Center for Advanced Multimodal Mobility Solutions and Education

Project ID: 2021 Project 04 & 2022 Project 11

Estimation of Pedestrian Compliance at Signalized Intersections Considering Demographic and Geographic Factors

Final Report

By

John Ivan (ORCID ID: https://orcid.org/0000-0002-8517-4354)
Professor, Department of Civil and Environmental Engineering
University of Connecticut
261 Glenbrook Road Unit, 3037, Storrs, CT 06269-3037
Phone: 860-486-0352; Email: john.ivan@uconn.edu

Amy Burnicki (ORCID ID: https://orcid.org/0000-0002-5232-4555)
Associate Professor in Residence, Department of Geography
University of Connecticut
Phone: 860-465-9644; Email: amy.burnicki@uconn.edu

Quinn Packer (ORCID ID: https://orcid.org/0000-0003-2219-9148)
Graduate Research Assistant, Department of Civil and Environmental Engineering
University of Connecticut
Phone: 860-996-2665; Email: quinn.packer@uconn.edu

for
Center for Advanced Multimodal Mobility Solutions and Education
(CAMMSE @ UNC Charlotte)
The University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC 28223

June 2022

ACKNOWLEDGMENTS

This project was funded by the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE @ UNC Charlotte), one of the Tier I University Transportation Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT), under the FAST Act. The authors are grateful to Vincent Huang for his help in the GIS data preparation.

DISCLAIMER

TABLE OF CONTENTS

TABLE OF CONTENTS **4**
LIST OF FIGURES **5**
LIST OF TABLES **5**
ABSTRACT **6**
1. INTRODUCTION **7**
2. LITERATURE REVIEW **9**
2.1 Impact of Land Use and “Built Environments” on Pedestrian Safety 9
2.2 Disparity of Income on Pedestrian Safety 9
2.3 Pedestrian Observations 10
2.4 Exclusive and Concurrent Signal Phasing 10
2.5 Traffic Volume 12
2.6 Crosswalk Characteristics 13
2.7 Regression Techniques 13
3. DATA PREPARATION **16**
3.1 Odds of Compliance 16
3.2 Buffer Zone Creation 17
3.3 Land Use Densities 17
3.4 Weighted Population Density 19
3.5 Poverty Ratio 22
4. METHODOLOGY **26**
4.1 Model Formulation 26
4.2 Spatial Autocorrelation (Local Moran’s I) 26
4.3 Data Analytics and Random Effects 29
5. RESULTS **31**
5.1 Correlation and Collinearity Analysis 31
5.2 Half-Mile Buffer Results 32
5.3 Quarter Mile Buffer Results 33
5.4 Eighth Mile Buffer Results 37
6. DISCUSSION **41**
7. CONCLUSION **45**
REFERENCES **47**

LIST OF FIGURES

FIGURE 1. ODDS OF COMPLIANCE FREQUENCY DISTRIBUTION 17
FIGURE 2. EXAMPLE OF ROAD NETWORK AND CROSSWALK LOCATIONS 18
FIGURE 3. EXAMPLE SERVICE AREA BUFFER 18
FIGURE 4. LAND USE DENSITY RASTER WITH BUFFERS EXAMPLE 20
FIGURE 5. BLOCKS WITH WEIGHTED POPULATION DENSITY 21
FIGURE 6. EXAMPLE WEIGHTED POVERTY RATIO: QUARTER MILE BUFFER 22
FIGURE 7. TRANSFORMATION OF THE ODDS OF PEDESTRIAN COMPLIANCE 27
FIGURE 8. QUARTILE-QUARTILE PLOT: LOGARITHMIC TRANSFORMATION OF COMPLIANCE 27
FIGURE 9. CONNECTIVITY GRAPH, 2000 FT THRESHOLD 28
FIGURE 10. MORAN'S I RESULTS: QUARTER MILE BUFFER 28
FIGURE 11. MORAN'S I SIGNIFICANCE MAP: QUARTER MILE BUFFER 29
FIGURE 12. CROSSWALK 63 30
FIGURE 13. INTERSECTION 192 42

LIST OF TABLES

TABLE 1. NLCD LAND USE DENSITIES DEFINITION 19
TABLE 2. DISCRETE VARIABLE DATA 23
TABLE 3. NON-BUFFER SPECIFIC CONTINUOUS VARIABLES 24
TABLE 4. HALF MILE BUFFER SPECIFIC VARIABLES 24
TABLE 5. QUARTER MILE BUFFER SPECIFIC VARIABLES 25
TABLE 6. EIGHTH MILE BUFFER SPECIFIC VARIABLES 25
TABLE 7. CORRELATION ANALYSIS BETWEEN CATEGORICAL VARIABLES 31
TABLE 8. PRELIMINARY MODEL FORMULATION: HALF MILE BUFFER 33
TABLE 9. PRELIMINARY MODEL FORMULATION: HALF MILE BUFFER 34
TABLE 10. FINAL MODEL: HALF MILE BUFFER 34
TABLE 11. PRELIMINARY MODEL FORMULATION: QUARTER MILE BUFFER 35
TABLE 12. PRELIMINARY MODEL FORMULATION: QUARTER MILE BUFFER 36
TABLE 13. FINAL MODEL FORMULATION: QUARTER MILE BUFFER 37
TABLE 14. PRELIMINARY MODEL FORMULATION: EIGHTH MILE BUFFER 38
TABLE 15. PRELIMINARY MODEL FORMULATION: EIGHTH MILE BUFFER 39
TABLE 16. FINAL MODEL FORMULATION: EIGHTH MILE BUFFER 40
TABLE 17. QUARTER MILE BUFFER: RELAXED COMPLIANCE 44

ABSTRACT

Walking as a main mode of transportation is continuously growing as a choice of mobility. This is due largely in part to its positive impacts on environmental and socioeconomic sustainability of the surrounding area. However, it is crucial that city planners and transportation engineers provide facilities that are safe for these vulnerable road users. Surrounding land use and demographic factors may be associated with pedestrian compliance with a given traffic signal.

This project seeks to create an understanding of how the surrounding land use and demographic characteristics of an intersection influence pedestrian compliance with a given crossing signal. This study uses pedestrian observation data collected from 145 crosswalks at 42 intersections in Connecticut. Pedestrians were recorded as being compliant if they crossed on the correct signal phasing and remained in the designated crosswalk for the entire crossing. The odds of compliance at each crosswalk are used as the response variable in a log-linear regression model which is being predicted by collected physical crosswalk site characteristics and geo-spatial data, including demographics and land use. The study analyzes three different buffer sizes for the geo-spatial data: half-mile, quarter-mile, and eighth-mile. The results from the study show that the quarter-mile land use buffer yielded the best model fit with all variables included being statistically significant at 95%. High density land use area, weighted population density, sidewalk presence, intersections with exclusive phasing, and day of the week all decrease the odds of pedestrian compliance with signal phasing. Medium density land use area, low density land use area, and crosswalk presence increase the odds of pedestrian compliance with signal phasing.

1. INTRODUCTION

Safety is a major concern in the field of transportation planning and engineering. How someone perceives the safety of a given mode of transportation may directly impact their decisions to use that mode. The boom of the automobile in the early 1900's brought with it a transportation engineering bias which provided facilities for cars while simultaneously destroying the safe environment which promoted walking. Pedestrians, who previously traversed the roadway network freely, had fallen victim to the impact of the automobile. The automobile was so dominant in transportation that the auto industry created a term for those who chose to walk outside of the designated crossing area: "jaywalkers." Jaywalking refers to one who crosses "between adjacent intersections controlled by traffic control signal devices or by police officers, pedestrians shall not cross the roadway at any place except in a crosswalk" [1]. This law, which still exists in much of the United States today, further impacted walking as a viable mode of transportation. Until recently, American streets were perceived to be restricted to motor-vehicle traffic only.

Pedestrians, who once roamed streets freely, were classified as "vulnerable road users" [2]. Vulnerable road users are those who travel unprotected by an external shield and could sustain greater injury from a collision with a motor vehicle. It was believed that pedestrians, who could no longer cross the street at will, needed to abide by marked crossing locations and crossing signals to be safe. Creating signals with adequate phasing which could provide the pedestrian with a safe crossing proved to be difficult, with pedestrians often getting caught in-between the transition from walking traffic to vehicle traffic [3]. One of the earliest known uses of the "Walk/Don't Walk" signal we know today dates to New York City in 1938 [4], yet there are no evaluations on how it impacted pedestrian safety.

The impact of the automobile on transportation was felt throughout the country. However, the largest impact occurred in the dense, urban environment of cities. Historic cities such as New York and Chicago, which had an established transportation network, had their roadways widened and sidewalks squeezed against buildings, yet the underlying roadway network remained [5].

Cities whose development coincided with the boom of the automobile were designed around car-centric infrastructure. Wide streets and large intersections were implemented with the goal of moving vehicles quickly and efficiently. This car-centric architecture can be found in Detroit, Michigan, and one of the cities evaluated in this study, Hartford, Connecticut. Hartford, like many other cities, fell victim to having its urban core dismantled and reorganized to accommodate the automobile [6]. American cities, which generally represent high density environments, claim 82% of all pedestrian-vehicle collisions when compared to the lower density rural and suburban environments [7]. This implies a considerable influence on the surrounding land use of an intersection and a pedestrian's safety while crossing.

This project explores how a combination of crosswalk characteristics, surrounding land use density, and population density influence pedestrian compliance with crossing signals found in Connecticut. The State of Connecticut utilizes two different pedestrian signal phasing options at intersections: exclusive phasing and concurrent phasing. These two phasing approaches can be defined as follows:

- Exclusive: Pedestrians cross simultaneously during a separate phase where all vehicular approaches have red indications [8]
- Concurrent: Pedestrians cross parallel to through vehicles during phases shared with vehicular traffic. [8]

Exclusive phasing is typically considered safer due to its separation of vehicle and pedestrian movement. However, because movement of pedestrian and vehicle traffic is completely separated via signal phasing, there is a larger negative impact on vehicle and pedestrian travel when compared to concurrent phasing. This is because concurrent phasing allows for pedestrian and automobile traffic traveling in the same direction to cross at the same time, which reduces the delay of all traffic movements.

It is important to note that Connecticut Department of Transportation (CTDOT) differentiates between side street green and concurrent phasing as follows:

- Side Street Green: Pedestrians cross while parallel motorists have a green signal indication. Both motorists and pedestrians share green ball indication. There is no pedestrian "WALK/ DON'T WALK" signal heads [9]
- Concurrent: Pedestrians cross while parallel motorists have a green signal indication. Pedestrians have a separate "WALK/DON'T WALK" signal head [9]

Although there is a slight difference between these two signal types, the movement of pedestrians is identical for both. Therefore, both signal types are categorized as concurrent in this study.

Pedestrian compliance with regulations and traffic signals is a factor influencing vehicle-pedestrian crashes [10]. This can be applied to pedestrian movement to interpret the safety a given crosswalk provides. Compliance with signals in this study is based on the laws in section 14-297, 14-300, and 14-300b of the General Statutes of the State of Connecticut which are summarized as follows:

- At intersections with the pedestrian-control signal "Walk/Don't Walk, pedestrians may only cross when indicated by the signal.
- Pedestrians may not cross at a pedestrian signal-controlled intersection against red or "Stop."
- Pedestrians must cross at marked crosswalks when present, during which time all vehicle traffic must yield.
- No pedestrian shall cross an intersection diagonally unless given authorization by police officer or pedestrian-control signal.
- No pedestrian shall cross a roadway between adjacent intersections which feature pedestrian-control signals.

Compliance at crosswalks is defined in this study using the statutes above. For concurrent signal phasing, pedestrians are only considered compliant if they cross within the designated crossing area and on the corresponding direction vehicle green. During exclusive phasing, pedestrians are only considered to be compliant if they cross within the designated crossing area and face a "Walk" signal. The total observations at each crosswalk location are then aggregated into a proportion of compliance which is used to calculate the odds of pedestrian compliance at the crosswalk. These odds of compliance with an intersection crossing will provide insight into what site characteristics and land use variables either positively or negatively influence pedestrian compliance and safety. With the inclusion of this data to predict pedestrian compliance, transportation planners may be able to understand what factors influence compliance.

2. LITERATURE REVIEW

Understanding what influences pedestrian compliance and safety at intersections has been a topic of research for decades. However, what makes a given pedestrian facility safe can be difficult to quantify and may not be apparent on the surface. Studies which analyze pedestrian crossing behavior and safety can provide insight into this study.

2.1 Impact of Land Use and "Built Environments" on Pedestrian Safety

The environment surrounding a crosswalk can influence pedestrian compliance with crossing signals and crossing behavior. However, it is important to understand how much of the land area and demographics surrounding the crosswalk influences pedestrian crossing behavior. A study published in 2011 by Miranda-Moreno et al. 2011 [11] uses land-use and demographic data surrounding the crosswalk in combination with characteristics from the roadway network to understand how they influence pedestrian activity and crashes at intersections. The study evaluated buffer sizes of 50, 150, 400, and 600 meters with the belief that the smaller buffers would encapsulate an intersection's immediate surroundings, with the larger buffers serving as proxies for characteristics of the land use at a walking distance or neighborhood level.

Miranda-Moreno et al. utilized land use data, which is referred to as "Built Environment," are gathered using GIS resources which are categorized into commercial, residential, industrial, parks, open space, and government areas in square meters. Demographics including population counts were gathered using census tract data and calculated by their intersection with each buffer. The results showed that the land use variables of commercial areas and open space increased pedestrian activity around intersections by 10.7% to 11.1%.

Pedestrian activity and built environment are further shown to be correlated in an earlier study by Pulugurtha et al. in 2008 [12]. Pedestrian volume and AADT were also correlated with pedestrian frequency. However, the results show that the effects of built environment on pedestrian crashes are largely dependent on pedestrian activity and traffic. Pulugurtha et al. used quarter mile, half mile, and mile buffers at 176 intersections in the city of Charlotte, North Carolina to estimate pedestrian activity at intersections using multi-linear regression. Land use type was divided into 22 possible categories. The results showed that positive coefficients of commercial areas and population again reinforce that pedestrian activity at intersections increases with these variables.

Finally, Ren et al. [13] performed an observational study of pedestrian crossing behavior at three cities in China. The results showed that when pedestrians cross in groups, non-compliant behavior at intersections increased. These studies indicate the tremendous influence population surrounding intersections has on pedestrian compliance with signals and safety.

2.2 Disparity of Income on Pedestrian Safety

The disproportional rate of pedestrian injury and deaths to motor-vehicle deaths hasn't gone unnoticed in recent years. There have been strives to improve pedestrian safety, but there is a concern that the problem is not being systematically addressed, leaving low-income areas behind. In 2014, the Safe Routes Partnership published a report which exemplified the disparity in safety between lower and higher income neighborhoods [14]. The results showed that in low-income metro areas, pedestrian fatality rates were more than double those of higher income neighborhoods (10.4 deaths per 100,000 people vs 5.0 deaths per 100,000 people respectively). This research points to a lack of pedestrian infrastructure as the main cause of high pedestrian collisions in the low-income neighborhoods.

A similar study published in 2019 by Lin et al. [15] used poverty rate collected at the block group level in combination with land use data and crosswalk characteristics to predict pedestrian crash frequency, pedestrian behavior, and possible injury. The results from the pedestrian crash frequency model suggest that crashes are more frequent in low-income areas with high populations. Lin et al. suggest that this is due to a larger zero-auto dependency when compared to higher income, lower population areas. High rates of pedestrian crashes in low-income areas when compared to higher income areas is further replicated in the study by Noland et al. in 2013 [16].

2.3 Pedestrian Observations

The two most common forms of observing pedestrian behavior come from on site, live observations and data collected from video recordings. As society progresses further into the digital age, video recordings and other utilizations of technology can lead to far greater quantities of data collection. However, the development of technology which can be used to recognize how a pedestrian is behaving automatically is still a work in progress with some inaccuracies between the automated video-assessment methods and the manual pedestrian observations [17] [18]. Therefore, in-person observations are still the most accurate methods of obtaining pedestrian behaviors.

2.4 Exclusive and Concurrent Signal Phasing

Pedestrian signal compliance at intersections has only recently seen a major increase in research even though it is estimated that 40% of all pedestrian crashes occur at intersections [19]. The signal phasing is often a point of focus when trying to understand pedestrian compliance and safety. Studies which include exclusive phasing are some of the most popular. The outcomes of the impact on pedestrian safety from exclusive signal phasing is mixed, since while exclusive phasing has the potential to improve pedestrian safety, it also significantly increases pedestrian and vehicle delays. This delay can lead pedestrians to disobey the signal phasing, negating any safety benefits exclusive phasing can provide. Due to this, it is important to understand how different signal phasing can influence pedestrian safety and compliance.

In 2009, a pilot study on exclusive phasing, referred to as "Pedestrian Scramble Operations" or "PSO", was conducted. In the study, exclusive phasing was implemented at two intersections in the downtown area of Calgary, Alberta, Canada with the purpose of understanding how exclusive phasing impacts pedestrian safety. The intersection crossings were categorized into "safe side" (concurrent with vehicle movement) and "unsafe side" crossings. The results showed that exposure variables such as peak hours, two-leg pedestrian crossings, and logarithm total number of pedestrians had a negative overall effect on pedestrian compliance. The results from that PSO study showed 13% of pedestrian non-compliant crossings occurred on the "safe-side" while only 2% were "unsafe-side" crossings. These results indicate that most non-compliant observations occur in the same direction as vehicle traffic. The results also showed that the newly implemented exclusive phasing significantly reduced the total number of pedestrian-vehicle crashes which occurred. However, the total amount of pedestrian signal violations increased with the switch from concurrent phasing to exclusive phasing. [20].

The results from that PSO study fall in line with a Bechtel et al. study performed in 2004 which suggested that approximately 25% of signal violations occurred on "safe-side" crossings. This PSO also analyzed the impact of exclusive phasing on pedestrian safety before and after installation and found that the introduction of exclusive pedestrian phasing to an intersection reduces the number of pedestrian-vehicle conflicts while simultaneously increasing the total number of instances where pedestrians were non-compliant with signal phasing. The study suggests the incorporation of “safe-side” crossings could potentially reduce the total number of signal violations and improve both pedestrian and vehicle traffic [21].

A study performed by Zegeer et al. took a broader approach to understanding the impact of pedestrian signal and signal timing on pedestrian crashes through the inclusion of concurrent, exclusive, and intersections which lacked pedestrian signals. The results from this study showed that there was no discernable difference in safety between intersections with concurrent phasing and intersections with no signal. In contrast, those intersections with exclusive phasing resulted in lower pedestrian crashes. This study investigated 2081 pedestrian crashes from 1297 intersections across the country. It is important to note that while pedestrian safety was superior for most exclusive phase signals, intersections where pedestrian volumes were less than 1200 per day did not see an improvement in safety. Concurrent phasing in the study also revealed a considerable proportion of pedestrian-vehicle crashes occurred due to vehicle turn-in maneuvers [22], defined as a crash between a pedestrian and a turning vehicle. This may indicate intersections with concurrent phasing may lack the safety provided by intersections with exclusive phasing.

Some of the most recent studies associated with pedestrian compliance with exclusive and concurrent phasing also were conducted in Connecticut. Zhang et al. published their findings on the safety impacts of exclusive and concurrent phasing at pedestrian crossings in 2015 [23]. The goal of that Connecticut study was to estimate pedestrian-vehicle crash counts and interaction severity prediction models for concurrent or exclusive phased intersections at 42 intersections across the state. Crash counts were gathered at each intersection over a period of eight years. The four conflict types were categorized based on the Swedish Traffic Conflict Technique developed by Lund University [24] and included undisturbed crossings, potential conflicts, minor conflicts, and major conflicts. Pedestrian observations were collected by trained observers during up to 6-hour time periods at each location. Site characteristics in addition to exclusive and concurrent signal phasing were gathered such as crossing distances, sidewalk presence, and crosswalk presence. The results from the study by Zhang et al. revealed that the probability of conflicts decrease at exclusive phased intersections when pedestrians crossed during the exclusive walk phase and increased when pedestrians jay-walked or crossed on green when exclusive phasing was an option. The probability of pedestrian-vehicle conflict occurring increased with exclusive phasing when compared to those intersections with concurrent phases.

It is important to note that the results from the Connecticut study contradict those found in the previous study in Calgary, Alberta, Canada [21] which found pedestrian vehicle conflict decreases with exclusive phasing at intersections. However, the Calgary pilot study only analyzed the impact of exclusive phasing at one intersection while Connecticut included 42 intersections. Therefore, it is possible the conflicting results from the Calgary study may be a result of another external factor, such as geo-spatial variables, which were not considered.

One of the biggest points to note from the Connecticut study is that although exclusive phasing appeared to have fewer crashes, those crashes which did occur tended to be far more severe in comparison to crashes which occurred at intersections with concurrent phasing. These results indicate that exclusive phasing is safer than concurrent only when pedestrians abide by the crossing signal, which was only 15% of the total observations. There was a negative correlation between pedestrian volume and conflicts and a positive correlation between crosswalk distance and conflicts. The study suggests that exclusive phasing may provide more adequate crossing safety in locations with low pedestrian volumes [23]. It is important to note that Zhang et al. did not consider the demographics and land-use surrounding the intersection.

A similar study, published in 2015, utilized the same 42 intersections as this study. The objective was to understand what influences compliance and create a binary regression model which could predict pedestrian compliance with exclusive and concurrent phased traffic signals. Ivan et al. [25] used 14,838 observed pedestrian behaviors and characteristics in combination with crosswalk location features to best understand which variables influence compliance using the same definition of compliance [8] to be used in this study. However, an alternate definition of compliance was also used, referred to as "Relaxed Compliance." Under this definition, pedestrians are considered compliant at exclusive phasing intersections if they wait for the "WALK" signal or if they cross on the vehicle green in the same direction. This definition of compliance treats observations at exclusive phasing intersections the same as those with concurrent phasing. This modification to the study resulted in the influence of signal type on pedestrian compliance becoming insignificant. Ivan et al. collected limited land use characteristic variables, similar to those used in this study. The land use surrounding the intersection was categorized into residential, non-residential, and mixed. It is important to note that land use categorization was significant in both the "strict" and "compliant" model formulations.

The results from the study further indicated that pedestrians are more likely to be compliant with signal phasing when crossing an intersection that allows pedestrians to cross with concurrent vehicle traffic, like the results from studies mentioned previously [20] [21]. The results from the regression using the "Relaxed Compliance" definition suggest that pedestrians are most likely going to cross an intersection when they perceive it is safe to do so and not necessarily wait until it is legally permitted. These studies suggest that the consideration of exclusive versus concurrent phasing can provide further insight into pedestrian signal compliance.

2.5 Traffic Volume

Pedestrian and vehicle traffic volumes are an important consideration in understanding pedestrian signal compliance. The inclusion of these variables has been used in numerous pedestrian safety and pedestrian compliance studies. In 2018, Kevin Diependaele, from the Belgian Road Safety Institute, investigated the frequency of non-compliance at intersections across nine of the most popular cities in Belgium [26]. It is important to note that the study in Belgium consisted of mainly high-density urban areas. The motivation behind the study came from 62% of pedestrian-vehicle crashes occurring at pedestrian crossings. What he discovered was that although 21% of pedestrians violate pedestrian crossing signals, there is a large variation in non-compliance depending on certain site characteristics. Pedestrian and motor-vehicle traffic volumes, which typically increase intersection complexity, had a negative effect on pedestrian compliance at intersections. Diependaele attributes site characteristics such as number of lanes and crossing distance to traffic volumes and therefore assumes that it is one of the most important variables to account for. The model estimates revealed that during each 15-minute time period, one in five (21%) of the 69,211 total pedestrians observed failed to cross on the correct signal. However, it is noted that as both pedestrian and vehicle traffic volume increased, the frequency in which pedestrians were compliant increased.

Indication that traffic volumes directly influence pedestrian compliance is also seen in W. Andrew Harrell's 2010 publication in the Journal of Social Psychology [27]. In Harrell's study, which used 571 pedestrian observations from signalized intersections in Edmonton, Alberta, Canada, it was again shown that the pedestrian's compliance with traffic signals varies greatly depending on traffic and pedestrian volume, much like the results from Diependaele's study [26]. However, the results from Harrell's yield conflicting results with pedestrians' compliance being observed at a higher rate at locations with lower traffic volumes. Harrell explains that compliance at crossing locations with lower traffic volumes is likely higher since vehicles are allowed to travel at much higher speeds than areas with high vehicle volume.

These results fall in-line with Per E. Garder's study on speed and other variables on pedestrian safety in which he found as speeds road width increase so do pedestrian-vehicle crashes [28]. This is most likely related to the lack of pedestrian compliance at these roadway locations. The influence of pedestrian volume on compliance again conflicted between the two studies. As pedestrian volume increased, pedestrian compliance decreased. Harrell goes on to reason that this is due to the "diffusion of responsibility" or "safety in numbers" effects [24]. He believes that pedestrians when solo feel more inclined to be non-compliant since they do not feel responsible for the safety of pedestrians around them and that pedestrians also feel they are better seen when crossing together.

The influence of pedestrian and vehicle traffic volumes on pedestrian compliance was also included in the study by Ivan et al. [25]. Garder found results which were both similar and dissimilar to those found in those discussed above [23] [24]. Ivan et al. discovered that while an increase in vehicle volume simultaneously increased pedestrian compliance, an opposite effect occurred with an increase in pedestrian volume as overall compliance decreased [25]. The varying results and interpretations of these studies indicate that additional research which incorporate traffic volumes is necessary to best understand how they impact pedestrian compliance.

2.6 Crosswalk Characteristics

The physical attributes of an intersection and how they influence pedestrian compliance with crossing signals are commonly investigated [11][13][16][23][25]. Sisiopiku et al. [28] performed an observational study of pedestrian behaviors at various urban crosswalks. The results from the study showed that although pedestrians are more likely to cross intersections with marked crosswalks, crosswalks have little influence on pedestrian compliance with signals. Sisiopiku et al. found the most influential factor on signal compliance to be crossing distance. Mukherjee et al. similarly looked at physical attributes of crosswalk intersections [29]. The results showed that crosswalk attributes such as on-street parking and absence of marked crosswalks lead to non-compliant intersection crossings.

2.7 Regression Techniques

The pedestrian observations used in this study are quantified in a binary response, compliant or non-compliant. However, the variables of interest are related to the land use and population density surrounding the crosswalk. To understand how the variables surrounding a crosswalk location influence pedestrian compliance, the observation data are aggregated by intersection and represented by the odds of pedestrian compliance with signal phasing. The log-odds of compliance at each crosswalk is used as the response variable in a log-linear regression which transforms the variable to obtain a normal distribution.

The aggregation of binary true/false response data is exemplified by Sackett et al [30] which highlights the usefulness of odds as a clinically useful measure of treatment. The odds or likelihood is often used as a measurement of risk. In a study performed by Sacket et al [30], the odds allow for a more comprehensive comparison of risk between multiple groups and attributes when compared to raw percentages which can be arbitrary. Therefore, the compliance, and inherently, the risk a pedestrian faces at an intersection could be compared more accurately across crosswalk site locations.

This study deals with pedestrian compliance at intersections, which is a relatively random phenomenon. However, pedestrian observations are aggregated into a percentage of compliance before being converted to odds of compliance at an intersection. The conversion from percentage of compliance with odds of compliance is due to linear modeling not being able to adequately predict probability. This is because as the probability does not increase as the predictors increase [31]. Odds can take any negative or positive number, unlike probabilities which must be between 0 and 1.0. This allows for linear model formulation. The predicted odds for the values of predictors can also be easily converted back into probabilities. Coefficients from the model, when exponentiated, can also be interpreted as the change of odds of an outcome per unit change for that predictor.

It is easy to make misleading interpretations when using odds. When the odds are greater than 1.0, the interpretation is intuitive. For example, an odds value of 6.0 indicates that six people experience an event for every one that does not. However, interpreting the odds with a value less than 1.0 can be less intuitive, which is explained in Davies, et al. study of the sometimes-misleading understandings of odds [32]. This misunderstanding is exemplified by using an odds value of 0.2 which is interpreted as 0.2 people experience the event for every one that does not, or there will be one pedestrian compliant for every five non-compliant observed (17%) [32]. The given odds of compliance at an intersection being less than 1.0 can lead to a convoluted understanding on how many pedestrians were observed versus how many were compliant. Understanding that the odds do not directly represent total pedestrians, but rather the percentage of pedestrians that are compliant given a set number of observations, is critical to interpreting the study properly. Another misinterpretation is to say "when odds ratio is less than 1.0 then it is always smaller than the relative risk. Conversely, if the odds ratio is greater than 1.0 then it is always greater than the relative risk" [32]. Understanding how to interpret the odds compliance and the implications it brings is important to this study.

When the distribution of the independent variable to be used in linear regression does not follow a normal distribution, transformations must be considered. When using the odds as the response variable, a logarithmic transformation allows for a more normal distribution centered around zero. Ronald Christensen describes the logarithmic rescaling best by saying how "probabilities between zero and one correspond to log odds between negative infinity and zero. Probabilities between one half and one correspond to log odds between zero and infinity" [33]. The odds, which previously followed a skewed right distribution, become symmetrical around zero. The use of logarithmic transformations of the dependent variable have been used extensively in the field of transportation. Examples can be seen in Wu et al. [34] and in the Federal Highway Administration's Modeling of Intersection Crash Counts and Traffic Volumes from 1997 [35]. Both use a logarithmic transformation of the dependent variable of the study to obtain a more normal distribution which can be used in linear modeling.

The sites of interest used in this study are all located in Hartford County, therefore, there is a strong possibility that the site characteristics of one crosswalk location may be similar to those of a neighboring crosswalk. To account for such correlation, a spatial auto-correlation analysis, namely Local Moran's I, is employed. Formally introduced in 1995 by Luc Anselin, Local Moran's I allow for researchers to identify local patterns or associations also referred to as "hotspots" [36]. The statistics resulting from a Local Moran's I analysis can provide insight into data at the local level. One takeaway from Moran's I analysis is it does not utilize traditional significance values. Instead, "pseudo" significance values are used via a randomization or permutation approach to significance testing.

A study published in 2020 by Ho et al. [37] uses a Moran's I analysis on pedestrian crashes in Hillsborough County, Florida. The goal of the study was to prove that pedestrian crashes are not randomized and occur at common locations. The study was able to find pedestrian crashes occurring at hotspot intersections and roadway segments.

In summation, the previous studies have included variables such as geospatial and demographic characteristics and the physical attributes of the crosswalk to predict pedestrian compliance with crossing signals and safety. Areas of lower income, higher population, and higher density land areas result in higher pedestrian crashes and lower compliance with pedestrian signaling. Analyzing several buffer sizes has helped to identify the best fitting models. However, no single study has included all of these variables mentioned to predict pedestrian compliance. This study seeks to use the combination of all variables mentioned above, calculated using three different buffer sizes, to predict their association with odds of pedestrian compliance with signalized crosswalk locations.

3. DATA PREPARATION

3.1 Odds of Compliance

The archived data used in this study contain individual observations of pedestrian behavior at crosswalk locations. These data were also used in previous studies, namely Zhang et al. 2015 [23] and Ivan et al. 2015 [25]. There are a total of 5753 pedestrian observations at 152 crosswalks located in Hartford County. The goal of this study is to predict the odds of pedestrian compliance at a crosswalk location based on surrounding land use densities and demographics; therefore, the observational data must be aggregated to represent the corresponding crosswalk that they were observed. The process of data aggregation is as follows:

- *Define Compliance*: Pedestrians are considered compliant in this study when they cross during the correct phase for the signal type and remain in the crosswalk for the entirety of the crossing. These observations are categorized as compliant and those which fail to meet this criterion are categorized as non-compliant.
- *Aggregation by Crosswalk:* The number of compliant observations at each crosswalk is divided by the total number of observations at the crosswalk. This process yields the proportion of compliant pedestrian observations at all 152 crosswalk locations ranging in values 0 to 1.0.
- *Odds of Compliance:* The percentage of compliance at each location is converted into the odds of compliance which provides the ability for linear modeling and a better interpretation of those who are compliant compared to those who are not. However, there are a few crosswalk locations where all pedestrian observations were compliant, or all observations were non-compliant. The log of 0 is negative infinity, which presents a problem in modeling.

To remedy this, a transformation proposed by Smithson and Verkuilen [38] is applied to the percentage of compliance prior to the calculation of odds, shown in **Equation 1** and **Equation 2**. Odds of compliance is calculated by the transformed percentage of observations which are compliant divided by the percentage of observations which are non-compliant **(Equation 3)** The distribution of the odds of pedestrian compliance with signal phasing is shown in **Figure 1.**

$$\boldsymbol{y^* = \frac{y(N-1)+1/C}{N}} \tag{1}$$

Where:
y = Percentage of Compliance at a Crosswalk Location
N= Number of Observations (152)
C=Number of Groups (2)

Therefore, the transformation of each observation becomes:

$$y^* = \frac{y(152-1)+\frac{1}{2}}{152} \tag{2}$$

$$Odds\ of\ Pedestrian\ Compliance = \frac{\%\ of\ Pedestrian\ Compliant\ Observations}{1-\%\ of\ Pedestrian\ Compliant\ Observations} \tag{3}$$

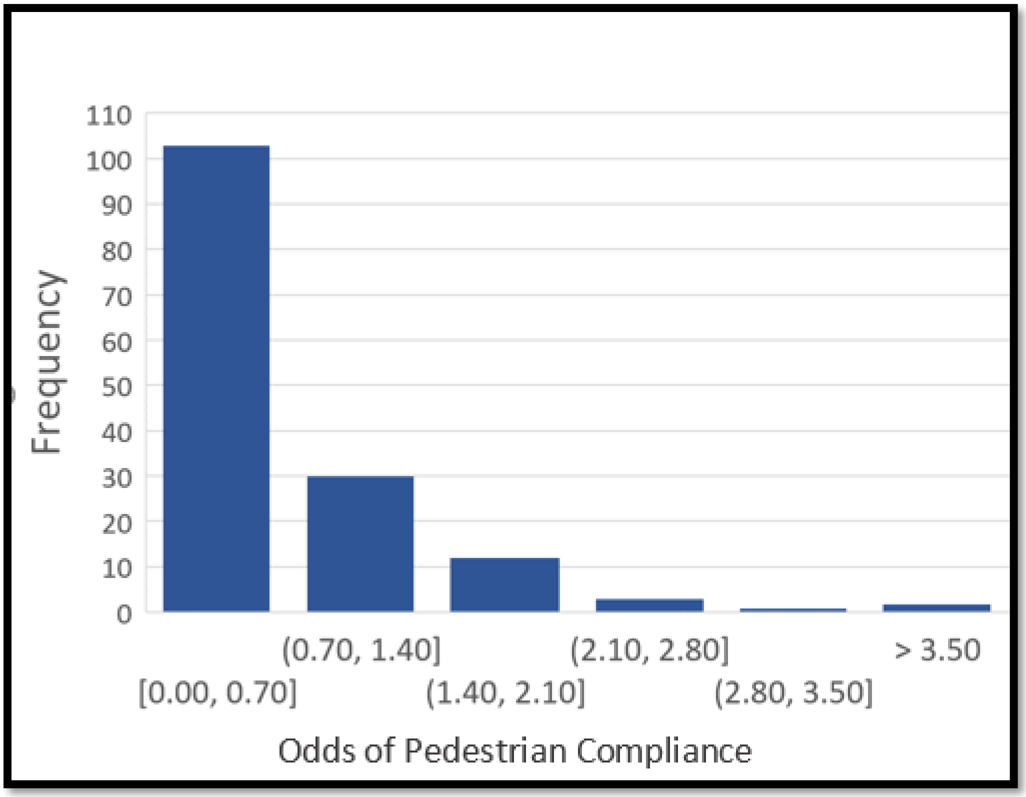

Figure 1. Odds of Compliance Frequency Distribution

3.2 Buffer Zone Creation

To encapsulate the land use densities and demographics surrounding an intersection, three buffers of 1/8th, 1/4th, and 1/2 miles must be considered. However, these buffers are not created arbitrarily. The "Network Analyst" extension in ArcGIS is the main tool utilized in this step. OpenStreetMap provided the network which was used to create the service areas. The network was cleaned to remove unnecessary roadways and was converted to a network dataset to generate the three service area buffers.

Figure 2 and **Figure 3** illustrate the network used in the analysis and the three service area buffers.

3.3 Land Use Densities

The buffers which were created are used to calculate the land use densities. There are three land use densities considered in this study: high-density development, medium-density development, and low-density development. Three raster files from the 2013 National Land Cover Database (NLCD) provided the land use densities via a binary grid (1 for corresponding high/mid/low density land development, 0 for other or none). There are 96 classifications of land use provided by the NLCD raster. The land use densities used in this study and their description are shown in **Table 1.**

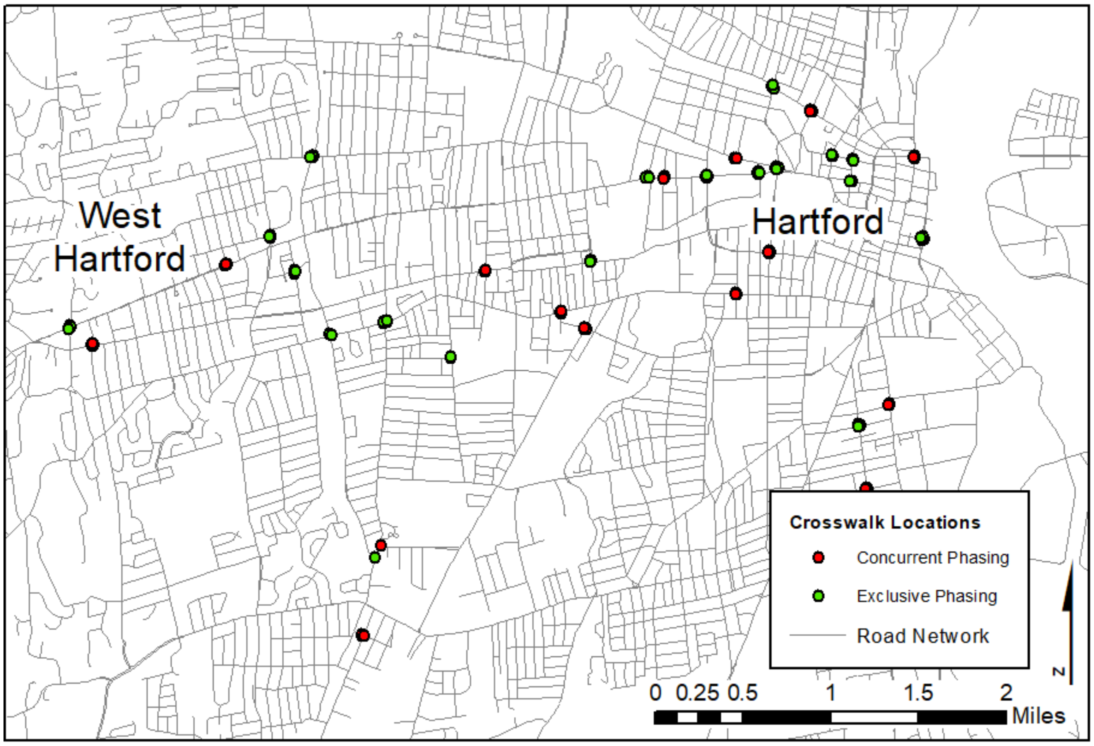

Figure 2. Example of Road Network and Crosswalk Locations

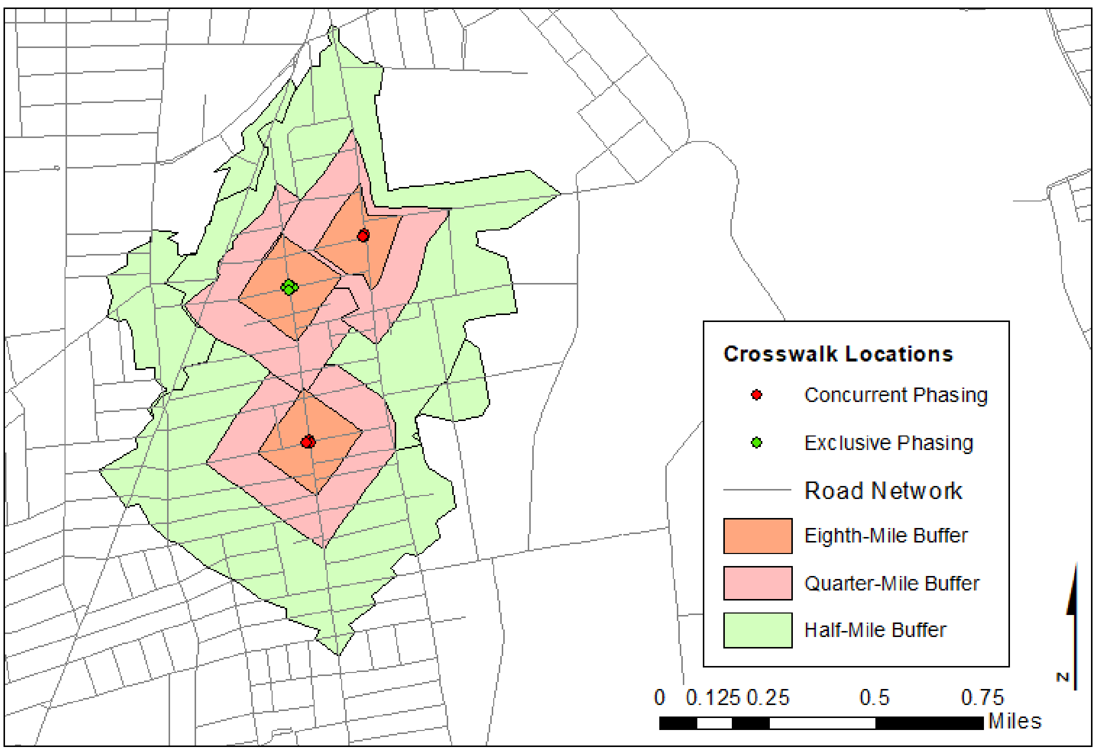

Figure 3. Example Service Area Buffer

Table 1. NLCD Land Use Densities Definition

22: Developed, Low Intensity Area	areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20% to 49% percent of total cover. These areas most commonly include single-family housing units.
23: Developed, Middle Intensity Area	areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50% to 79% of the total cover. These areas most commonly include single-family housing units.
24: Developed, High Intensity Area	highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80% to 100% of the total cover.

The “Zonal Statistics” operation in ArcGIS was used to calculate the number of “1s” in each of the three buffers. The sum of which provided the total count of each land use density in the buffer. Each raster cell represents a 30x30 meter (900 m^2) area, so the total sum of the land density cells is multiplied by 900 to get the total land cover in square meters. To obtain the area in square feet, the area in meters squared was multiplied by 10.7639. An example calculation for the high-density area in a half-mile buffer is shown below in **Figure 4**.

3.4 Weighted Population Density

The weighted population density utilizes table data and TIGER/Line shapefiles which were collected from the United States Census Bureau. Population was gathered at the block level to provide the highest accuracy. Table P1: Total Population Decennial Census Data provided the population data and 15-digit Block-ID’s. Block GIS shapefiles were downloaded from the 2010 Tiger/Line Shapefiles database. The population table and block shapefile are joined via the Block-ID in ArcGIS. The population density of each block is calculated by dividing the total population of the block by the total area of the block. The “Intersect” tool is used on the three buffer and the block group layers. The area of the new intersection is calculated, and the weighted population density is calculated by multiplying the population density of the block by the area created by the intersection of block groups and buffers. The sum of the weighted population densities for each intersection is calculated in the buffers to give the total weighted population density within the given buffer **(Figure 5)**.

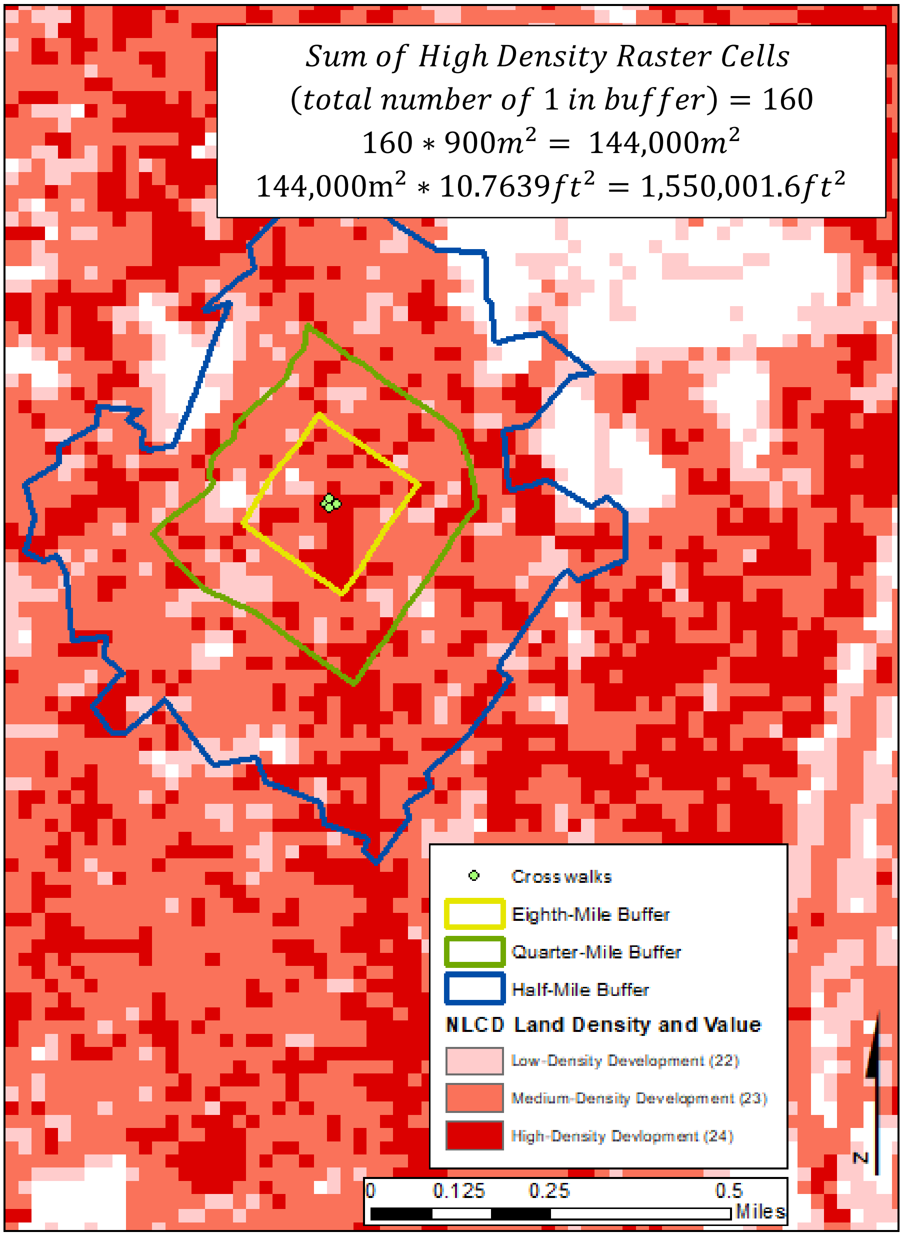

Figure 4. Land Use Density Raster with Buffers Example

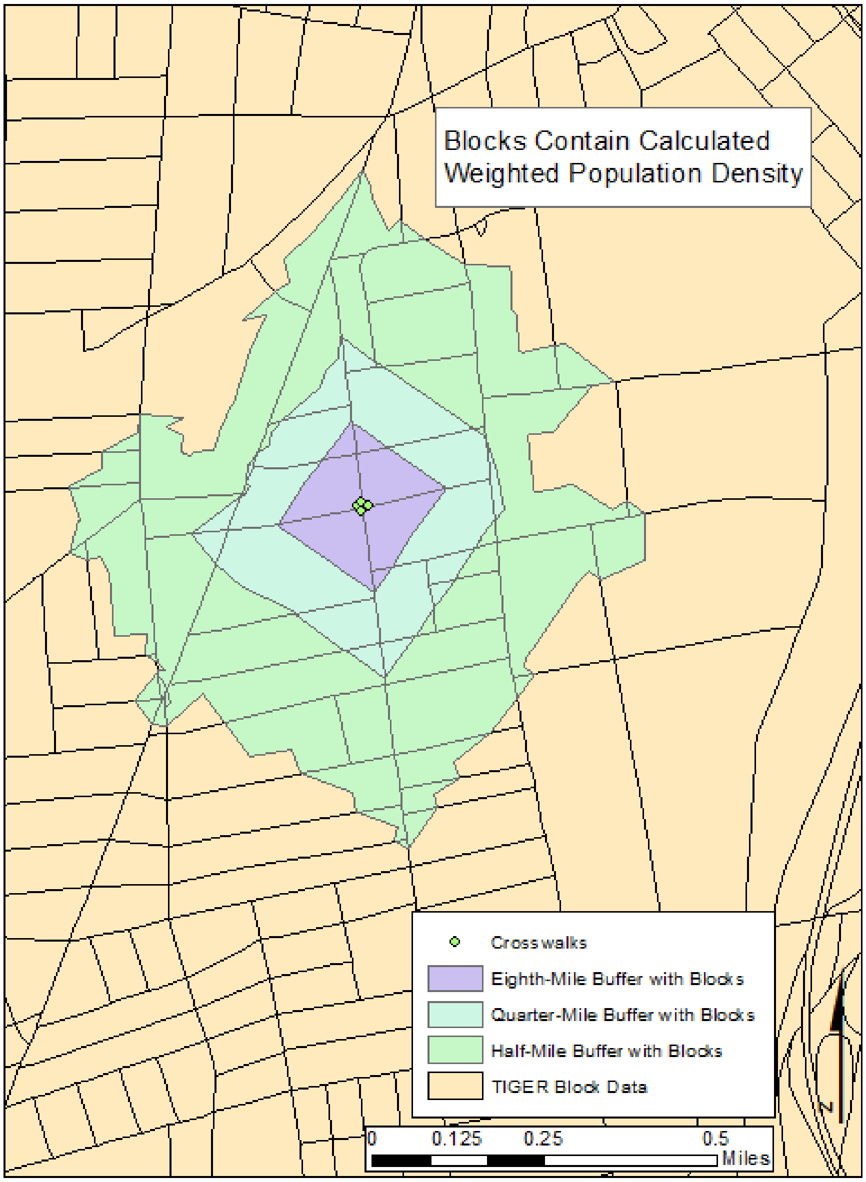

Figure 5. Blocks with Weighted Population Density

3.5 Poverty Ratio

The theory behind the inclusion of a poverty ratio in this study is its ability to capture the rate of people falling below the poverty line in an area in relation to those who are above it. Table *C17002: Ratio of Income to Poverty Level in the Past 12 Months* for the year 2015 was used to calculate the poverty ratio was collected from the United States Census Bureau. Poverty is defined as those whose income falls below the threshold value which is the "dollar amounts set by the U.S. government to indicate the least amount of income a person or family needs to meet their basic needs" [34]. The poverty line for each block group is calculated in this table and is represented as a "1". The total poverty ratio is calculated as the total count of people who fall below the poverty line divided by the total count of people who are above it. The three buffers contain multiple block groups so the weighted average of the poverty ratio for buffer is calculated. Weighted average is obtained by multiplying the poverty ratio of a block group within the buffer by the percentage of total area taken up by the block group, as seen in **Figure 6**.

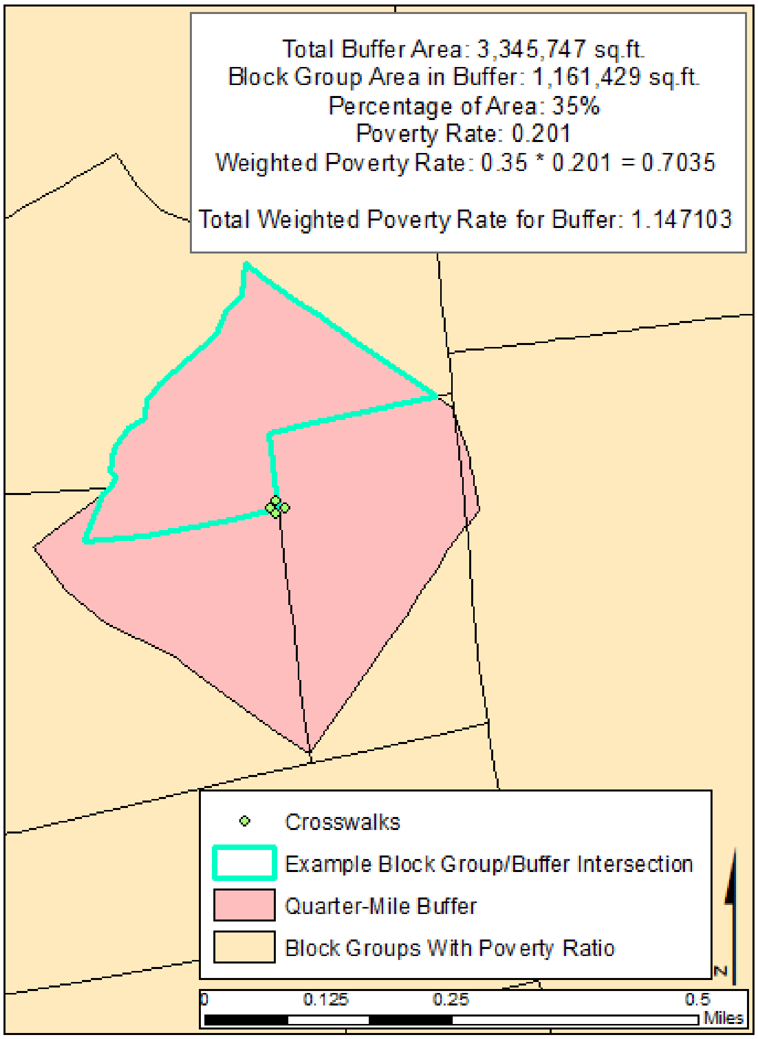

Figure 6. Example Weighted Poverty Ratio: Quarter Mile Buffer

The odds of pedestrian compliance at each crosswalk [Section 4.1], is combined with the 1/2, 1/4, and 1/8-mile buffers surrounding the intersections which have the calculated high, medium, and low-density development [Section 4.3], weighted population densities [Section 4.4] and the poverty ratios [Section 4.5]. Various transformations were used on continuous variables to obtain a normal distribution. Summary Statistics for the buffer specific and non-buffer specific variables can be found below in **Table 2, Table 3,**

Table 4**,**

Table 5, and Table 6.

Table 2. Discrete Variable Data

Variables with Levels	Number of Observations	Percentage of Observations
Signal Type		
Exclusive	93	64%
Concurrent	52	36%
Median Presence		
Yes	3	2%
No	142	98%
Speed Limit		
25	81	56%
30	39	20%
35	25	24%
On Street Parking		
Yes	74	51%
No	71	49%
Crosswalk Presence		
Yes	130	90%
No	15	10%
Weather		
Sunny	106	73%
Cloudy	35	24%
Rainy	4	3%
Sidewalk		
Yes	138	95%
No	7	5%
Day of the Week		
Monday	42	29%
Tuesday	36	25%
Wednesday	28	19%
Thursday	20	14%
Friday	19	13%

Table 3. Non-Buffer Specific Continuous Variables

Variables	Minimum	Mean	Max	Standard Deviation
Log Odds of Pedestrian Compliance (Response)	-5.7071	-1.0596	1.7165	1.2113
Crossing Distance (Feet)	23.7000	45.3890	86.0000	12.6685
Square Root Average Annual Daily Traffic (AADT)	21.7720	71.5139	142.2670	27.4776
Log Average Pedestrian Volume per Hour	0.28770	2.4602	4.8991	1.0281
Square Root Average Vehicle Volume Per Hour	6.63320	20.3203	38.9615	7.4329

Table 4. Half Mile Buffer Specific Variables

	Variable	Minimum	Mean	Maximum
Area (1000 ft²)	*High Density Land Development*	184.1000	3692.6000	6238.8000
	Medium Density Land Development	1705.0000	5449.4000	7565.9000
	Low Density Land Development	678.1000	1971.1000	5657.5000
Continuous	*Weighted Population Density*	706.8000	4164.6000	8883.4000
	Poverty Ratio	0.0244	0.5923	1.6639

Table 5. Quarter Mile Buffer Specific Variables

	Variable	Minimum	Mean	Maximum
Area (1000 ft^{2})	*High Density Land Development*	0	809.5000	2305.6000
	Medium Density Land Development	494.1000	1392.1000	2111.9000
	Low Density Land Development	29.1000	545.4000	1714.7000
Continuous	*Weighted Population Density*	120.2000	1156.9000	2921.7000
	Poverty Ratio	0.0196	0.6380	2.8453

Table 6. Eighth Mile Buffer Specific Variables

	Variable	Minimum	Mean	Maximum
Area (1000 ft^{2})	*High Density Land Development*	0	230.3690	590.9381
	Medium Density Land Development	145.3127	366.3816	639.3763
	Low Density Land Development	0	127.6814	416.5629
Continuous	*Weighted Population Density*	0	38.0000	214.8634
	Poverty Ratio	0.0244	0.0196	0.0196

4. METHODOLOGY

4.1 Model Formulation

The odds of pedestrian compliance with signal phasing, the dependent variable, is a continuous variable which can take any value between zero and positive infinity, which allows for multiple-linear regression analysis. Multiple linear regression allows for an understanding of how the dependent variable is influenced by the independent variables, specifically land use and demographic characteristics. When analyzing the odds of pedestrian compliance with crossing signals, it is apparent that it does not follow a normal distribution **(**
Figure 2**),** which is essential for multiple linear regression. A logarithmic transformation of the odds of compliance is implemented. and **Figure 7** and **Figure 8** show the values to appear to follow a normal distribution. The results from a Shapiro-Wilk yielded a statistic of 0.0567 with a p-value <0.001. Therefore, we fail to reject the null hypothesis that the logarithmic transformation is normally distributed. These tests show less than a 0.1% probability that the data do not follow a normal distribution. The basic model form is as follows (**Equation 3)**:

$$\ln(Y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_m x_{i,m} \quad \textbf{(3)}$$

Where:

$$\beta_m = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \textbf{(4)}$$

Y_i is the response variable, odds of pedestrian compliance at crosswalk location *i*. β_0 is the intercept value when all independent variables are zero. **Equation 4** displays the calculation of the regression coefficients, β_m. Each beta represents the change in the response variable per 1.0 unit change of each $x_{i,m}$, independent variables used in the model.

4.2 Spatial Autocorrelation (Local Moran's I)

The crosswalks used in this study are all located at intersections in Hartford County. Some of these intersections are less than 1000 feet apart from each other, indicating that some crosswalk features could be similar between these adjacent intersections. To account for this, a spatial autocorrelation analysis was performed on the residuals from the model estimations. Developed by Anselin in 1995 [36], Local Moran's I analysis identifies both local positive and local negative spatial autocorrelation. Positive (clustered) spatial autocorrelation is identified if the residuals are high or low clustered around a local point *i*. Negative (dispersed) spatial autocorrelation is also possible and occurs when a local point *i* is surrounded by dissimilar values, i.e., an intersection with positive residuals is surrounded by intersections with negative residuals. Instances of positive or negative spatial autocorrelation indicate the possibility of a geographical influence on pedestrian compliance with signal phasing. A distance-based weight matrix was created with queens contiguity which considers neighboring intersections in all directions. A designated threshold value of 2000 feet was used as it best encapsulated the connectivity between neighboring intersections **(Figure 9)**. The threshold value requires discretion so various distances were tried. The results from the analysis for various buffer sizes yielded insignificant results with no spatial autocorrelation between crosswalks under 0.001 **(Figure 10**), the recommended significance level for Local Moran's I analysis, and a Moran's I value of 0.014 **(Figure 11).**

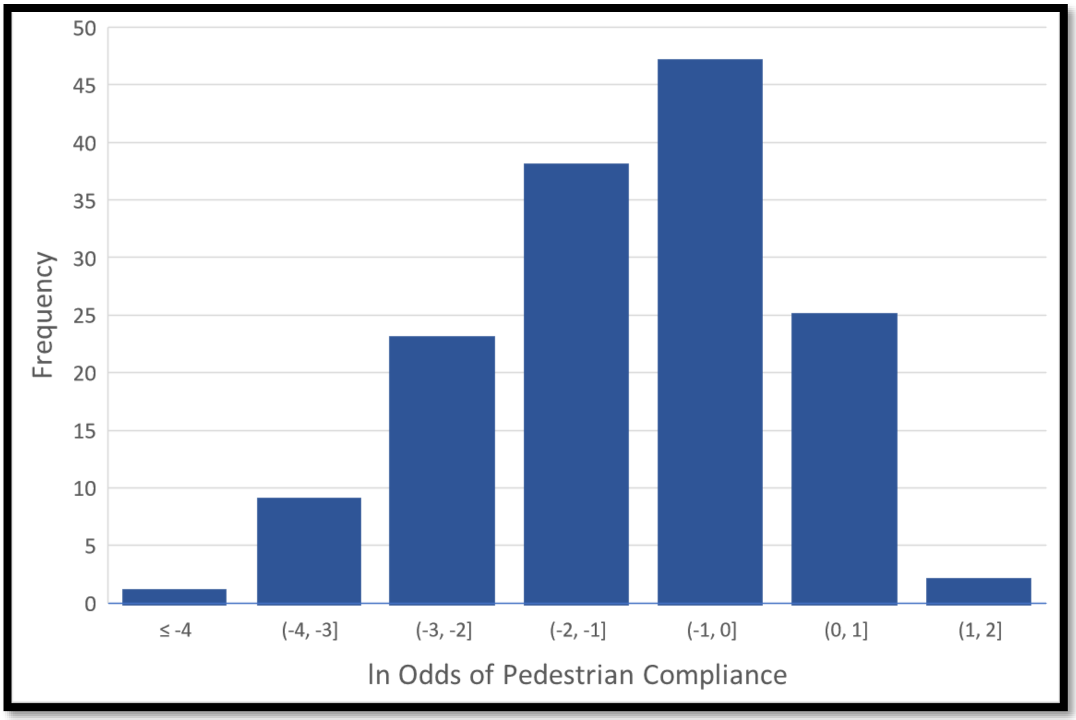

Figure 7. Transformation of the Odds of Pedestrian Compliance

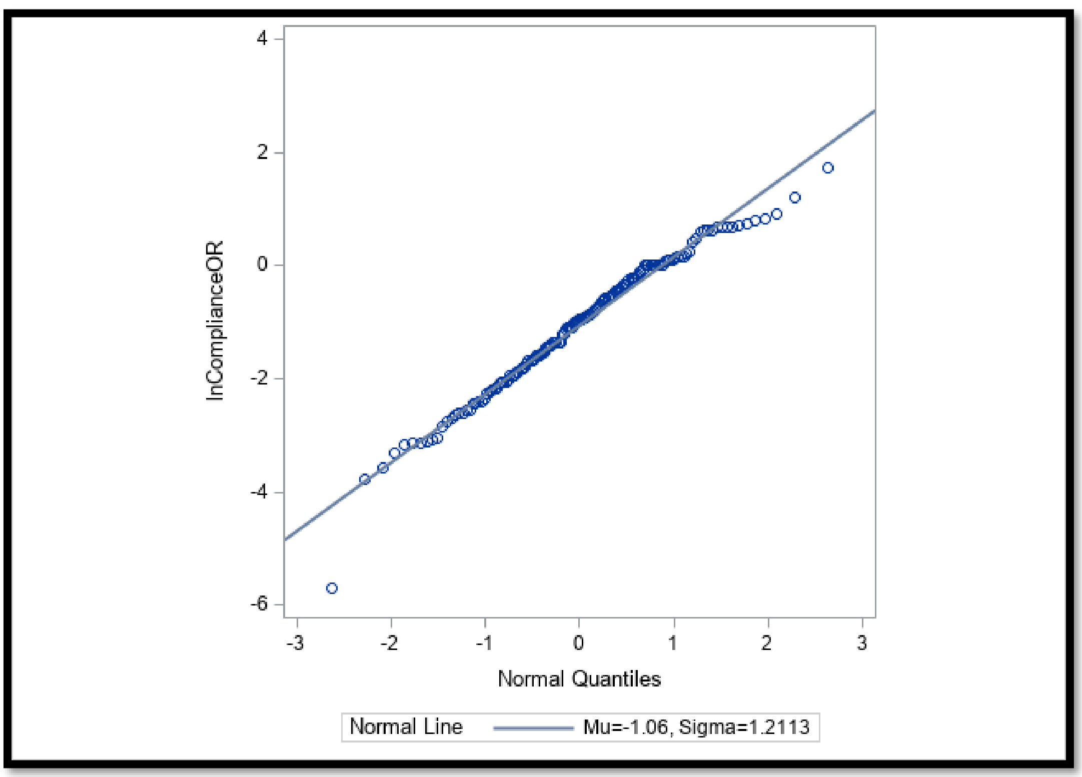

Figure 8. Quartile-Quartile Plot: Logarithmic Transformation of Compliance

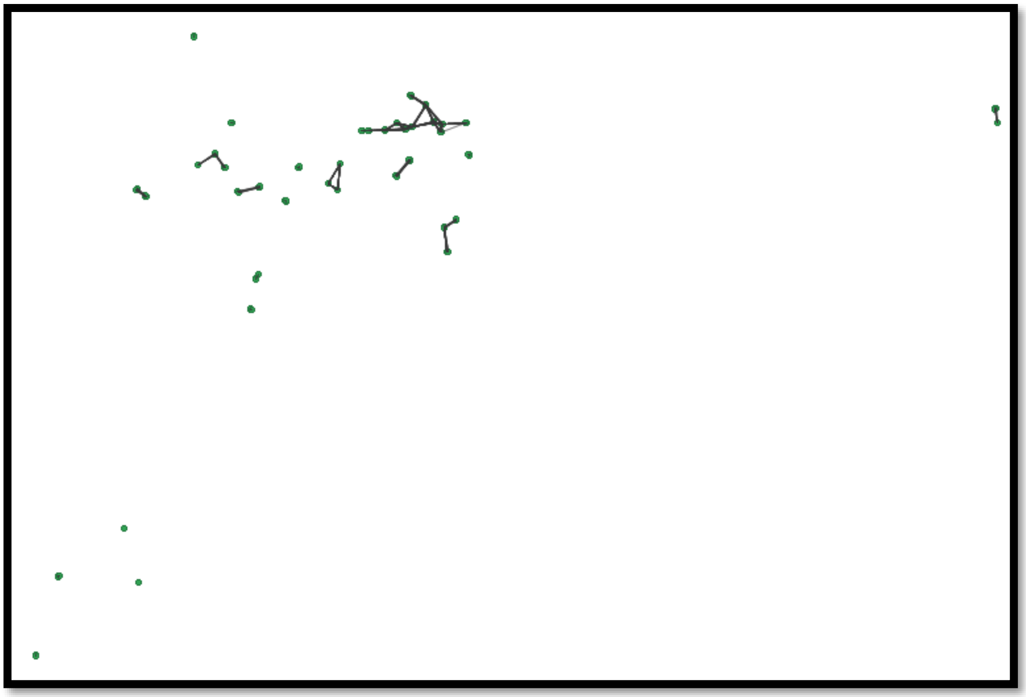

Figure 9. Connectivity Graph, 2000 ft Threshold

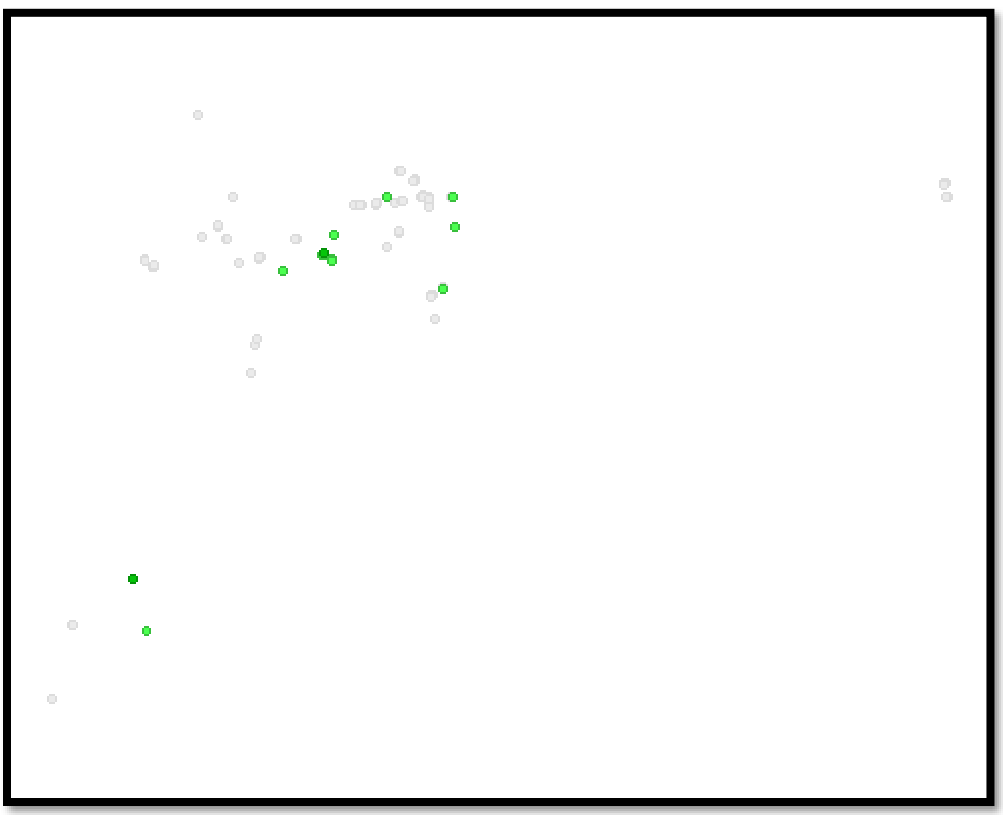

Figure 10. Moran’s I Results: Quarter Mile Buffer

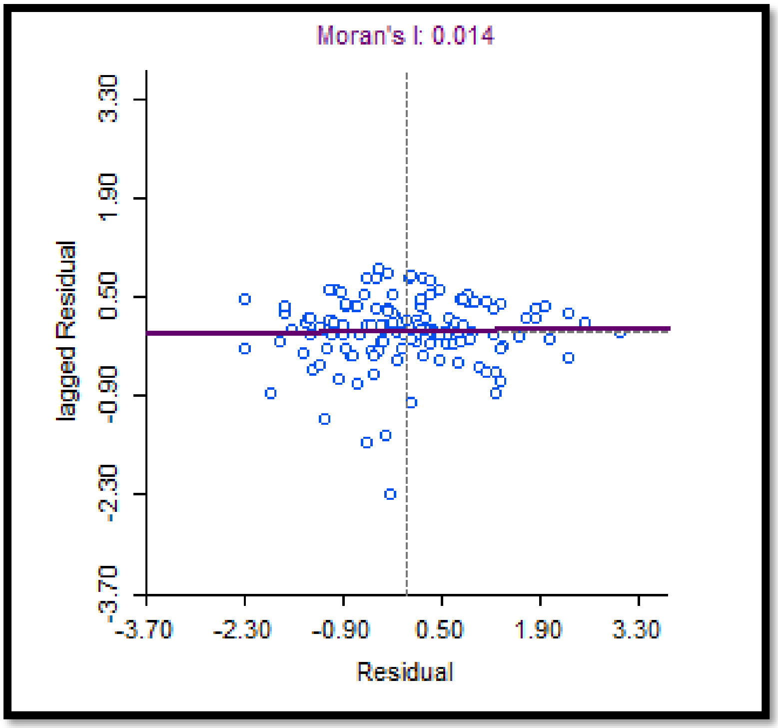

Figure 11. Moran's I Significance Map: Quarter Mile Buffer

4.3 Data Analytics and Random Effects

There are pedestrian observations at 152 crosswalks considered in this study. However, when aggregating the observation data to obtain the odds of pedestrian compliance, there were seven crosswalk locations with the log-odds of compliance being either -999 or positive 999, indicating either entirely noncompliant or perfect compliance. These observations presented as outliers in the regression analysis and significantly influenced the model. Investigating the sites explained why six of the seven crosswalks were outliers. Observations at crosswalk locations 94, 107, 114, 142, and 146 revealed no proper signaling or crosswalk marking, therefore the observations were labeled as non-compliant since there was technically no way to cross legally there. Consequently, these crosswalks were dropped from the study to avoid this bias. Crosswalk location 75 had only two observations, both of which were labeled fully compliant, yet this site also did not have a signal or marked crosswalk and was therefore dropped from the study since the compliance outcomes were erroneous.

The reason for there being only noncompliant observations at crosswalk 63 at intersection 410, located at the intersection of Ridgewood Road and Boulevard in West Hartford **(**

Figure 12**)** was not as clear as the previously mentioned crosswalks. This crosswalk had complete crosswalk markings with pedestrian signaling. To account for the unique pedestrian behavior observed at this location, it was assigned a dummy variable which accounts for variation that was unaccounted for in the regression. We chose to keep this crosswalk in the analysis because, unlike the others we dropped, there is no evidence that these observations were erroneous or improperly identified.

The spatial autocorrelation discussed above showed no correlation between adjacent intersections. However, a few intersections feature crosswalks where residuals are either all positive or all negative, indicating that there is some kind of shared effed which is unaccounted for by other variables in the model. To account for this apparent spatial autocorrelation, intersection dummy variables are introduced to the model for the crosswalks at each of these intersections. The intersection dummy variables which remain significant with 95% confidence ($p<0.05$) are left in the model and those which do not are removed. The updated model format is as follows **(Equation 5)**:

$$\ln(Y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_m x_{i,m} + C_i \quad \textbf{(5)}$$

where C_i represents the random effects at either crosswalk location 63 and/or the intersection j in which crosswalks I are located. Note that for most crosswalks, C_i is equal to 0.

Figure 12. Aerial View of Crosswalk 63

5. RESULTS

There are three different buffer sizes considered in this study, 1/2, 1/4, and 1/8 miles in distance from the intersection location. Therefore, three different model formulations are necessary to determine which buffer yields the best model results and incorporates as much of the land use density and demographic variables as possible. The log-linear modeling with random effects was performed using SAS 9.4. The PROC REG segment was utilized as it allows for correlation analysis, graphic displays of model outputs, detailed estimate outputs, residuals, and nested F testing between different model forms.

The ANOVA table output from the PROC REG program displays the F-statistic which was used to determine whether the variables used in the model were significant. The F-statistic value was considered among different model formulations, as the higher the value, the more significant the variables in the model are. The Adjusted R^2 value was also considered in the model formulation as it represents the "goodness of fit" of the regression model. Those models with the highest F-statistic and Adjusted R^2 were considered the model of best fit for each buffer size.

5.1 Correlation and Collinearity Analysis

As with any statistical model formulation, there is the possibility of interactions between independent variables. To prevent this, a correlation and collinearity analysis between independent and dependent variables was performed. The Fisher's-exact test provided the correlation results between categorical variables and the corresponding significance level. It was found that we fail to reject the null hypothesis (P>0.05) and that there was significant correlation between the following variables: Weather and Day of the Week, Speed Limit and Crosswalk Presence, Speed Limit and Parking, and Sidewalk and Parking **(Table 7)**.

Table 7. Correlation Analysis Between Categorical Variables

Fisher Exact Tests				
	Speed Limit & Sidewalk	Speed Limit & Parking	On Street Parking & Sidewalk	Weather & Day of the Week
Probability	0.0047	0.0007	0.0058	<.0001
Significance of Correlation	0.0133	0.0249	0.0058	<.0001

Interaction variables were created to best incorporate those variables which were correlated. These variables were defined to describe the indicated combination of conditions. These include:

- Sidewalks and Parking: The presence of sidewalk facilities with on street parking
- Cloudy Tuesdays: Cloudy weather on Tuesdays
- Speed Limit and Crosswalk: 25/30/35 mph Speed limit and crosswalk presence

None of the interaction variables above were significant in any of the model formulations.

The Pearson Correlation Coefficient, also known as Pearson's r, was used to analyze the correlation between continuous variables. There was no significant correlation found between any continuous variable used in this study.

5.2 Half-Mile Buffer Results

The half mile buffer was the first buffer analyzed in this study with the belief that it would best incorporate the land use and demographic characteristics due to its size. The first variables included in the model are the land use and demographic characteristics of the buffer, as they are the most important to include in the model. Descriptive statistics on the continuous variables were consulted to determine possible transformations on the independent variables and are represented in the model formulations accordingly.

The following models highlight several models that were estimated. Model 1 features signal type and the land use and demographic variables. Model 2 added the average vehicle volume per hour, crossing distance, and average pedestrian volume per hour **(Table 8)**. Model 3 adds Speed limit and Weather. Finally, Model 4 adds the day of the week, sidewalk, and crosswalk **(**

Parameters	Model 1			Model 2		
	Estimate	t Value	Pr > |t|	Estimate	t Value	Pr > |t|
Intercept	-2.7580	-0.69	0.4929	-2.7220	-0.66	0.5133
Exclusive Signal	-1.2541	-7.83	<.0001	-1.3082	-8.05	<.0001
High Density Area (Millions of Feet^2)	-0.1923	-2.56	0.0116	-0.1933	-2.60	0.0103
Mid Density Area (Millions of Feet^2)	0.2173	2.08	0.0394	0.1869	1.78	0.0779
LOG (Low Density Area)	0.5951	2.12	0.0355	0.4924	1.67	0.0981
LOG (Weighted Population Density)	-0.8721	-3.33	0.0011	-0.7464	-2.74	0.0069
Poverty Ratio	0.5954	2.53	0.0124	0.5202	2.22	0.0281
Crossing Distance (Feet)	-	-	-	0.0165	2.36	0.0199
LOG (Average Hourly Pedestrian Volume)	-	-	-	-0.0249	-0.25	0.8003
Square Root (Average Hourly vehicle Volume)	-	-	-	-0.00008	-0.01	0.9946
Crosswalk 63	-6.2130	-6.68	<.0001	-6.2092	-6.75	<.0001
F Value	18.36			14.24		
Pr > F	<.0001			<.0001		
Adjusted R^2	0.4577			0.4789		

5.3 Quarter Mile Buffer Results

The quarter mile buffer model formulation process was like the formulation of the half mile buffer. As with the previous formulation, the goal is to include as much of the land use and demographic variables as possible. Therefore, they are included in the model first and are kept in the model when they are significant. Model 1 highlights only the land use and demographic variables. Model 2 includes other continuous variables such as crossing distance, the log of the average hourly pedestrian volume, and the square root of the average hourly vehicle volume **(Table 11)**. Model 3 removes the insignificant continuous variables added in Model 2 and adds sidewalk presence and day of the week **(Table 12)**. Model 3 has 11 variables which are all 95% significant. However, at this point poverty ratio falls out of significance. Model 4 features the same variables as Model 3 but with the crosswalk presence variable removed. In this model, poverty ratio is significant. The final model shows the final variable selection with random effect parameters for intersections **(Table 13).**

Table 9**)**. Variables were added or removed in the model formulation to account for correlation found in Section 6.1. The final model represents the model which best fit the data. In this model, all variables which fail to be 95% significant are removed. It includes four dummy parameters to account for over or under predicted crosswalks located at the same intersection **(Table 6).**

The results from the half mile buffer **(Table 10)** indicate that middle and low land use densities as well as poverty ratio are not significant parameters, at 95% confidence, in the model. Exclusive signal phasing, high density land use area, and the log of the weighted population density have negative coefficients indicating that when there is an increase in these parameters, the odds of pedestrian compliance with signaling decreases. The only variable which does increase the odds of pedestrian compliance in this model formulation is crosswalk presence. Nested F tests were performed between final models that had middle density land use, low density land use, and poverty ratio, but they were found to be insignificant in improving the fit of the model.

Intersections 91, 92, 112, and 192 are assigned dummy variables in the final model to account characteristics of the intersection which are unaccounted for by the regression. The crosswalks at these intersections had all positive or all negative residuals indicating the regression model was over or underpredicting pedestrian compliance with signal phasing at these locations. For example, after the regression model was formulated, the residuals for the crosswalks at Intersection 192 in Hartford were all negative. This indicates that the model was under predicting compliance with signal phasing at this location. The introduction of a dummy variable represents the difference between the prediction of the other intersections in relation to Intersection 192.

To interpret the percentage impact of a 1.0 unit change in a given variable on the odds of pedestrian compliance, the coefficient must be exponentiated, subtract one, and multiplied by 100. For example, high density land density area in millions of feet has a coefficient value of -0.1601. For a one-unit decrease in high density land area $(e^{-0.1601} - 1) * 100 = 14.7\%$ change in the odds of pedestrian compliance with signal phasing. For logarithmically transformed variables, such as the log of the weighted population density, the conversion is slightly less complicated. For example, the coefficient value of -0.37 indicates that for every 1% increase in the population density, the odds of pedestrian compliance with crossing signals decreases by 37%. The Adjusted R^2 of 0.57 indicates that 57% of variance in the model is accounted for by the variables used in the model.

Table 8. Preliminary Model Formulation: Half Mile Buffer

Parameters	Model 1			Model 2		
	Estimate	t Value	Pr > |t|	Estimate	t Value	Pr > |t|
Intercept	-2.7580	-0.69	0.4929	-2.7220	-0.66	0.5133
Exclusive Signal	-1.2541	-7.83	<.0001	-1.3082	-8.05	<.0001
High Density Area (Millions of Feet^2)	-0.1923	-2.56	0.0116	-0.1933	-2.60	0.0103
Mid Density Area (Millions of Feet^2)	0.2173	2.08	0.0394	0.1869	1.78	0.0779
LOG (Low Density Area)	0.5951	2.12	0.0355	0.4924	1.67	0.0981
LOG (Weighted Population Density)	-0.8721	-3.33	0.0011	-0.7464	-2.74	0.0069
Poverty Ratio	0.5954	2.53	0.0124	0.5202	2.22	0.0281
Crossing Distance (Feet)	-	-	-	0.0165	2.36	0.0199
LOG (Average Hourly Pedestrian Volume)	-	-	-	-0.0249	-0.25	0.8003
Square Root (Average Hourly vehicle Volume)	-	-	-	-0.00008	-0.01	0.9946
Crosswalk 63	-6.2130	-6.68	<.0001	-6.2092	-6.75	<.0001
F Value	18.36			14.24		
Pr > F	<.0001			<.0001		
Adjusted R^2	0.4577			0.4789		

5.3 Quarter Mile Buffer Results

The quarter mile buffer model formulation process was like the formulation of the half mile buffer. As with the previous formulation, the goal is to include as much of the land use and demographic variables as possible. Therefore, they are included in the model first and are kept in the model when they are significant. Model 1 highlights only the land use and demographic variables. Model 2 includes other continuous variables such as crossing distance, the log of the average hourly pedestrian volume, and the square root of the average hourly vehicle volume **(Table 11)**. Model 3 removes the insignificant continuous variables added in Model 2 and adds sidewalk presence and day of the week **(Table 12)**. Model 3 has 11 variables which are all 95% significant. However, at this point poverty ratio falls out of significance. Model 4 features the same variables as Model 3 but with the crosswalk presence variable removed. In this model, poverty ratio is significant. The final model shows the final variable selection with random effect parameters for intersections **(Table 13).**

Table 9. Preliminary Model Formulation: Half Mile Buffer

Parameters	Model 3			Model 4		
	Estimate	t Value	Pr > |t|	Estimate	t Value	Pr > |t|
Intercept	-2.4235	-0.62	0.5344	-1.7517	-0.46	0.6429
Exclusive Signal	-1.3293	-8.57	<.0001	-1.3203	-9.00	<.0001
High Density Area (Millions of Feet^2)	-0.2053	-2.75	0.0068	-0.2228	-2.80	0.0059
Mid Density Area (Millions of Feet^2)	0.2201	2.17	0.0319	0.1667	1.66	0.0984
LOG (Low Density Area)	0.5082	1.87	0.0635	0.4313	1.61	0.1104
LOG (Weighted Population Density)	-0.8309	-3.17	0.0019	-0.5603	-2.11	0.0370
Poverty Ratio	0.5843	2.53	0.0126	0.2004	0.85	0.3966
Crossing Distance (Feet)	0.0159	2.80	0.0059	0.0095	1.71	0.0901
Sidewalk Presence	-0.5021	-1.43	0.1538	-0.9529	-2.67	0.0085
Crosswalk Presence	0.4799	1.95	0.0527	0.4973	2.14	0.0345
Monday	-	-	-	-0.9130	-3.55	0.0005
Tuesday	-	-	-	-0.5048	-2.45	0.0156
Wednesday	-	-	-	-0.5386	-2.53	0.0127
Friday	-	-	-	-1.0365	-4.06	<.0001
Crosswalk 63	-6.1664	-6.90	<.0001	-5.7534	-6.72	<.0001
F Value	15.49			14.04		
Pr > F	<.0001			<.0001		
Adjusted R^2	0.5016			0.5591		

Table 10. Final Model: Half Mile Buffer

Parameters	Estimate	t Value	Pr > |t|
Intercept	3.0878	3.10	0.0024
Exclusive Signal Phasing	-0.9947	-6.63	<.0001
High Density Area (Millions of Feet^2)	-0.1609	-2.78	0.0062
LOG (Weighted Population Density)	-0.3724	-3.02	0.0030
Crosswalk Presence	0.6011	2.71	0.0077
Monday	-0.5792	-2.49	0.0139
Tuesday	-0.3987	-2.04	0.0438
Wednesday	-0.3696	-1.80	0.0739
Friday	-0.8168	-3.26	0.0014
Crosswalk 63	-5.8001	-7.00	<.0001
Intersection 91	-1.5253	-2.99	0.0033
Intersection 92	-1.1836	-2.50	0.0136
Intersection 112	-1.3894	-3.01	0.0032
Intersection 192	-1.0996	-2.34	0.0206
F Value	16.00		
Pr > F	<.0001		
Adjusted R^2	0.5753		

Table 11. Preliminary Model Formulation: Quarter Mile Buffer

Parameters	Model 1			Model 2		
	Estimate	t Value	Pr > |t|	Estimate	t Value	Pr > |t|
Intercept	-1.1033	-0.58	0.5662	-1.9482	-0.92	0.3582
Exclusive Signal	-1.1884	-7.32	<.0001	-1.2315	-7.55	<.0001
High Density Area (Millions of Feet^2)	-0.4537	-1.89	0.0608	-0.3882	-1.63	0.1065
Mid Density Area (Millions of Feet^2)	0.4916	1.75	0.0824	0.5142	1.84	0.0686
LOG (Low Density Area)	0.3039	1.94	0.0550	0.2837	1.71	0.0900
LOG (Weighted Population Density)	-0.5091	-3.27	0.0014	-0.4432	-2.75	0.0068
Poverty Ratio	0.0585	0.45	0.6561	0.0231	0.18	0.8608
Crossing Distance (Feet)	-	-	-	0.0130	1.87	0.0640
LOG (Average Hourly Pedestrian Volume)	-	-	-	-0.0387	-0.39	0.6965
Square Root (Average Hourly vehicle Volume)	-	-	-	0.0065	0.53	0.5982
Crosswalk 63	-6.0783	-6.47	<.0001	-6.0170	-6.44	<.0001
F Value	17.93			13.67		
Pr > F	<.0001			<.0001		
Adjusted R^2	0.4515			0.4680		

The final model results show that all land use densities and the weighted population density are significant at 95% confidence (Pr> 0.05). Poverty ratio, while on the verge of meeting the significance criteria, failed, and was removed. When poverty ratio was included in the model, and the random effects intersection dummy variables were added, it lowered the Adjusted R^2 to 0.59 and the high-density land use and crosswalk presence variables were no longer significant. Nested F tests were also performed between the full model which had poverty ratio and the reduced (final) model to verify the removal of poverty ratio. The results indicated that the inclusion of the variable did not add to the significance of the model, therefore its removal was justified. This model also features dummy variables for three intersections which have either all negative or positive residuals indicating characteristics which are unaccounted for by the model.

Exclusive pedestrian signal phasing, high density land use area, weighted population density, sidewalk presence, and day of the week all have negative coefficients. This indicates an inverse relationship between the odds of pedestrian compliance and these variables. For example, as the weighted population density and high-density land use area go up, the odds of pedestrian compliance with signal phasing decreases. Middle-density, low density, and crosswalk presence all increase the odds of pedestrian compliance. This means that as middle density and low-density land use area increases, so does the odds of a pedestrian complying to signal phasing.

Table 12. Preliminary Model Formulation: Quarter Mile Buffer

Parameters	Model 3			Model 4		
	Estimate	t Value	Pr > |t|	Estimate	t Value	Pr > |t|
Intercept	-0.4936	-0.28	0.7762	-0.6952	-0.39	0.6957
Exclusive Signal	-1.2098	-8.47	<.0001	-1.1827	-8.11	<.0001
High Density Area (Millions of Feet^2)	-0.5954	-2.56	0.0114	-0.4891	-2.09	0.0388
Mid Density Area (Millions of Feet^2)	0.6488	2.56	0.0118	0.6414	2.47	0.0149
LOG (Low Density Area)	0.2853	2.02	0.0458	0.3218	2.23	0.0273
LOG (Weighted Population Density)	-0.4268	-3.03	0.0030	-0.3745	-2.62	0.0099
Poverty Ratio	-0.2232	-1.72	0.0870	-0.2866	-2.20	0.0298
Sidewalk Presence	-0.8614	-2.51	0.0132	-0.9088	-2.59	0.0106
Crosswalk Presence	0.6646	2.73	0.0073	-	-	-
Monday	-1.0026	-3.89	0.0002	-1.0690	-4.07	<.0001
Tuesday	-0.5322	-2.76	0.0066	-0.6076	-3.11	0.0023
Wednesday	-0.8452	-4.03	<.0001	-0.8591	-4.01	0.0001
Friday	-1.3997	-5.83	<.0001	-1.3746	-5.60	<.0001
Crosswalk 63	-5.3988	-6.39	<.0001	-5.3533	-6.19	<.0001
F Value	16.43			16.39		
Pr > F	<.0001			<.0001		
Adjusted R^2	0.5822			0.5618		

The logarithmic transformation of low-density land use area and weighted population was performed to obtain a normal distribution. The coefficient estimate value for the log of the low-density area of 0.33 indicates that there is a 33% increase in the odds of pedestrian compliance per 1% increase low density land use area. There is a 32% decrease in the odds of pedestrian compliance per 1% increase in weighted population density. Similar methods to that seen in Section 6.1 can be used to interpret the non-transformed continuous variables. For 1.0 unit decrease in high density land area $(e^{-0.56190} - 1) * 100 = 75\%$ change in the odds of pedestrian compliance with signal phasing.

Intersections 70, 71, and 192 are assigned dummy variables in the final model to account characteristics of the intersection which are unaccounted for by the regression. The crosswalks at these intersections had all positive or all negative residuals indicating the regression model was over or underpredicting pedestrian compliance with signal phasing at these locations. For example, after the regression model was formulated, the residuals for the crosswalks at Intersection 71 in West Hartford were all positive. This indicates that the model was over predicting compliance with signal phasing at this location. The introduction of a dummy variable represents the difference between the prediction of the other intersections in relation to Intersection 71.

The Adjusted R^2 of 0.632 signifies that about 63% of all variances in the data can be explained by the model. The F-value of 16 with a significance of <0.001 also shows that the variables used in the model significantly add to the statistical model when compared to the model when all coefficients are equal to 0.

Table 13. Final Model Formulation: Quarter Mile Buffer

Parameters	Estimate	t Value	Pr > |t|
Intercept	-2.5094	1.7366	0.1509
Exclusive Signal Phasing	-0.9823	0.1477	<.0001
High Density Area (Millions of Feet^2)	-0.5619	0.2187	0.0113
Middle Density Area (Millions of Feet^2)	0.6687	0.2437	0.0069
LOG (Low Density Area)	0.3332	0.1352	0.0150
LOG (Weighted Population Density)	-0.3156	0.1344	0.0204
Sidewalk Presence	-0.9093	0.3209	0.0053
Crosswalk Presence	0.8747	0.2316	0.0002
Monday	-0.7821	0.2367	0.0012
Tuesday	-0.5282	0.1850	0.0050
Wednesday	-0.6673	0.1910	0.0007
Friday	-0.9727	0.2345	<.0001
Crosswalk 63	-5.3569	0.7932	<.0001
Intersection 70	1.8973	0.5403	0.0006
Intersection 71	0.8696	0.4133	0.0373
Intersection 192	-1.2878	0.4494	0.0049
F Value		17.50	
Pr > F		<.0001	
Adjusted R^2		0.6322	

5.4 Eighth Mile Buffer Results

The model formulation for the eighth mile buffer began similarly to the process used in the half mile, Section 6.1, and quarter mile, Section 6.2. However, unlike the previous buffers, the eighth mile buffer has a minimum value of zero for high density land use area, low density land use area, weighted population density, and poverty ratio **(Table 6).** The distribution of these variables was also skewed to the right with majority of observations around zero. To remedy this, a logarithmic transformation of variables was used where necessary. Model formulation steps for the eighth mile buffer can be seen in **Table 14** and **Table 15.**

Model 1 features the land use and demographic variables alone. Model 2 features the non-buffer specific continuous variables. It is important to note that only crossing distance and exclusive phasing are significant at 95%. Model 3 adds categorical variables such as crosswalk presence, sidewalk presence, speed limit, and weather. Rainy weather was found to be significant but falls out when day of the week is added to the model. Model 4 adds day of the week and removed variables such as speed limit and weather. The final model **(Table 16)** features the final variable selection with P-values less than 0.05 and intersection dummy variables to account for characteristics unaccounted for in the regression.

Table 14. Preliminary Model Formulation: Eighth Mile Buffer

Parameters	Model 1			Model 2		
	Estimate	**t Value**	**Pr > |t|**	**Estimate**	**t Value**	**Pr > |t|**
Intercept	-1.8310	-0.83	0.4101	-3.7769	-1.62	0.1076
Exclusive Signal	-1.0905	-6.30	<.0001	-1.1621	-7.01	<.0001
High Density Area (Millions of Feet^2)	-2.1520	-2.66	0.0087	-1.4486	-1.81	0.0719
Mid Density Area (Millions of Feet^2)	0.0558	0.08	0.9371	0.7169	1.03	0.3038
LOG (Low Density Area)	0.1769	1.04	0.3017	0.2251	1.34	0.1823
Weighted Population Density	-0.0019	-1.34	0.1838	-0.0016	-1.17	0.2461
LOG (Poverty Ratio)	-0.0267	-0.55	0.5834	-0.0378	-0.82	0.4165
Crossing Distance (Feet)	-	-	-	0.0223	2.95	0.0038
LOG (Average Hourly Pedestrian Volume)	-	-	-	-0.0637	-0.63	0.5282
Square Root (Average Hourly vehicle Volume)	-	-	-	0.0078	0.61	0.5428
Crosswalk 63	-6.2783	-6.47	<.0001	-6.1664	-6.61	<.0001
F Value	15.44			13.79		
Pr > F	<.0001			<.0001		
Adjusted R^2	0.4193			0.4774		

The final model for the eighth mile buffer features only one land use variable, high density land use area, and no demographic variables. Much like the half mile and quarter mile buffer models, the negative coefficient associated with high density land use indicates a decrease in the odds of pedestrian compliance with crossing signals with an increase in high density land use. Exclusive pedestrian phasing and day of the week also have negative coefficients. Crosswalk presence increases pedestrian compliance, which falls in line with the previous buffer model formulations. This model also features dummy variables for three intersections which have either all negative or positive residuals indicating characteristics which are unaccounted for by the model.

Intersections 70, 91, 102, and 192 are assigned dummy variables in the final model to account characteristics of the intersection which are unaccounted for by the regression. The crosswalks at these intersections had all positive or all negative residuals indicating the regression model was over or underpredicting pedestrian compliance with signal phasing at these locations. For example, after the regression model was formulated, the residuals for the crosswalks at Intersection 91 in West Hartford were all negative. This indicates that the model was under predicting compliance with signal phasing at this location. The introduction of a dummy variable represents the difference between the prediction of the other intersections in relation to Intersection 91.

The final Adjusted R^2 of 0.6009 means that about 60% of all variances in the data can be explained by the model. The F-value of 19.07 has a significance <0.001, indicating the model with the variables included is significantly better than the intercept only model where all variables are equal to zero.

Table 15. Preliminary Model Formulation: Eighth Mile Buffer

Parameters	Model 3			Model 4		
	Estimate	t Value	Pr > |t|	Estimate	t Value	Pr > |t|
Intercept	-5.2478	-2.26	0.0258	-3.0792	-1.45	0.1485
Exclusive Signal	-1.0063	-5.45	<.0001	-1.1416	-7.49	<.0001
High Density Area (Millions of Feet^2)	-1.5362	-1.90	0.0602	-2.0124	-2.55	0.0120
Mid Density Area (Millions of Feet^2)	0.9637	1.39	0.1677	0.8738	1.29	0.2000
LOG (Low Density Area)	0.3243	1.86	0.0650	0.2652	1.66	0.0991
LOG (Weighted Population Density)	-0.0015	-1.09	0.2770	-0.0022	-1.75	0.0832
Log (Poverty Ratio)	-0.0383	-0.84	0.4046	-0.0452	-1.05	0.2968
Crossing Distance (Feet)	0.0246	3.65	0.0004	0.0180	2.93	0.0040
Crosswalk Presence	0.2945	1.19	0.2371	0.4390	1.89	0.0614
Sidewalk Presence	-0.4092	-1.11	0.2697	-0.8269	-2.36	0.0199
Monday	-	-	-	-0.8314	-3.29	0.0013
Tuesday	-	-	-	-0.3616	-1.74	0.0835
Wednesday	-	-	-	-0.6776	-3.21	0.0017
Friday	-	-	-	-1.0909	-4.51	<.0001
Speed Limit: 25	0.1413	0.78	0.4357	-	-	-
Speed Limit: 35	0.1548	0.64	0.5219	-	-	-
Cloudy	0.2280	1.06	0.2927	-	-	-
Rainy	1.0147	2.04	0.0436	-	-	-
Crosswalk 63	-6.0415	-6.58	<.0001	-5.6555	-6.56	<.0001
F Value	10.77			13.80		
Pr > F	<.0001			<.0001		
Adjusted R^2	0.4943			0.5614		

Table 16. Final Model Formulation: Eighth Mile Buffer

Parameters	Estimate	t Value	Pr > |t|
Intercept	-0.0928	-0.35	0.7291
Exclusive Signal Phasing	-0.7226	-5.06	<.0001
High Density Area (Millions of Feet^2)	-3.0476	-6.18	<.0001
Crosswalk Presence	0.8379	3.72	0.0003
Monday	-0.4001	-1.71	0.0896
Tuesday	-0.6972	-3.76	0.0003
Wednesday	-0.4867	-2.45	0.0158
Friday	-0.9462	-3.92	0.0001
Crosswalk 63	-5.5061	-6.92	<.0001
Intersection 70	2.0492	4.09	<.0001
Intersection 91	-1.0675	-2.39	0.0182
Intersection 102	-1.7250	-3.53	0.0006
Intersection 192	-1.1723	-2.62	0.0097
F Value	19.07		
Pr > F	<.0001		
Adjusted R^2	0.6009		

6. DISCUSSION

Log-linear regression provided the framework for the model formulation for the half mile, quarter mile, and eighth mile buffers with the goal of accurately predicting the odds of pedestrian compliance with crosswalk phasing using geo-spatial characteristics surrounding the crosswalk locations. Model formulation considered correlation and collinearity between variables, spatial autocorrelation, and shared effects at intersections. Models were created using variable transformations and were selected based on the F-value, adjusted R^2, and a criterion of all parameters in the model achieving a significance <0.05 to be included in the model.

The half mile buffer incorporates high density land use area and weighted population density. The eighth mile buffer model only features high-density land use area. High density land use area was found to be significant among all model formulations. This falls in line with previous studies which suggest pedestrian activity and collisions increase at intersections surrounded by high density "built environments" [11][12]. The inclusion of a poverty ratio variable failed to be significant in any of the three final models. These results fail to replicate the safety disparity found in lower income areas found in previous studies [14][15][16]. This may be because poverty ratio is correlated with the density variables, which were better at capturing the related effects. This is notable, however, for recognizing income disparity in pedestrian safety outcomes. The average hour vehicle volume, average hour pedestrian volume, crossing distance, speed limit, and weather are also not significant to the model despite being significant predictors in previous studies which used the same data to estimate pedestrian compliance [23][25]. This could be due to the aggregation of observations and the inclusion of the land use and demographic characteristics.

The quarter mile buffer provided the best model results with an F-value of 17.5 and the highest adjusted R^2 value of 0.6322. The quarter mile buffer providing the best results falls in line with the study by Miranda-Moreno et al. [11], which used 50-, 150-, 400-, and 600-meter buffers to encapsulate built environment and crosswalk characteristics surrounding the location. In that study, the 400-meter buffer (which is the closest to quarter mile) provided the best results with the belief that it served as a proxy for how characteristics within walking distance of an intersection increase pedestrian collisions.

Crosswalk and sidewalk presence also have an influencing factor on pedestrian compliance with crossing signals. When a crosswalk is present, pedestrians are more likely to be compliant with crossing signals. This could be due to pedestrians being more willing to wait for signals that have a marked crosswalk location, which contradicts the study by Mukherjee et al. that found crosswalk presence decreased the odds of pedestrian compliance with signal phasing [29]. However, sidewalk presence decreases pedestrian compliance. This is most likely caused by pedestrians not having an area of refuge to feel safe prior to crossing the road at locations without sidewalks and therefore cross the road more readily.

Low density land use area increases pedestrian compliance with crossing signals. This could be because vehicle speed and crossing distances increase in low density areas and deter pedestrians from defying crossing signals. As the weighted population density surrounding the crosswalk location increases, the odds of a pedestrian complying to crossing signals decreases. This is as expected and has been found in previous studies [13], whereas population increases, non-compliant behavior to crossing signals also increases. These results could be from numerous factors. Pedestrians may be less inclined to comply to pedestrian crossing signals when there are more pedestrians present. This phenomenon, known as "safety in numbers," is the phenomenon where pedestrians are likely to do what those surrounding them are doing and that they feel more noticeable when crossing with a larger group. Population density may also decrease pedestrian compliance with crossing signals because the vehicle speeds are lower in this area and pedestrians do not feel the same sense of danger as high-speed intersections.

There were instances of outlier intersections in all model formulations. These outlier intersections were introduced to the regression to account for crosswalks with all positive or all negative residuals. There were no instances where these outlier intersections were clustered within proximity to each other. However, Intersection 192 **Figure 13.** Intersection 192**Figure 13)** was found to be an outlier in all model formulations with pedestrian compliance with signal phasing being underpredicted in each model. This indicates that there is something occurring at this intersection influencing pedestrian signal compliance that was not accounted for by any of the other covariates in the models. Further investigation of the observation at this intersection with exclusive phasing shows most pedestrians cross when the parallel vehicle traffic receives a green light indication. This may be because two of the approaches are single lane one-way exits. Pedestrians may be significantly less likely to comply with signal phasing here since interactions with bidirectional crossing traffic is greatly reduced.

Figure 13. Intersection 192 (Franklin Ave and Benton St Hartford, CT, 06114, 41.747208, -72.677095)

Exclusive pedestrian phasing is one of the most significant variables included in all 3 buffer models. This is most likely due to pedestrians being less compliant with crosswalks with exclusive phasing. These results replicate those found in the study by Ivan et al. [25] which also found pedestrians are less likely to be compliant at crosswalks with exclusive phasing. This study attributed pedestrian compliance with concurrent phasing to be higher due to the increase in complexity associated with concurrent intersections. In that study, which used the same observations used in this study, there was an additional model formulation using an alternative definition of pedestrian compliance referred to as "relaxed compliance." Under this definition, pedestrians are considered compliant at exclusive phased intersections if the cross within the designated area and on the parallel vehicle green. These guidelines treat compliance with exclusive phasing identically to compliance with concurrent signaling.

The results from the study using the relaxed definition of compliance showed that when compliance between these two signals is treated equally, the type of significance is no longer important. To replicate these results, the relaxed compliance definition has been applied to the quarter mile buffer and a model has been formulated accordingly **(Table 17).** In the model formulation, the type of signal phasing is no longer significant in the model. Although the adjusted R^2 of 0.48 and F value of 12.26 are lower than the quarter mile buffer with strict compliance, it is important to note that exclusive phasing is no longer significant. One variable which is now significant unlike the previous model formulations is the average hourly vehicle volume. The negative coefficient associated with average hourly vehicle volume infers that as vehicle volume increases, the odds of pedestrian compliance with signal crossing decreases. This is contradictory as one might expect vehicle traffic to motivate pedestrians to be compliant with signal phasing. One explanation could be higher vehicle volume is associated with a denser environment, such as a city where speeds are lower. It is important to note that when the relaxed definition of compliance is applied to all observations, Intersection 192, an outlier in the previous model formulations, is no longer an outlier. As noted above, pedestrians at this intersection were nearly all observed crossing on the green light with parallel traffic, so this is not surprising.

Table 17. Quarter Mile Buffer: RELAXED COMPLIANCE

Parameters	Estimate	t Value	Pr > |t|
Intercept	-3.6768	-3.43	0.0008
Square Root Average Hourly Vehicle Volume	-0.0324	-3.96	0.0001
Mid Density Area (Millions of Feet^2)	0.6140	3.01	0.0032
LOG (Low Density Area)	0.4871	6.07	<.0001
LOG (Weighted Population Density)	-0.5037	-4.58	<.0001
Monday	-0.4548	-2.31	0.0225
Tuesday	-0.3747	-2.33	0.0215
Wednesday	-0.6749	-3.95	0.0001
Friday	-0.8634	-4.03	<.0001
Crosswalk Presence	0.8073	3.78	0.0002
Crosswalk 63	-5.3597	-7.40	<.0001
Intersection 70	1.3117	2.81	0.0058
Intersection 191	1.5227	3.84	0.0002
F Value		12.26	
Pr > F		<.0001	
Adjusted R^2		0.4840	

7. CONCLUSION

Understanding what factors influence pedestrian compliance with crossing signals allows transportation planners to plan how to improve pedestrian compliance. The focus of this study was to incorporate land use and demographic variables in a log-linear statistical model with the goal of predicting the odds of pedestrian compliance with crossing signals. The quarter mile buffer provided the best results by incorporating the high-density land use area, middle density land use area, low density land use area, and weighted population density were included in the final model with a significance of < 0.05, meaning we can say with 95% confidence that these variables significantly increase the accuracy of the model in predicting the odds of pedestrian compliance with crossing signals.

The negative coefficients associated with exclusive signal phasing, high density, sidewalk presence, and day of the week indicate that when these variables are present or increase, the odds of pedestrians complying with crossing signals at crosswalks decrease. High density land use area and weighted population density are typically associated with downtown urban locations which often feature close high-rise buildings that have high capacities. These dense urban environments are the location of many of the observations used in this study. The negative coefficients associated with these variables indicate that pedestrians in dense urban environments are far less likely to comply with signal phasing. This may be an indication that the wait time at these signals should be reduced to better accommodate pedestrian behavior. Many of the intersections with exclusive phasing are in these dense, highly populated urban environments **(** **Figure** 2**)**. The separation of vehicle and pedestrian movement and the increase in wait time associated with exclusive phasing may be a deterrent to pedestrians to comply at crosswalks with exclusive phasing. Therefore, converting from exclusive phasing to concurrent phasing may also increase pedestrian compliance with signal phasing.

The positive coefficients associated with middle density land use area, low density land use area, and crosswalk presence indicate that when these variables are present or increase, pedestrian compliance with crossing signals increase. The average hourly vehicle volumes and average hourly pedestrian volumes failed to make it into any model using the strict compliance definition. This comes as a surprise as one would assume that these variables would highly influence the odds of pedestrian compliance according to previous studies [24] [25] [26] [28]. However, weighted population density may be explaining some of the influence that the average hourly pedestrian volume has on the odds of pedestrian compliance.

Signal type was found to be one of the most influential variables in predicting pedestrian compliance with signal phasing, like the results found by Ivan et al. [26]. However, this may only provide a partial explanation in understanding pedestrian compliance. A better variable to introduce to this study could be the amount of time pedestrians wait to cross at each intersection. This is because pedestrians are probably not influenced by the signal type itself, rather the amount of time waiting to cross. This is indicated by the negative coefficient associated with exclusive phasing for each model formulation. Exclusive phasing typically has longer cycle lengths and longer delays for the movement of traffic [40]. Therefore, pedestrians may be less willing to comply with signal phasing after a certain threshold of time.

It is important to mention that there are some possible shortcomings when it comes to this study which may be crucial to include in future research. Firstly, the observations used in this study are at least eight years old. Pedestrian behavior since then may have changed significantly and the results could differ with the new data. There was an attempt to create the three buffer size models using data recorded from the summer of 2021, yet, due to a lack of observations, the aggregated data yielded inconclusive results which did not add to the study.

Hartford is the location of many of the observations in this study. This may present an issue because much of the traffic in Hartford comes from commuters, meaning there is the potentially a large variation in traffic and pedestrian behavior from weekday to weekend. All observations in this study were collected during the work week, Monday – Friday. Therefore, much of the variation in pedestrian behavior and compliance with signal phasing could be missed.

The next issue also comes with the time in which the data was collected, which was during the summer. This impacts the study because observations only included weather variables such as sunny, cloudy, and rainy. However, the temperature or snow could highly impact the odds of pedestrian compliance with crossing signals. There have been studies which analyze the impact of weather on pedestrian activity considering extreme hot and extreme cold weather [39]. However, no studies that analyze the influence of extreme weather patterns on pedestrian compliance with signal phasing could be found. Therefore, it would be beneficial to include observations with more variation in the weather.

The observations used are all from Hartford County in Connecticut. This presents a possible lack of variation in pedestrian crossing behavior in different locations. For example, pedestrian behavior in Hartford County may be drastically different than pedestrian behavior in another area of the country or the world. Again, there was an attempt to include variation in locations with the inclusion of the pedestrian observation behavior from 2021. Most of the new data had observations from southern Connecticut, specifically the Bridgeport area. The addition of more sites and observations would increase the accuracy of the study through the exposure of more pedestrian observation behaviors.

Observing pedestrian behavior is a relatively subjective topic, even with observer calibration to ensure that observers record behavior as accurately as possible. Human error is always going to be a factor. Although the poverty ratio variable was not a significant inclusion to this study, there is still reason to believe that the inclusion of an income variable may be influential to pedestrian compliance with crossing signals since death rates in lower income locations are noticeably higher than those in higher income areas [14][15][16].

The inclusion of more demographic variables could improve the model because they may offer insight to how different groups of people behave at crosswalks. A variable worth consideration is vehicle ownership in the area as well as a survey which encapsulates a household or individuals' willingness to walk to and from destinations. The theory behind the inclusion of these variables is their ability to represent an individual's perception of walking as a mode of transportation and how they may abide by signal phasing. Therefore, the inclusion of more, newer data, more location variation, and different seasons may add to the significance of future studies.

Pedestrian compliance with crossing signals is an important subject to study in the field of transportation engineering. It can provide insight into how transportation planners can increase compliance and provide safer facilities for those who choose walking as their mode of transportation.

REFERENCES

[1] “Code Section.” *Law Section*, 1959, https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=VEH§ionNum=21955.

[2] *Safety.fhwa.dot.gov*. https://safety.fhwa.dot.gov/ped_bike/docs/designsafety.pdf.

[3] *Pedestrian Walk Light System - Fhwa.dot.gov*. https://fhwa.dot.gov/infrastructure/walk02.cfm.

[4] Weingroff, Richard Weingroff F. “Highway History.” *U.S. Department of Transportation/Federal Highway Administration*, https://www.fhwa.dot.gov/infrastructure/walk.cfm.

[5] Melosi, Martin v. “The Automobile Shapes the City.” *The Automobile Shapes the City: From "Walking Cities" to "Automobile Cities"*, http://www.autolife.umd.umich.edu/Environment/E_Casestudy/E_casestudy3.htm.

[6] “I-84 Hartford Project.” *Goody Clancy*, https://www.goodyclancy.com/projects/i-84-hartford-project/.

[7] “Pedestrians and Car Crashes.” *Injury Facts*, 30 Mar. 2022, https://injuryfacts.nsc.org/motor-vehicle/road-users/pedestrians/#:~:text=According%20to%20NHTSA%20data%2C%20in,dark%20lighting%20conditions%20(80%25).

[8] “Revision 4 January 2021 - Connecticut.” *Https://Portal.ct.gov/-/Media/DOT/Documents/Dtrafficdesign/TrafficControlSignalDesignManual.pdf*, 2009, https://portal.ct.gov/-/media/DOT/documents/dtrafficdesign/TrafficControlSignalDesignManual.pdf.

[9] *Introducing Concurrent Pedestrian Phasing at State Owned ... - Connecticut*. https://portal.ct.gov/-/media/DOT/documents/dtrafficdesign/Concurrent-Pedestrian-Phasing-January-2021.pdf.

[10] Park, Joon H, et al. *PEDESTRIAN CROSSING BEHAVIOR AND COMPLIANCE AT SIGNALIZED INTERSECTIONS*. https://www.diva-portal.org/smash/get/diva2:925046/FULLTEXT01.pdf.

[11] Miranda-Moreno, Luis F., et al. “The Link between Built Environment, Pedestrian Activity and Pedestrian–Vehicle Collision Occurrence at Signalized Intersections.” *Accident Analysis & Prevention*, vol. 43, no. 5, 2011, pp. 1624–1634., https://doi.org/10.1016/j.aap.2011.02.005.

[12] Pulugurtha, Srinivas S., and Sudha R. Repaka. “Assessment of Models to Measure Pedestrian Activity at Signalized Intersections.” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2073, no. 1, 2008, pp. 39–48., https://doi.org/10.3141/2073-05.

[13] Ren, Gang, et al. “Crossing Behaviors of Pedestrians at Signalized Intersections: Observational Study and Survey in China.” *Transportation Research Record: Journal of*

[14] “America's Poor Neighborhoods Plagued by Pedestrian Death.” *Safe Routes Partnership*, 1 Jan. 1970, https://www.saferoutespartnership.org/resources/research/poor-neighborhoods-pedestrian-death.

[15] Lin, Pei-Sung, et al. “Development of Countermeasures to Effectively Improve Pedestrian Safety in Low-Income Areas.” *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 6, no. 2, 2019, pp. 162–174., https://doi.org/10.1016/j.jtte.2019.02.001.

[16] Noland, Robert B., et al. "Do Lower Income Areas Have More Pedestrian Casualties?" *Accident Analysis & Prevention*, vol. 59, 2013, pp. 337–345., https://doi.org/10.1016/j.aap.2013.06.009.
[17] Schulz, Andreas, and Rainer Stiefelhagen. "Video-Based Pedestrian Head Pose Estimation for Risk Assessment." *2012 15th International IEEE Conference on Intelligent Transportation Systems*, 2012, https://doi.org/10.1109/itsc.2012.6338829.
[18] Sisiopiku, V.P, and D Akin. "Pedestrian Behaviors at and Perceptions towards Various Pedestrian Facilities: An Examination Based on Observation and Survey Data." *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 6, no. 4, 2003, pp. 249–274., https://doi.org/10.1016/j.trf.2003.06.001.
[19] Lord, Dominique, et al. "Pedestrian Accidents with Left-Turning Traffic at ..." *Https://Safety.fhwa.dot.gov/ped_bike/Docs/00674.Pdf*, http://pedbikeinfo.org/cms/downloads/PedAccidents_LeftTurn.pdf.
[20] Kattan, Lina, et al. "Pedestrian Scramble Operations." *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2140, no. 1, 2009, pp. 79–84., https://doi.org/10.3141/2140-08.
[21] Bechtel, Allyson K., et al. "Pedestrian Scramble Signal in Chinatown Neighborhood of Oakland, California: An Evaluation." *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1878, no. 1, 2004, pp. 19–26., https://doi.org/10.3141/1878-03.
[22] ZEGEER, CHARLES V, et al. "Effect of Pedestrian Signals and Signal Timing on Pedestrian Accidents."
[23] Zhang, Yaohua, et al. "Safety Effects of Exclusive and Concurrent Signal Phasing for Pedestrian Crossing." *Accident Analysis & Prevention*, vol. 83, 2015, pp. 26–36., https://doi.org/10.1016/j.aap.2015.06.010.
[24] Hydén, C., and L. Linderholm. "The Swedish Traffic-Conflicts Technique." *International Calibration Study of Traffic Conflict Techniques*, 1984, pp. 133–139., https://doi.org/10.1007/978-3-642-82109-7_12.
[25] Ivan, John N., et al. "A Study of Pedestrian Compliance with Traffic Signals for Exclusive and Concurrent Phasing." *Accident Analysis & Prevention*, vol. 98, 2017, pp. 157–166., https://doi.org/10.1016/j.aap.2016.10.003.
[26] Diependaele, Kevin. "Non-Compliance with Pedestrian Traffic Lights in Belgian Cities." *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 67, 2019, pp. 230–241., https://doi.org/10.1016/j.trf.2016.11.017.
[27] Harrell, W. Andrew. "Factors Influencing Pedestrian Cautiousness in Crossing Streets." *The Journal of Social Psychology*, vol. 131, no. 3, 1991, pp. 367–372., https://doi.org/10.1080/00224545.1991.9713863.
[28] Gårder, Per E. "The Impact of Speed and Other Variables on Pedestrian Safety in Maine." *Accident Analysis & Prevention*, vol. 36, no. 4, 2004, pp. 533–542., https://doi.org/10.1016/s0001-4575(03)00059-9.
[29] Mukherjee, Dipanjan, and Sudeshna Mitra. "A Comprehensive Study on Factors Influencing Pedestrian Signal Violation Behaviour: Experience from Kolkata City, India." *Safety Science*, vol. 124, 2020, p. 104610., https://doi.org/10.1016/j.ssci.2020.104610.
[30] Cook, R. J, and D. L Sackett. "The Number Needed to Treat: A Clinically Useful Measure of Treatment Effect." *BMJ*, vol. 310, no. 6977, 1995, pp. 452–454., https://doi.org/10.1136/bmj.310.6977.452.

[31] "Why Statisticians like Odds." *Statistics.com: Data Science, Analytics & Statistics Courses*, https://www.statistics.com/why-statisticians-like-odds/#:~:text=A%20probability%20must%20lie%20between,linear%20model%20can%20be%20fit.
[32] Davies, H. T., et al. "When Can Odds Ratios Mislead?" *BMJ*, vol. 316, no. 7136, 1998, pp. 989–991., https://doi.org/10.1136/bmj.316.7136.989.
[33] Christensen, Ronald R. *Log-Linear Models and Logistic Regression*. Springer, 1997.
[34] Wu, Jianqing, and Hao Xu. "Annual Average Daily Traffic Prediction Model for Minor Roads at Intersections." *Journal of Transportation Engineering, Part A: Systems*, vol. 145, no. 10, 2019, p. 04019041., https://doi.org/10.1061/jtepbs.0000262.
[35] "Modeling Intersection Crash Counts and Traffic Volume - Final Report." *U.S. Department of Transportation/Federal Highway Administration*, 1997, https://www.fhwa.dot.gov/publications/research/safety/98096/ch07.cfm.
[36] Anselin, Luc. "Local Indicators of Spatial Association-Lisa." *Geographical Analysis*, vol. 27, no. 2, 2010, pp. 93–115., https://doi.org/10.1111/j.1538-4632.1995.tb00338.x. [37] Ho, Katherine, and Rebecca Loraamm. "Analyzing Pedestrian-Involved Crash Hotspots in Hillsborough County, Florida." *Southeastern Geographer*, vol. 60, no. 2, 2020, pp. 121–140., https://doi.org/10.1353/sgo.2020.0011.
[38] "Supplemental Material for a Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables." *Psychological Methods*, 2006, https://doi.org/10.1037/1082-989x.11.1.54.supp.
[39] Runa, Ferdousy, and Patrick A. Singleton. "Assessing the Impacts of Weather on Pedestrian Signal Activity at 49 Signalized Intersections in Northern Utah." *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2675, no. 6, 2021, pp. 406–419., https://doi.org/10.1177/0361198121994111.
[40] Asante, Seth, and Andrew Nagle. "1 Introduction." *Pedestrian Signal Phasing Study*, https://www.ctps.org/data/html/studies/bikeped/ped_signal_phase/Literature_Pedestrian_Signal.html.