, and B. K. Allery. Relationships between safety and both congestion and number of lanes on urban freeways. *Transportation Research Record,* Vol. 2083, No. 1, 2008, pp. 26-39.
[61] Quddus, M. A., C. Wang, and S. G. Ison. Road traffic congestion and crash severity: econometric analysis using ordered response models. *Journal of Transportation Engineering,* Vol. 136, No. 5, 2010, pp. 424-435.
[62] Sutherland, M., M. McKenney, and A. Elkbuli. Vehicle related injury patterns during the COVID-19 pandemic: What has changed? *The American Journal of Emergency Medicine,* Vol. 38, No. 9, 2020, pp. 1710-1714.
[63] Carter, D. Effects of COVID-19 Shutdown on Crashes and Travel in NC. North Carolina.In, Department of Transportation, Transportation Research Board (TRB) Webinar, 2020.

[64] Lin, L., F. Shi, and W. Li. Assessing Road Traffic Safety Under COVID-19: Inequality, Irregularity, and Severity. *arXiv preprint arXiv:2011.02289*, 2020.
[65] OHS. *Motor Vehicle Fatality Rates up 14% in March, Despite COVID-19.* Occupational Health and Safety. https://ohsonline.com/articles/2020/05/22/motor-vehicle-fatality-rates-up-14-percent-in-march-despite-covid19.aspx. Accessed February 13, 2021.
[66] Noland, R. B., and M. A. Quddus. Congestion and safety: A spatial analysis of London. *Transportation Research Part A: Policy and Practice,* Vol. 39, No. 7-9, 2005, pp. 737-754.
[67] Wang, C., M. Quddus, and S. Ison. A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK. *Transportmetrica A: Transport Science,* Vol. 9, No. 2, 2013, pp. 124-148.
[68] Clark, D. E., and B. M. Cushing. Rural and urban traffic fatalities, vehicle miles, and population density. *Accident Analysis & Prevention,* Vol. 36, No. 6, 2004, pp. 967-972.
[69] Yeo, J., S. Park, and K. Jang. Effects of urban sprawl and vehicle miles traveled on traffic fatalities. *Traffic injury prevention,* Vol. 16, No. 4, 2015, pp. 397-403.
[70] Doucette, M. L., A. Tucker, M. E. Auguste, A. Watkins, C. Green, F. E. Pereira, K. T. Borrup, D. Shapiro, and G. Lapidus. Initial impact of COVID-19's stay-at-home order on motor vehicle traffic and crash patterns in Connecticut: an interrupted time series analysis. *Injury prevention,* Vol. 27, No. 1, 2021, pp. 3-9.
[71] Van Benthem, A. What is the optimal speed limit on freeways? *Journal of Public Economics,* Vol. 124, 2015, pp. 44-62.
[72] Doucette, M. L., A. Tucker, M. E. Auguste, J. D. Gates, D. Shapiro, J. P. Ehsani, and K. T. Borrup. Evaluation of motor vehicle crash rates during and after the COVID-19-associated stay-at-home order in Connecticut. *Accident Analysis & Prevention,* Vol. 162, 2021, p. 106399.
[73] Sebego, M., R. B. Naumann, R. A. Rudd, K. Voetsch, A. M. Dellinger, and C. Ndlovu. The impact of alcohol and road traffic policies on crash rates in Botswana, 2004–2011: a time-series analysis. *Accident Analysis & Prevention,* Vol. 70, 2014, pp. 33-39.
[74] Khattak, A., and F. Targa. Injury severity and total harm in truck-involved work zone crashes. *Transportation Research Record,* Vol. 1877, No. 1, 2004, pp. 106-116.
[75] Shankar, V., F. Mannering, and W. Barfield. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis & Prevention,* Vol. 27, No. 3, 1995, pp. 371-389.
[76] Abdulhafedh, A. Road crash prediction models: Different statistical modeling approaches. *Journal of transportation technologies,* Vol. 7, No. 02, 2017, p. 190.
[77] Abdel-Aty, M. A., and A. E. Radwan. Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention,* Vol. 32, No. 5, 2000, pp. 633-642.
[78] Arvin, R., M. Kamrani, and A. J. Khattak. How instantaneous driving behavior contributes to crashes at intersections: extracting useful information from connected vehicle message data. *Accident Analysis & Prevention,* Vol. 127, 2019, pp. 118-133.
[79] BTS. *Trips by Distance.* US Bureau of Transportation Statistics. https://data.bts.gov/Research-and-Statistics/Trips-by-Distance/w96p-f2qv. Accessed October 27, 2021.
[80] TN.GOV. *COVID-19 Timeline*. https://www.tn.gov/governor/covid-19/covid19timeline.html. Accessed September 2, 2022, 2022.

[81] Harmon, T., G. B. Bahar, and F. B. Gross. Crash costs for highway safety analysis.In, United States. Federal Highway Administration. Office of Safety, 2018.
[82] Adegbite, Q., K. Billah, H. Sharif, and S. Dessouky. Urban Intersections and Traffic Safety in the City of San Antonio.In *MATEC Web of Conferences, No. 271*, EDP Sciences, 2019. p. 06003.
[83] Greene, W. H. *Econometric analysis*. Pearson Education India, 2003.
[84] Woo, H., B. N. Eskelson, and V. J. Monleon. Tree height increment models for national forest inventory data in the Pacific Northwest, USA. *Forests,* Vol. 11, No. 1, 2019, p. 2.
[85] Atkinson, A. C. Plots, transformations and regression; an introduction to graphical methods of diagnostic regression analysis.In, Clarendon Press: Oxford, UK, 1985.
[86] Haque, A. M., I. Mahdinia, A. L. Patwary, and A. J. Khattak. Are Damages to Remainder Parcels in Right-of-Way Acquisitions Stationary? A Spatial Analysis of Appraisal Report Data. *Transportation Research Record*, 2022, p. 03611981221105073.
[87] Wali, B., A. J. Khattak, J. Waters, D. Chimba, and X. Li. Development of safety performance functions: incorporating unobserved heterogeneity and functional form analysis. *Transportation Research Record,* Vol. 2672, No. 30, 2018, pp. 9-20.
[88] Mohammadnazar, A., I. Mahdinia, N. Ahmad, A. J. Khattak, and J. Liu. Understanding how relationships between crash frequency and correlates vary for multilane rural highways: Estimating geographically and temporally weighted regression models. *Accident Analysis & Prevention,* Vol. 157, 2021, p. 106146.
[89] Baruya, A. Speed-accident relationships on European roads.In *9th International Conference on Road Safety in Europe*, 1998. pp. 1-19.
[90] Lave, C. A. Speeding, coordination, and the 55 mph limit. *The American Economic Review,* Vol. 75, No. 5, 1985, pp. 1159-1164.
[91] Imprialou, M.-I. M., M. Quddus, D. E. Pitfield, and D. Lord. Re-visiting crash–speed relationships: A new perspective in crash modelling. *Accident Analysis & Prevention,* Vol. 86, 2016, pp. 173-185.
[92] Tefft, B. C., L. Villavicencio, A. Benson, L. Arnold, W. Kim, V. Añorve, and W. J. Horrey. Self-Reported Risky Driving in Relation to Changes in Amount of Driving During the COVID-19 Pandemic. *AAA Foundation for Traffic Safety*, 2022.
[93] Katrakazas, C., E. Michelaraki, M. Sekadakis, A. Ziakopoulos, A. Kontaxi, and G. Yannis. Identifying the impact of the COVID-19 pandemic on driving behavior using naturalistic driving data and time series forecasting. *Journal of safety research,* Vol. 78, 2021, pp. 189-202.
[94] Patwary, A. L., and A. Khattak. A Study of Truck-Involved Crashes and Fatalities Before and During the COVID-19 Pandemic: Evidence of Traffic Safety from Tennessee.In *101st Annual Meeting of Transportation Research Board (TRB)*, 2022.
[95] Zhang, J., B. Feng, Y. Wu, P. Xu, R. Ke, and N. Dong. The effect of human mobility and control measures on traffic safety during COVID-19 pandemic. *PLoS one,* Vol. 16, No. 3, 2021, p. e0243263.
[96] Oei, H.-L., and P. Polak. Intelligent speed adaptation (ISA) and road safety. *IATSS research,* Vol. 26, No. 2, 2002, pp. 45-51.
[97] Warner, H. W., and L. Åberg. The long-term effects of an ISA speed-warning device on drivers' speeding behaviour. *Transportation research part F: traffic psychology and behaviour,* Vol. 11, No. 2, 2008, pp. 96-107.

[98] Lian, Y., G. Zhang, J. Lee, and H. Huang. Review on big data applications in safety research of intelligent transportation systems and connected/automated vehicles. *Accident Analysis & Prevention,* Vol. 146, 2020, p. 105711.
[99] Lee, J., F. Baig, and X. Li. Media influence, trust, and the public adoption of automated vehicles. *IEEE Intelligent Transportation Systems Magazine*, 2021.
[100] Haque, A. M., R. Arvin, and A. Khattak. Harnessing Basic Safety Messages to Identify High-Risk Hotspots for Transit Buses.In, 2021.
[101] Corsaro, N., D. W. Gerard, R. S. Engel, and J. E. Eck. Not by accident: An analytical approach to traffic crash harm reduction. *Journal of Criminal Justice,* Vol. 40, No. 6, 2012, pp. 502-514.
[102] USDOT. National Roadway Safety Strategy.In, Washington, DC, 2022.
[103] White-House. *Justice40 Initiative - Environmental Justice - The White House*. https://www.whitehouse.gov/environmentaljustice/justice40/. Accessed December 30, 2022.
[104] USDOT. *Transportation Disadvantaged Census Tracts (Historically Disadvantaged Communities)*. https://usdot.maps.arcgis.com/apps/dashboards/d6f90dfcc8b44525b04c7ce748a3674a. Accessed 5.24.2022.
[105] Liu, J., and A. J. Khattak. Delivering improved alerts, warnings, and control assistance using basic safety messages transmitted between connected vehicles. *Transportation research part C: emerging technologies,* Vol. 68, 2016, pp. 83-100.
[106] Khattak, A. J., and B. Wali. Analysis of volatility in driving regimes extracted from basic safety messages transmitted between connected vehicles. *Transportation research part C: emerging technologies,* Vol. 84, 2017, pp. 48-73.
[107] Wali, B., A. J. Khattak, H. Bozdogan, and M. Kamrani. How is driving volatility related to intersection safety? A Bayesian heterogeneity-based analysis of instrumented vehicles data. *Transportation Research Part C: Emerging Technologies,* Vol. 92, 2018, pp. 504-524.
[108] Kamrani, M., R. Arvin, and A. J. Khattak. Extracting useful information from Basic Safety Message Data: an empirical study of driving volatility measures and crash frequency at intersections. *Transportation Research Record*, 2018, p. 0361198118773869.
[109] Patwary, A. L. Bus safety during COVID-19: analyzing bus-involved crashes and economic harm. *International Journal of Crashworthiness*, 2023, pp. 1-8.
[110] Kadilar, G. O. Effect of driver, roadway, collision, and vehicle characteristics on crash severity: a conditional logistic regression approach. *International journal of injury control and safety promotion,* Vol. 23, No. 2, 2016, pp. 135-144.
[111] Ahmed, M. M., R. Franke, K. Ksaibati, and D. S. Shinstine. Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models. *Accident Analysis & Prevention,* Vol. 117, 2018, pp. 106-113.
[112] Mohammadnazar, A., A. L. Patwary, N. Moradloo, R. Arvin, and A. J. Khattak. Incorporating driving volatility measures in safety performance functions: Improving safety at signalized intersections. *Accident Analysis & Prevention,* Vol. 178, 2022, p. 106872.
[113] Saha, S., P. Schramm, A. Nolan, and J. Hess. Adverse weather conditions and fatal motor vehicle crashes in the United States, 1994-2012. *Environmental health,* Vol. 15, No. 1, 2016, pp. 1-9.

[114] Eisenberg, D., and K. E. Warner. Effects of snowfalls on motor vehicle collisions, injuries, and fatalities. *American journal of public health,* Vol. 95, No. 1, 2005, pp. 120-124.
[115] Pulugurtha, S. S., V. R. Duddu, and Y. Kotagiri. Traffic analysis zone level crash estimation models based on land use characteristics. *Accident Analysis & Prevention,* Vol. 50, 2013, pp. 678-687.
[116] Schneider, R. J. United States Pedestrian Fatality Trends, 1977 to 2016. *Transportation Research Record,* Vol. 2674, No. 9, 2020, pp. 1069-1083.
[117] Baireddy, R., H. Zhou, and M. Jalayer. Multiple Correspondence Analysis of Pedestrian Crashes in Rural Illinois. *Transportation Research Record,* Vol. 2672, No. 38, 2018, pp. 116-127.
[118] Loeb, P. D., and W. A. Clarke. The cell phone effect on pedestrian fatalities. *Transportation Research Part E: Logistics and Transportation Review,* Vol. 45, No. 1, 2009, pp. 284-290.
[119] Rudisill, T. M., L. O. Barbee, and B. Hendricks. Characteristics of Fatal, Pedestrian-Involved, Motor Vehicle Crashes in West Virginia: A Cross-Sectional and Spatial Analysis. *International journal of environmental research and public health,* Vol. 20, No. 7, 2023, p. 5251.
[120] Patwary, A. L., and A. J. Khattak. Crash harm before and during the COVID-19 pandemic: Evidence for spatial heterogeneity in Tennessee. *Accident Analysis & Prevention,* Vol. 183, 2023, p. 106988.
[121] Pour, M. H., J. Prasetijo, A. S. Yahaya, and S. M. R. Ghadiri. Modeling Vehicle-pedestrian Crashes With Excess Zero Along Malaysia Federal Roads. *Procedia - Social and Behavioral Sciences,* Vol. 53, 2012, pp. 1216-1225.
[122] Pei, X., S. C. Wong, and N. N. Sze. The roles of exposure and speed in road safety analysis. *Accident Analysis & Prevention,* Vol. 48, 2012, pp. 464-471.
[123] Stamatiadis, N., and G. Puccini. Fatal crash rates in the southeastern United States: Why are they higher? *Transportation research record,* Vol. 1665, No. 1, 1999, pp. 118-124.
[124] Lee, J., M. Abdel-Aty, and K. Choi. Analysis of residence characteristics of at-fault drivers in traffic crashes. *Safety science,* Vol. 68, 2014, pp. 6-13.
[125] Nantulya, V. M., and M. R. Reich. Equity dimensions of road traffic injuries in low-and middle-income countries. *Injury control and safety promotion,* Vol. 10, No. 1-2, 2003, pp. 13-20.
[126] Traynor, T. L. The relationship between regional economic conditions and the severity of traffic crashes. *Traffic injury prevention,* Vol. 10, No. 4, 2009, pp. 368-374.
[127] Chimba, D., A. Musinguzi, and E. Kidando. Associating pedestrian crashes with demographic and socioeconomic factors. *Case studies on transport policy,* Vol. 6, No. 1, 2018, pp. 11-16.
[128] Christie, N., H. Ward, and R. Kimberlee. Road traffic injury risk in disadvantaged communities: evaluation of the neighbourhood road safety initiative. 2010.
[129] Azetsop, J. Social justice approach to road safety in Kenya: addressing the uneven distribution of road traffic injuries and deaths across population groups. *Public health ethics,* Vol. 3, No. 2, 2010, pp. 115-127.
[130] Campos-Outcalt, D., C. Bay, A. Dellapena, and M. K. Cota. Motor vehicle crash fatalities by race/ethnicity in Arizona, 1990–96. *Injury Prevention,* Vol. 9, No. 3, 2003, pp. 251-256.

[131] Sanders, R. L., and R. J. Schneider. An exploration of pedestrian fatalities by race in the United States. *Transportation Research Part D: Transport and Environment,* Vol. 107, 2022, p. 103298.
[132] Li, X., S. Yu, X. Huang, B. Dadashova, W. Cui, and Z. Zhang. Do underserved and socially vulnerable communities observe more crashes? A spatial examination of social vulnerability and crash risks in Texas. *Accident Analysis & Prevention,* Vol. 173, 2022, p. 106721.
[133] Roudsari, B., S. Ramisetty-Mikler, and L. A. Rodriguez. Ethnicity, age, and trends in alcohol-related driver fatalities in the United States. *Traffic injury prevention,* Vol. 10, No. 5, 2009, pp. 410-414.
[134] Naimi, T. S., Z. Xuan, V. Sarda, S. E. Hadland, M. C. Lira, M. H. Swahn, R. B. Voas, and T. C. Heeren. Association of state alcohol policies with alcohol-related motor vehicle crash fatalities among US adults. *JAMA internal medicine,* Vol. 178, No. 7, 2018, pp. 894-901.
[135] Noland, R. B., N. J. Klein, and N. K. Tulach. Do lower income areas have more pedestrian casualties? *Accident Analysis & Prevention,* Vol. 59, 2013, pp. 337-345.
[136] Jacobsen, P. L., and H. Rutter. *Cycling safety*. MIT Press Cambridge, MA, 2012.
[137] Giles-Corti, B., and R. J. Donovan. Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment. *Preventive medicine,* Vol. 35, No. 6, 2002, pp. 601-611.
[138] Gibbs, K., S. J. Slater, L. Nicholson, D. Barker, and F. J. Chaloupka. Income disparities in street features that encourage walking. *Bridging the Gap Program*, 2012.
[139] Buehler, R., and J. Pucher. Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. *Transportation,* Vol. 39, 2012, pp. 409-432.
[140] Dill, J., and T. Carr. Bicycle commuting and facilities in major US cities: if you build them, commuters will use them. *Transportation Research Record,* Vol. 1828, No. 1, 2003, pp. 116-123.
[141] Barajas, J. M. Not all crashes are created equal. *Journal of Transport and Land Use,* Vol. 11, No. 1, 2018, pp. 865-882.
[142] Cradock, A. L., P. J. Troped, B. Fields, S. J. Melly, S. V. Simms, F. Gimmler, and M. Fowler. Factors associated with federal transportation funding for local pedestrian and bicycle programming and facilities. *Journal of Public Health Policy,* Vol. 30, 2009, pp. S38-S72.
[143] Yu, C.-Y., X. Zhu, and C. Lee. Income and racial disparity and the role of the built environment in pedestrian injuries. *Journal of Planning Education and Research,* Vol. 42, No. 2, 2022, pp. 136-149.
[144] Shin, E. J. Decomposing neighborhood disparities in bicycle crashes: A Gelbach decomposition analysis. *Transport Policy,* Vol. 131, 2023, pp. 156-172.
[145] Chen, C., H. Lin, and B. P. Loo. Exploring the impacts of safety culture on immigrants' vulnerability in non-motorized crashes: a cross-sectional study. *Journal of urban health,* Vol. 89, 2012, pp. 138-152.
[146] Nguyen-Hoang, P., and R. Yeung. Dollars for lives: The effect of highway capital investments on traffic fatalities. *Journal of safety research,* Vol. 51, 2014, pp. 109-115.
[147] Rogers, M. M., and W. L. Weber. Evaluating CO2 emissions and fatalities tradeoffs in truck transport. *International Journal of Physical Distribution & Logistics Management*, 2011.

[148] Lord, D., S. P. Washington, and J. N. Ivan. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention,* Vol. 37, No. 1, 2005, pp. 35-46.
[149] Lord, D., S. Washington, and J. N. Ivan. Further notes on the application of zero-inflated models in highway safety. *Accident Analysis & Prevention,* Vol. 39, No. 1, 2007, pp. 53-57.
[150] Son, H. D., Y.-J. Kweon, and B. B. Park. Development of crash prediction models with individual vehicular data. *Transportation Research Part C: Emerging Technologies,* Vol. 19, No. 6, 2011, pp. 1353-1363.
[151] Cai, Q., J. Lee, N. Eluru, and M. Abdel-Aty. Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. *Accident Analysis & Prevention,* Vol. 93, 2016, pp. 14-22.
[152] Hosseinpour, M., J. Prasetijo, A. S. Yahaya, and S. M. R. Ghadiri. A comparative study of count models: Application to pedestrian-vehicle crashes along Malaysia federal roads. *Traffic injury prevention,* Vol. 14, No. 6, 2013, pp. 630-638.
[153] Hosseinpour, M., A. S. Yahaya, and A. F. Sadullah. Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian Federal Roads. *Accident Analysis & Prevention,* Vol. 62, 2014, pp. 209-222.
[154] BRFSS. *PLACES: Census Tract Data (GIS Friendly Format), 2022 release*. https://chronicdata.cdc.gov/500-Cities-Places/PLACES-Census-Tract-Data-GIS-Friendly-Format-2022-/yjkw-uj5s. Accessed June 25, 2023.
[155] Finlay, J., M. Robert, E. Michael, K. Anam, L. Mao, G.-L. Iris, C. Philippa, and C. Megan. *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]. https://doi.org/10.3886/E160262V1. Accessed July 13, 2023.
[156] FHWA. *Highway Statistics Series*. https://www.fhwa.dot.gov/policyinformation/statistics/2019/hm60.cfm. Accessed June 25, 2023.
[157] NIAAA. *Alcohol Consumption by State*. https://wisevoter.com/state-rankings/alcohol-consumption-by-state/. Accessed June 25, 2023.
[158] IIHS. *Cellphone use laws*. https://www.iihs.org/topics/distracted-driving/cellphone-use-laws. Accessed June 25, 2023.
[159] NHTSA. *Seat Belt Use in 2021 - Use Rates in the States and Territories*. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813307. Accessed June 25, 2023.
[160] USDOT. *Transportation Disadvantaged Census Tracts (Historically Disadvantaged Communities) Interim Definition Methodology*. https://www.transportation.gov/priorities/equity/justice40/transportation-disadvantaged-census-tracts-historically-disadvantaged. Accessed January 25, 2023.
[161] Šinkovec, H., G. Heinze, R. Blagus, and A. Geroldinger. To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets. *BMC Medical Research Methodology,* Vol. 21, 2021, pp. 1-15.
[162] Krapp, A. Transportation Equity Project Prioritization Criteria. *MUP Capstone*, 2020.
[163] Voas, R. B., A. S. Tippetts, and J. Fell. The relationship of alcohol safety laws to drinking drivers in fatal crashes. *Accident Analysis & Prevention,* Vol. 32, No. 4, 2000, pp. 483-492.
[164] Lim, S. H., and J. Chi. Cellphone bans and fatal motor vehicle crash rates in the United States. *Journal of Public Health Policy,* Vol. 34, No. 2, 2013, pp. 197-212.

[165] Pollack, K. M., S. Frattaroli, J. L. Young, G. Dana-Sacco, and A. C. Gielen. Motor vehicle deaths among American Indian and Alaska Native populations. *Epidemiologic reviews,* Vol. 34, No. 1, 2012, pp. 73-88.
[166] Murphy, T., P. Pokhrel, A. Worthington, H. Billie, M. Sewell, and N. Bill. Unintentional injury mortality among American Indians and Alaska natives in the United States, 1990–2009. *American journal of public health,* Vol. 104, No. S3, 2014, pp. S470-S480.
[167] Naimi, T. S., N. Cobb, D. Boyd, D. W. Jarman, R. Brewer, D. E. Nelson, J. Holt, D. Espey, P. Snesrud, and P. Chavez. 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,* Vol. 57, No. 34, 2008, pp. 938-941.
[168] Dharmaratne, S. D., A. U. Jayatilleke, and A. C. Jayatilleke. Road traffic crashes, injury and fatality trends in Sri Lanka: 1938-2013. *Bulletin of the World Health Organization,* Vol. 93, 2015, pp. 640-647.
[169] Munteanu, P. L., M. Rosu, V. Panaitescu, and A. Pungă. Human and environmental factors contributing to fatal road accidents in a Romanian population. *Rom J Leg Med,* Vol. 22, No. 2, 2014, pp. 97-100.
[170] Stamatiadis, N., and G. Puccini. Socioeconomic descriptors of fatal crash rates in the Southeast USA. *Injury control and safety promotion,* Vol. 7, No. 3, 2000, pp. 165-173.
[171] Statista. *Smartphone penetration rate as share of the population in the United States from 2010 to 2021*. https://www.statista.com/statistics/201183/forecast-of-smartphone-penetration-in-the-us/. Accessed July 20, 2022.
[172] Huang, Z., B. P. Y. Loo, and K. W. Axhausen. Travel behaviour changes under Work-from-home (WFH) arrangements during COVID-19. *Travel behaviour and society,* Vol. 30, 2023, pp. 202-211.
[173] Hensher, D. A., E. Wei, and M. J. Beck. The impact of COVID-19 and working from home on the workspace retained at the main location office space and the future use of satellite offices. *Transport Policy,* Vol. 130, 2023, pp. 184-195.
[174] Hensher, D. A., M. J. Beck, and C. Balbontin. Working from home 22 months on from the beginning of COVID-19: What have we learned for the future provision of transport services? *Research in Transportation Economics*, 2023, p. 101271.
[175] Meister, A., C. Winkler, B. Schmid, and K. Axhausen. In-store or online grocery shopping before and during the COVID-19 pandemic. *Travel behaviour and society,* Vol. 30, 2023, pp. 291-301.
[176] *Monthly Retail Trade*, US Census Bureau. https://www.census.gov/retail/index.html. Accessed July 20, 2022.
[177] Mokhtarian, P. L. Telecommunications and travel: The case for complementarity. *Journal of industrial ecology,* Vol. 6, No. 2, 2002, pp. 43-57.
[178] Kim, W., and X. C. Wang. To be online or in-store: Analysis of retail, grocery, and food shopping in New York city. *Transportation Research Part C: Emerging Technologies,* Vol. 126, 2021, p. 103052.
[179] MorganStanley. *Can Food Delivery Apps Deliver Profits for Investors?* https://www.morganstanley.com/ideas/food-delivery-app-profits. Accessed July 22, 2022.
[180] Justice40. *Justice40: A WHOLE-OF-GOVERNMENT INITIATIVE*, The White House. https://www.whitehouse.gov/environmentaljustice/justice40/. Accessed July 10, 2022.

[181] Wang, D., and F. Y. T. Law. Impacts of Information and Communication Technologies (ICT) on time use and travel behavior: a structural equations analysis. *Transportation,* Vol. 34, No. 4, 2007, pp. 513-527.
[182] Loo, B. P. Y., and B. Wang. Factors associated with home-based e-working and e-shopping in Nanjing, China. *Transportation,* Vol. 45, No. 2, 2018, pp. 365-384.
[183] Sim, L. L., and S. M. Koi. Singapore's Internet shoppers and their impact on traditional shopping patterns. *Journal of retailing and consumer services,* Vol. 9, No. 2, 2002, pp. 115-124.
[184] Farag, S., T. Schwanen, M. Dijst, and J. Faber. Shopping online and/or in-store? A structural equation model of the relationships between e-shopping and in-store shopping. *Transportation research part A: policy practice,* Vol. 41, No. 2, 2007, pp. 125-141.
[185] Spurlock, C. A., A. Todd-Blick, G. Wong-Parodi, and V. Walker. Children, income, and the impact of home delivery on household shopping trips. *Transportation Research Record,* Vol. 2674, No. 10, 2020, pp. 335-350.
[186] Bezirgani, A., and U. Lachapelle. Online grocery shopping for the elderly in Quebec, Canada: The role of mobility impediments and past online shopping experience. *Travel behaviour and society,* Vol. 25, 2021, pp. 133-143.
[187] ---. Qualitative Study on Factors Influencing Aging Population's Online Grocery Shopping and Mode Choice When Grocery Shopping in Person. *Transportation Research Record,* Vol. 2675, No. 1, 2021, pp. 79-92.
[188] Morganosky, M. A., and B. J. Cude. Consumer response to online grocery shopping. *International Journal of Retail & Distribution Management*, 2000.
[189] Zheng, Q., J. Chen, R. Zhang, and H. H. Wang. What factors affect Chinese consumers' online grocery shopping? Product attributes, e-vendor characteristics and consumer perceptions. *China Agricultural Economic Review,* Vol. 12, No. 2, 2020, pp. 193-213.
[190] Suel, E., S. Le Vine, and J. Polak. Empirical application of expenditure diary instrument to quantify relationships between in-store and online grocery shopping: case study of Greater London. *Transportation Research Record,* Vol. 2496, No. 1, 2015, pp. 45-54.
[191] Berg, J., and M. Henriksson. In search of the 'good life': Understanding online grocery shopping and everyday mobility as social practices. *Journal of Transport Geography,* Vol. 83, 2020, p. 102633.
[192] Young, M., and U. Lachapelle. Transportation behaviours of the growing Canadian single-person households. *Transport Policy,* Vol. 57, 2017, pp. 41-50.
[193] Bhat, C. R. A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B: Methodological,* Vol. 79, 2015, pp. 50-77.
[194] PHTS. *Household Travel Survey.* Puget Sound Regional Council. https://household-travel-survey-psregcncl.hub.arcgis.com/search?collection=Dataset.
[195] USDOT. *Justice40 Initiative.* US Department of Transportation. https://www.transportation.gov/equity-Justice40.
[196] Hatef, E., X. Ma, Y. Shaikh, H. Kharrazi, J. P. Weiner, and D. J. Gaskin. Internet access, social risk factors, and web-based social support seeking behavior: assessing correlates of the "digital divide" across neighborhoods in the State of Maryland. *Journal of medical systems,* Vol. 45, 2021, pp. 1-14.

[197] Araque, J. C., R. P. Maiden, N. Bravo, I. Estrada, R. Evans, K. Hubchik, K. Kirby, and M. Reddy. Computer usage and access in low-income urban communities. *Computers in Human Behavior,* Vol. 29, No. 4, 2013, pp. 1393-1401.
[198] Pick, J., A. Sarkar, and E. Parrish. Internet use and online activities in US States: geographic disparities and socio-economic influences. 2018.
[199] Rudolph, T., B. Rosenbloom, and T. Wagner. Barriers to online shopping in Switzerland. *Journal of International Consumer Marketing,* Vol. 16, No. 3, 2004, pp. 55-74.
[200] Yazdanifard, R., T. Agodi, and S. Alizadeh. How unreliable delivery system affects e-marketing effectiveness.In *Proceedings of 2011 International Conference on Information Communication and Management*, 2011. pp. 10-14.
[201] Ha, N., and T. Nguyen. The effect of trust on consumers' online purchase intention: An integration of TAM and TPB. *Management Science Letters,* Vol. 9, No. 9, 2019, pp. 1451-1460.
[202] Hilmers, A., D. C. Hilmers, and J. Dave. Neighborhood disparities in access to healthy foods and their effects on environmental justice. *American journal of public health,* Vol. 102, No. 9, 2012, pp. 1644-1654.
[203] Bhat, C. R. Analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B: Methodological,* Vol. 32, No. 6, 1998, pp. 361-371.
[204] Galante, N., E. G. López, and S. Monroe. The future of online grocery in Europe. *McKinsey & Company*, 2013, pp. 22-31.
[205] Neren, J. *The Impact of E-Commerce on the Retail Industry.* Forbes. https://www.forbes.com/sites/forbestechcouncil/2021/02/19/the-future-of-e-commerce-grocery-has-arrived-2021-industry-outlook/?sh=251c5e3089b9. Accessed February 3, 2023.
[206] Jensen, K. L., J. Yenerall, X. Chen, and T. E. Yu. US consumers' online shopping behaviors and intentions during and after the COVID-19 pandemic. *Journal of Agricultural and Applied Economics,* Vol. 53, No. 3, 2021, pp. 416-434.
[207] Lavieri, P. S., Q. Dai, and C. R. Bhat. Using virtual accessibility and physical accessibility as joint predictors of activity-travel behavior. *Transportation Research Part A: Policy and Practice,* Vol. 118, 2018, pp. 527-544.
[208] Ali, N. I., S. Samsuri, M. Sadry, I. A. Brohi, and A. Shah. Online shopping satisfaction in Malaysia: A framework for security, trust and cybercrime.In *2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*, IEEE, 2016. pp. 194-198.
[209] EIA. *U.S. energy facts explained*. https://www.eia.gov/energyexplained/us-energy-facts/. Accessed July 18, 2020.
[210] EPA. *Sources of Greenhouse Gas Emissions*. https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions. Accessed July 18, 2022.
[211] Globenewswire. *Alternative Fuel Vehicle Market Size Worth US$ 1,331.5 Bn by 2030*, Globe Newswire. https://www.globenewswire.com/en/news-release/2021/12/15/2353153/0/en/Alternative-Fuel-Vehicle-Market-Size-Worth-US-1-331-5-Bn-by-2030.html. Accessed July 18, 2022.
[212] Jia, J. Analysis of alternative fuel vehicle (AFV) adoption utilizing different machine learning methods: a case study of 2017 NHTS. *IEEE Access,* Vol. 7, 2019, pp. 112726-112735.

[213] Li, X., C. Liu, and J. Jia. Ownership and usage analysis of alternative fuel vehicles in the United States with the 2017 national household travel survey data. *Sustainability,* Vol. 11, No. 8, 2019, p. 2262.
[214] Jia, W., Z. Jiang, T. D. Chen, and R. Paleti. Evaluating fuel tax revenue impacts of electric vehicle adoption in virginia counties: Application of a bivariate linear mixed count model. *Transportation Research Record,* Vol. 2673, No. 9, 2019, pp. 548-561.
[215] Canepa, K., S. Hardman, and G. Tal. An early look at plug-in electric vehicle adoption in disadvantaged communities in California. *Transport Policy,* Vol. 78, 2019, pp. 19-30.
[216] Hardman, S., K. Fleming, E. Khare, and M. M. Ramadan. A perspective on equity in the transition to electric vehicle. *MIT Science Policy Review,* Vol. 2, 2021, pp. 46-54.
[217] Bae, Y., S. K. Mitra, C. R. Rindt, and S. G. Ritchie. Factors influencing alternative fuel adoption decisions in heavy-duty vehicle fleets. *Transportation Research Part D: Transport and Environment,* Vol. 102, 2022, p. 103150.
[218] Rogers, E. M., A. Singhal, and M. M. Quinlan. Diffusion of innovations.In *An integrated approach to communication theory and research*, Routledge, 2014. pp. 432-448.
[219] Gallagher, K. S., and E. Muehlegger. Giving green to get green? Incentives and consumer adoption of hybrid vehicle technology. *Journal of environmental economics and management,* Vol. 61, No. 1, 2011, pp. 1-15.
[220] Tal, G., and M. A. Nicholas. Studying the PEV market in California: Comparing the PEV, PHEV and hybrid markets.In *2013 World Electric Vehicle Symposium and Exhibition (EVS27)*, IEEE, 2013. pp. 1-10.
[221] Jansson, J., A. Marell, and A. Nordlund. Green consumer behavior: determinants of curtailment and eco-innovation adoption. *Journal of consumer marketing*, 2010.
[222] Zhu, X., and C. Liu. Investigating the neighborhood effect on hybrid vehicle adoption. *Transportation Research Record,* Vol. 2385, No. 1, 2013, pp. 37-44.
[223] Egbue, O., and S. Long. Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. *Energy Policy,* Vol. 48, 2012, pp. 717-729.
[224] Sierzchula, W., S. Bakker, K. Maat, and B. Van Wee. The influence of financial incentives and other socio-economic factors on electric vehicle adoption. *Energy Policy,* Vol. 68, 2014, pp. 183-194.
[225] Jansson, J., T. Pettersson, A. Mannberg, R. Brännlund, and U. Lindgren. Adoption of alternative fuel vehicles: Influence from neighbors, family and coworkers. *Transportation Research Part D: Transport and Environment,* Vol. 54, 2017, pp. 61-73.
[226] Klöckner, C. A., A. Nayum, and M. Mehmetoglu. Positive and negative spillover effects from electric car purchase to car use. *Transportation Research Part D: Transport and Environment,* Vol. 21, 2013, pp. 32-38.
[227] Hidrue, M. K., G. R. Parsons, W. Kempton, and M. P. Gardner. Willingness to pay for electric vehicles and their attributes. *Resource and energy economics,* Vol. 33, No. 3, 2011, pp. 686-705.
[228] Javid, R. J., and A. Nejat. A comprehensive model of regional electric vehicle adoption and penetration. *Transport Policy,* Vol. 54, 2017, pp. 30-42.
[229] Nesbitt, K., and J. Davies. From the top of the organization to the bottom line: Understanding the fleet market for plug-in electric vehicles.In *2013 World Electric Vehicle Symposium and Exhibition (EVS27)*, IEEE, 2013. pp. 1-14.

[230] Jeong, I.-J. An optimal approach for a set covering version of the refueling-station location problem and its application to a diffusion model. *International Journal of Sustainable Transportation,* Vol. 11, No. 2, 2017, pp. 86-97.
[231] White, L. V., A. L. Carrel, W. Shi, and N. D. Sintov. Why are charging stations associated with electric vehicle adoption? Untangling effects in three United States metropolitan areas. *Energy Research & Social Science,* Vol. 89, 2022, p. 102663.
[232] Axsen, J., D. C. Mountain, and M. Jaccard. Combining stated and revealed choice research to simulate the neighbor effect: The case of hybrid-electric vehicles. *Resource and energy economics,* Vol. 31, No. 3, 2009, pp. 221-238.
[233] Mau, P., J. Eyzaguirre, M. Jaccard, C. Collins-Dodd, and K. Tiedemann. The 'neighbor effect': Simulating dynamics in consumer preferences for new vehicle technologies. *Ecological Economics,* Vol. 68, No. 1-2, 2008, pp. 504-516.
[234] Axsen, J., and K. S. Kurani. Interpersonal influence in the early plug-in hybrid market: Observing social interactions with an exploratory multi-method approach. *Transportation Research Part D: Transport and Environment,* Vol. 16, No. 2, 2011, pp. 150-159.
[235] Carlton, G. J., and S. Sultana. Electric vehicle charging station accessibility and land use clustering: A case study of the Chicago region. *Journal of Urban Mobility,* Vol. 2, 2022, p. 100019.
[236] Hathaway, Z., H. Polis, J. Loomis, J. Boroski, A. Milano, and J. Ouyang. A Utility Roadmap for Expanding Customer Adoption of Electric Vehicles. *World Electric Vehicle Journal,* Vol. 12, No. 2, 2021, p. 81.
[237] AFDC. *Alternative Fuels Data Center: Electric Vehicle Charging Station Locations*. https://afdc.energy.gov/fuels/electricity_locations.html#/find/nearest. Accessed September 30, 2022.
[238] ACS. *Code Lists, Definitions, and Accuracy.* US Census Bureau. https://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/2020_ACSSubjectDefinitions.pdf. Accessed July 22, 2022.
[239] Lee, D., J. Mulrow, C. J. Haboucha, S. Derrible, and Y. Shiftan. Attitudes on autonomous vehicle adoption using interpretable gradient boosting machine. *Transportation Research Record,* Vol. 2673, No. 11, 2019, pp. 865-878.
[240] Nvidia. *What is XGBOOST?*, Data Science. https://www.nvidia.com/en-us/glossary/data-science/xgboost/#:~:text=XGBoost%20is%20a%20scalable%20and,model%20performance%20and%20computational%20speed. Accessed July 15, 2022.
[241] Chang, I., H. Park, E. Hong, J. Lee, and N. Kwon. Predicting effects of built environment on fatal pedestrian accidents at location-specific level: Application of XGBoost and SHAP. *Accident Analysis & Prevention,* Vol. 166, 2022, p. 106545.
[242] Meng, H., X. Wang, and X. Wang. Expressway crash prediction based on traffic big data.In *Proceedings of the 2018 International Conference on Signal Processing and Machine Learning*, 2018. pp. 11-16.
[243] Lucas, A., R. Barranco, and N. Refa. Ev idle time estimation on charging infrastructure, comparing supervised machine learning regressions. *Energies,* Vol. 12, No. 2, 2019, p. 269.
[244] Ullah, I., K. Liu, T. Yamamoto, M. Zahid, and A. Jamal. Prediction of electric vehicle charging duration time using ensemble machine learning algorithm and Shapley additive

explanations. *International Journal of Energy Research,* Vol. 46, No. 11, 2022, pp. 15211-15230.
[245] Lundberg, S. M., and S.-I. Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems,* Vol. 30, 2017.
[246] Liu, J., A. J. Khattak, X. Li, and X. Fu. A spatial analysis of the ownership of alternative fuel and hybrid vehicles. *Transportation Research Part D: Transport and Environment,* Vol. 77, 2019, pp. 106-119.
[247] Li, J., J. L. Walker, S. Srinivasan, and W. P. Anderson. Modeling private car ownership in China: Investigation of urban form impact across megacities. *Transportation Research Record,* Vol. 2193, No. 1, 2010, pp. 76-84.
[248] Ma, L., and S. Srinivasan. Impact of individuals' immigrant status on household auto ownership. *Transportation Research Record,* Vol. 2156, No. 1, 2010, pp. 36-46.
[249] Whelan, G. Modelling car ownership in Great Britain. *Transportation Research Part A: Policy and Practice,* Vol. 41, No. 3, 2007, pp. 205-219.
[250] Misra, A., and K. Watkins. Modeling cyclist route choice using revealed preference data: an age and gender perspective. *Transportation Research Record,* Vol. 2672, No. 3, 2018, pp. 145-154.
[251] Shah, N. R., and C. R. Cherry. Different Safety Awareness and Route Choice between Frequent and Infrequent Bicyclists: Findings from Revealed Preference Study Using Bikeshare Data. *Transportation Research Record*, 2021, p. 03611981211017136.
[252] Sarstedt, M., C. M. Ringle, and J. F. Hair. Partial least squares structural equation modeling.In *Handbook of market research*, Springer, 2021. pp. 587-632.
[253] Jann, B. Predictive margins and marginal effects in Stata. 2013.
[254] Lee, J. H., S. J. Hardman, and G. Tal. Who is buying electric vehicles in California? Characterising early adopter heterogeneity and forecasting market diffusion. *Energy Research & Social Science,* Vol. 55, 2019, pp. 218-226.
[255] Bas, J., C. Cirillo, and E. Cherchi. Classification of potential electric vehicle purchasers: A machine learning approach. *Technological Forecasting and Social Change,* Vol. 168, 2021, p. 120759.
[256] Khattak, Z. H., and A. J. Khattak. Spatial and unobserved heterogeneity in consumer preferences for adoption of electric and hybrid vehicles: A Bayesian hierarchical modeling approach. *International Journal of Sustainable Transportation*, 2021, pp. 1-14.
[257] Forrest, R., and A. Kearns. Social cohesion, social capital and the neighbourhood. *Urban studies,* Vol. 38, No. 12, 2001, pp. 2125-2143.
[258] Plötz, P., U. Schneider, J. Globisch, and E. Dütschke. Who will buy electric vehicles? Identifying early adopters in Germany. *Transportation Research Part A: Policy and Practice,* Vol. 67, 2014, pp. 96-109.
[259] AFDC. *Alternative Fuels Data Center: Federal Laws and Incentives*. https://afdc.energy.gov/laws/all?state=US. Accessed January 25, 2023.
[260] Fueleconomy.gov. *Federal Tax Credits for Plug-in Electric and Fuel Cell Electric Vehicles Purchased in 2023 or After*. https://www.fueleconomy.gov/feg/tax2023.shtml. Accessed February 28, 2023.
[261] Electrification-Coalition. *INFLATION REDUCTION ACT IMPACT ON ELECTRIC VEHICLES*. https://electrificationcoalition.org/work/federal-ev-policy/inflation-reduction-

act/#:~:text=For%20the%20first%20time%2C%20used,Income%20caps%20also%20apply.
Accessed February 28, 2023.
[262] Loveday, S. *2021 Chevrolet Bolt EV Discounts Soar Prior To Redesign Reveal*.
https://insideevs.com/features/489389/2021-chevrolet-bolt-ev-deals-soar-ahead-2022-model/.
Accessed February 26, 2023.

VITA

A. Latif Patwary grew up in Chandpur, a small city in Bangladesh. After high school and college education, he attended Bangladesh University of Engineering and Technology (BUET) and received his bachelor's degree in urban and regional planning. Before graduating with his undergraduate degree, he knew he wanted to attend graduate school. He started his Doctor of Philosophy program in Civil Engineering with a concentration in Transportation Engineering in 2020 at the University of Tennessee, Knoxville (UTK). He also earned two master's degrees: one in natural resource economics and another in Statistics from UTK. His main fields of research interests include transportation data and planning, traffic safety, travel behavior analysis, smart mobility, sustainability, and transportation economics. He authored seven journal articles, two technical reports, and 16 conference proceedings. Additionally, he serves as a Reviewer for the Transportation Research Board Annual Meeting and Transportation Research Record, having completed over 20 reviews. He is incredibly grateful for all the support from his family, friends, advisors, and colleagues.