TRCLC 14-05

November 15, 2016

Development of Decision Support Tools to Assess Pedestrian and Bicycle Safety: Development of Safety Performance Functions

FINAL REPORT

Valerian Kwigizile, Jun-Seok Oh, Ron Van Houten, Keneth Morgan Kwayu

Transportation Research Center for Livable Communities

Western Michigan University

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
TRCLC 14-05	N/A	N/A

4. Title and Subtitle Development of Decision Support Tools to Assess Pedestrian and Bicycle Safety: Development of Safety Performance Functions	**5. Report Date** November 15, 2016 **6. Performing Organization Code** N/A
7. Author(s) Valerian Kwigizile, Jun-Seok Oh, Ron Van Houten, Keneth Morgan Kwayu	**8. Performing Org. Report No.** N/A
9. Performing Organization Name and Address Western Michigan University 1903 West Michigan Avenue Kalamazoo, MI 49008	**10. Work Unit No. (TRAIS)** N/A **11. Contract No.** TRCLC 14-05
12. Sponsoring Agency Name and Address Transportation Research Center for Livable Communities (TRCLC) 1903 W. Michigan Ave., Kalamazoo, MI 49008-5316	**13. Type of Report & Period Covered** Final Report 7/1/2014 - 9/30/2016 **14. Sponsoring Agency Code** N/A

15. Supplementary Notes

16. Abstract

While a number of studies have developed Safety Performance Functions (SPFs) for motorized traffic, there has been a very limited focus on developing SPFs for non-motorized traffic. Lack of exposure measures for pedestrians and bicyclists has been cited often as the obstacle to developing reliable SPFs. This project aimed at developing statewide SPFs for bicyclists and pedestrians at urban intersections. The SPFs developed can be used as a decision supporting tool for planners and practitioners in prioritizing areas for safety improvements. In order to capture crash patterns despite, a carefully designed sampling approach was implemented and improvements on other aspects of modeling were made. Specifically, we implemented structural equation modeling in the efforts to establish proxy measures of non-motorized exposures. SPFs for pedestrians and bicyclists at urban intersections, are recommended.

17. Key Words	18. Distribution Statement
Pedestrians, Bicyclists, Safety Performance Functions	No restrictions. This document is available to the public through the Michigan Department of Transportation.

19. Security Classification - report	20. Security Classification - page	21. No. of Pages	22. Price
Unclassified	Unclassified		N/A

Disclaimer

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the information presented herein. This publication is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. This report does not necessarily reflect the official views or policies of the U.S. government, or the Transportation Research Center for Livable Communities, who assume no liability for the contents or use thereof. This report does not represent standards, specifications, or regulations.

Acknowledgments

This research was funded by the US Department of Transportation through the Transportation Research Center for Livable Communities (TRCLC), a Tier 1 University Transportation Center.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Research Problem

Walking and biking are forms of transportation that offer basic mobility for all people. In totality, walking and bicycling improve quality of life in many ways such as increasing physical activity and active lifestyles which consequently result into health benefits. In communities where walking and biking is encouraged, the number of motor vehicle trips, which are often the cause of air pollution and congestion, are reduced. Walking and biking can also boost local economy by inviting retail merchant to invest in places near homes and working places. In USA, trips that are done by walking and bicycling rose from 9.5 percent in 2001 to 11.9 percent in 2009 (National Household Travel Survey, 2009).

On the other hand, bicyclists and pedestrians are 2.3 and 1.5 times, respectively, more likely be killed in a crash for each trip as compared to vehicle occupants (Beck et al, 2007). Therefore, transportation agencies have several prevailing concerns with respect to pedestrian and bicycle safety. Resource constraints make it imperative that such agencies develop a framework to identify locations that are at highest risk for pedestrian- and bicyclist-involved accidents. Most importantly, the ability to not only develop, but also to evaluate effectiveness of appropriate countermeasures, is crucial for ensuring safety of pedestrians and bicyclists. Against this backdrop, safety performance functions (SPFs) provide a promising approach for quantifying the risk for pedestrian accidents at specific intersections or road segments.

Currently in Michigan, there is no robust SPF reflecting statewide non-motorized safety. The difficulty has been mostly lack of necessary data (especially pedestrian bicyclist counts) for model development. Another set of data which is necessary, but very often difficult to obtain, is the motorized traffic volume, especially for collectors and local roads, where pedestrians and bicyclists are commonly found. These data are essential part of the model as they explain most of variation of the non-motorized crashes occurring at different locations. Therefore a careful sampling plan which capture the randomness of non-motorized crashes and inclusion of reliable proxy exposure measure for pedestrians and bicyclists will help developing the robust statewide SPF for bicyclist and pedestrians. These SPFs can be modified over time as more planning agencies within the state are starting to collect pedestrian and bicycle volumes for planning purposes in their jurisdiction.

1.2 Objective of the project

The main purpose of this project was to develop a methodology for estimating statewide safety performance functions for pedestrians and bicyclists at intersection. A case study for this research was all urban collector and arterial roads intersections in Michigan. Specifically, the methodology addressed the following:

- Proper sampling procedure to establish an unbiased sample size for model development
- Developing proxy measures of pedestrian and bicyclist exposure using data that are readily available at statewide level.
- Assessment of SPF performance using cross-validation technique.

1.3 Project Scope

The SPFs developed focused on urban intersections in Michigan, specifically for collector and arterial roads. Methodology formulated in this research can be used to develop non-motorized SPFs at county level, census tract, census block group and at corridor level, for instance at the road mid-blocks areas. Transferability of the model to other state is possible provided that proper local calibration factors are applied.

2 Literature Review

2.1 Overview

This section presents review of past studies that have focused on different aspects of non-motorized safety as listed below:

- Non-motorized performance measures that use crash data
- Different exposure measures used in past studies to account for level of risk that the pedestrians and bicyclists experience as they interact with other road users.
- The use of structural equation modelling aspects in modeling and explaining factors that are associated with pedestrian and bicycle crashes.

2.2 Non-motorized performance measures

Performance measures for non-motorized safety refers to the factors that can be used to quantify the level of risk that the pedestrians and bicyclists are experiencing for a given roadway environment. Over the past years, different performance measures developed have been relatively simple to complicated ones. Names have been assigned to those performance measures depending on the type of data and the methodology that was used. With regards to data, performance measures have been mainly developed using crash data, behavioral data and safety ratings. This review mainly focused on studies that have used crash data for developing safety performance measures.

Level of risk experienced by pedestrians and bicyclists on given road infrastructure have been often quantified using non-motorized crash data. Crash data are observed incidents and therefore represent the actual facts. However they are rare and random events. As a consequence, it has been a challenge to develop robust modelling approaches to compute the observed variation of non-motorized crashes at a given location.

Using non-motorized crash data, safety performance measures have been developed using two main approaches, as summarized in Table 1:

- Quantifying non-motorized risk by normalizing the crash data with the exposure measure such as pedestrian volume, distance walked and time spent walking.
- Developing models which relate the number of non-motorized crashes with the roadways, demographic, social economic and non-motorized facility characteristics. Performance

measures developed using this approach are commonly referred to as Safety Performance functions (SPFs).

Table 1: Summary of literatures that have used crash data to develop non-motorized safety performance measures

Author	Modelling approach	Model outcome
Schneider et al., 2010	Pedestrian at signalized intersections Use of crash rates to quantify risk(crashes per 10 million pedestrians crossing) Negative binomial regression to identify geometric characteristics that have significant relationship with pedestrian crashes.	Factors associated with increase in pedestrian crashes: • Vehicle volume • Number of pedestrian crossing • Number of right turn movement by vehicles • Non-residential driveways within 50 feet • Commercial properties within 0.1miles • Percentage of young residents (age <18 years) within 0.25miles Factors associated with decrease in pedestrian crashes: • Raised medians
Nordback, K., Marshall, W. E., & Janson, B. N., 2014	Bicyclist SPFs for intersections Negative binomial model using generalized linear model with log link function.	Factors increasing pedestrian crashes: • Bicyclist volume (Annual average daily bicyclist, AADB) • Traffic volume (AADT) Intersections with more than 200 entering cyclists had fewer collision per cyclist. This demonstrated safety in number concept.
Minikel, 2012	Relative collision rate for bicycle facility running parallel to the arterials.	Collision rates on bicycle boulevards are 2-8 times lower than on parallel, adjacent arterial routes. Factors associated with diminished bicyclist safety include: • Vehicle speed and volume • Presence of heavy vehicle
Oh et al., 2013	Poisson regression model-pedestrian intersection SPF. Negative Binomial regression-bicyclist intersection SPF.	Increase in pedestrian crashes at intersection were significantly related with: • Decrease in total number of lanes for minor roads • Increase in average daily traffic approaching to the intersection • Increase in number of bars • Decrease in number of people with graduate degree within ¼ mile Increase in bicycle crashes at intersection were significantly related with increase in: • Number of right turn lane on the major approach • Bicycle volume • Average daily traffic volume • Presence of bus stop • Business land use

Author	Modelling approach	Model outcome
Oh et al., 2013	Negative Binomial method was used for both pedestrian and bicycle midblock SPFs	Increase in pedestrian crashes at the midblock was significantly associated with • Increase in number of access points • Increase in Average Daily Traffic • Increase in pedestrian volume • Decrease in speed limit • Increase in length of the segment Pedestrian crashes in the midblock was significantly associated with: • Increase in bicycle volume • Decrease in speed limit • Increase in number of bus stops • Increased number bike commuters • Presence of bike lane(decreases)
Turner et al, 2011	Generalized Linear model-Poisson or Negative Binomial SPFs were developed by crash type	Increase in crashes was significantly related with the following variables: • Increase in bicycle and vehicle volumes • Decrease in total intersection approach width • Absence of advanced stop boxes • Increase in intersection depth • Decrease in bicycle lane width, curbside lane width • Increase in midblock length
Jonsson, 2013	Generalized Linear model- Negative binomial distribution Non-motorized SPFs for midblock SPF for bike-bike, pedestrian alone using hospital data.	Variables that were significantly increasing crashes include • Segment length • Traffic volume • Different mix of land use
McArthur, A., Savolainen, P., & Gates, T. (2014).	• SPF for child pedestrian at school zone(1mile radius) • Negative binomial distribution	• Census data; average family size, children ages 5 to 14(increase crashes), average parents per household (decrease crashes), median family income $1000 (decrease crashes), population density (increase crashes) and proportion of non-whites households (decrease crashes) • Number of students enrolled (increase crashes) • School located on local roadway (increase crashes)

2.3 Exposure measures for pedestrians and bicyclists

The Federal Highway Administration (FHWA) and the National Highway Traffic Safety Administration (NHTSA) have identified bicyclist and pedestrian exposure among the top four most important research area (Hedlund. J, 2000). Planners and safety advocates have been using crash data alone in assigning risks that are associated with pedestrians and bicyclists at different facilities. This has led to misallocation of effort to improve the non-motorized safety. Better comparison of risk across different facilities and models of transportation could be obtained using non-motorized crashes normalized by either of the following

- Population density
- Number of pedestrians using the facility
- Time spent walking
- Distance walked
- Number of trips
- Other surrogate measures such as number of potential collision etc.

In essence, there is no single measure that is most suitable to represent pedestrian exposure to traffic unless there was continuous monitoring of pedestrian movement at all time. The choice of exposure is dependent upon the intended purpose of the study. For example distance travelled by a pedestrian will be preferred if analyzing the effectiveness of the sidewalks (Greene-Roesel et al, 2010). Different exposure measures are discussed in the section below as they have been used in different studies.

2.3.1 Population data

Population data is mostly represented as population density in a particular geographic unit. It is used with the underlying assumption that crashes between a pedestrian/bicyclist and motor vehicles will likely to occur as number of residents, drivers, bicyclist and pedestrians increase in a given area. It has been widely used as it is readily available from the census data. However, it is recommended not to use this exposure measure, unless it is impractical to obtain other granular measure of exposure, because it is a crude measure of pedestrian exposure and only provide coarse picture of non-motorized safety. Malino (2000) commented on the insensitivity of population density to location specific factors such as changes in travel behaviors of bicyclist and pedestrian. Population exposure also assumes the pedestrian exposure is uniform for a given population and does not account for the number of people who actually walk or bike. Distance and time a pedestrian or bicyclist is exposed to traffic are not taken into consideration.

2.3.2 Pedestrian/bicyclist volume

Pedestrian/bicyclist volume is defined as the number of pedestrian/bicyclist observed in a roadway at a given location in a specified duration. This exposure measure can be incorporated in pedestrian/bicyclist safety studies as hourly count, or it can be annualized to account for the time of the day, day of the week, and month of the year. It can be collected using different ways such as manual count, video data and through other sophisticated technologies such as the use of active and passive infrared, inductive loops, pneumatic tubes and computer visioning. The choice of which method to use for counting will depend on the purpose of the count, required level of accuracy and overall cost.

Statistical models have been developed which relate pedestrian and bicycle volume with geometric characteristics of the road, facility information, socio-economic and demographic factors. Raford et al., (2003) develop a space syntax pedestrian volume modeling tool for the Oakland City in California. The method utilizes data such as connectivity of street grids, population density employment density and pedestrian count at some key locations within the pedestrian grid network. The space syntax software correlate and extrapolate the aforementioned data to estimate pedestrian volume at street level. Nordback et al., (2014) used negative binomial volume model to estimate bicycle hourly count. Independent variables were hourly temperature, parameter to account for working and nonworking days in a year, solar radiation and school days parameter. Available continuous count were used to calibrate the model. Oh et al (2013) developed a model to estimate pedestrian and bicyclist volume as the function of land use, demographic and socio-economic characteristics. Based on the nature of sample data collected, log-linear model and negative binomial model were used for developing pedestrian and bicyclist volume models, respectively. Qin and Ivan (2001) used generalized linear regression model to predict pedestrian volume in rural areas as the function of population density, site characteristics, demographic characteristics, land use characteristics and road characteristics. In addition to pedestrian and bicycle volumes, the interaction of pedestrian volume and traffic volume at intersection has been used in evaluating the risk associated with variety of pedestrian characteristics and behaviors (Davis et al, 1987 and Tobey at al, 1983). The shortcoming of this exposure metric is that it does not account for the time of separation and how close a motorist is from a pedestrian. A situation might happen when a pedestrian was crossing the road at different time when motorist passes and pedestrian might be walking far away from the moving traffic thus reducing the chances of the crash to occur (Molino et al, 2000).

2.3.3 Number of trips

The number of trips made, as an exposure metric, can be obtained from survey data such as the National Household Travel Survey (NHTS), U.S. Census Journey to Work and America Community Survey. It is mostly used to assess the changes in pedestrian behavior across different jurisdictions. Using number of trips as an exposure measure offers flexibility in analysis since trips can be analyzed at individual, household or location level. However, since most of the information is from survey, the reliability of the data is usually questioned as most of the pedestrian and bicyclist trips are underreported in surveys (Schwartz, 2000).

2.3.4 Distance travelled

This is the distance that the pedestrian walks while exposed to vehicular traffic. It is mostly expressed as million-person-miles travelled when analyzed at individual level (Chu, 2003). It can be obtained in aggregated format by summing the pedestrian distances travelled in a given defined area to get total pedestrian miles travelled. Molino et al (2012) estimated annual pedestrian and bicyclist exposure for Washington DC defined as 100million pedestrian/bicyclist miles travelled. Distances that pedestrians and bicyclists travel on the shared facility with motor vehicles and 15 minutes raw count data were used for development of this exposure metric.

Using distance travelled to account for non-motorized exposure also has its own setbacks. Pedestrians are not exposed to traffic every time when walking. Aggregating distance travelled by pedestrian/bicyclist in a certain geographical unit might overestimate the actual level of exposure that the pedestrians are experiencing. It does not account for the difference in speed among individuals who are walking, which could moderate the individual risk level to traffic. One mile of walking represents greater effect than one mile of a person riding on a passenger car because of difference in travelling speed between these two travelling modes (Chu, 2003).

2.3.5 *Time spent walking*

This is defined as the time taken by the pedestrian when walking while exposed to vehicular traffic. This exposure metric has been used in comparing pedestrian risks across different transportation modes and in different social groups based on age and sex (Greene-Roesel et al, 2010). Other useful application can be on quantifying the risk that pedestrians are facing while crossing the intersection. Knoblauch et al (1996) suggest that time spent crossing at intersection can be a better representation of exposure than a volume count because it takes into account pedestrian age, gender, weather condition, compliance with signal control and signal length. However, it is difficult to obtain this granular data when dealing with large geographical area. Always cost constrain is an impediment for collecting this exposure metric.

2.4 The Use of Structural Equation Modeling in Traffic Safety Studies

Structural Equation Modeling (SEM) is a multivariate modeling technique that has been applied for creating and testing the causal models. It is a combination of confirmatory factor analysis, path analysis and regression analysis. In most cases, SEM is used as confirmatory tool that test the theory the researcher has hypothesized during model construction. For this reason, the researcher has to establish the causality between different variables involved in the model. SEM will then test how well the sample used by the researcher support the model specification. Schumacker et al (2004) provide a good introduction to structural equation modeling for beginners. Step by step procedures are elaborated on how to develop the structural equation model including model specification and identification, model fit, model estimation, testing and assessing the goodness of fit. It is not the aim of this research to explore in details such steps. Rather the goal is to leverage the benefits of SEM in developing tools for assessing non-motorized safety.

SEM has been widely used due to its ability to model complex phenomena, incorporate latent variables in the model and advance in statistical software with minimal coding efforts. Latent construct can be estimated in the model as a function of measurable variables.

Endogeneity effect among variables is explicitly accounted for in the process of explaining complex phenomenon between variables using SEM. Endogeneity exists when there is a loop of causality between variables. In traffic safety studies that involve modeling of crash frequency, often times researchers have been getting results which can be easily judged as counterintuitive. A good example was the one provided by Jonsson (2005) whereby road with low speed limit were highly associated with high non-motorized crashes as compared to high speed limit roads. This can be due to high pedestrian levels in those low speed roads, which in turn increases the conflicts between vehicular movements and pedestrians. Thus more chances for collision to occur

2.5 Examples of Structural Equation Modeling Applications in Traffic Safety

2.5.1 Modeling of Crashes

Wang, K., & Qin, X. (2014) used SEM to model severity of single vehicle crashes. Force and speed were introduced as the latent variables which in turn were hypothesized to influence the crash severity. Manifest variables that were used to measure force that include the type of object that was hit by the vehicle. Speed as latent construct was estimated using roadway, weather and lighting condition, gender and age. By using this model technique, it was possible to explain some of the relationship that could not be revealed using normal ordinal models. Inclement weather, poor lighting condition, and poor pavement surface condition were found to reduce speed (latent variable), which in turn reduced the injury severity.

Initially, SEM was designed for continuous variables whereby the estimation was done in a sample variance-covariance matrix. Therefore it was impossible at that time to incorporate other data format such as nominal, ordinal and intervals. Overtime SEM has been modified to handle the aforementioned data format but introducing a link function which defines the type of data used. Application of this type of modification can be found in the study done by Xie et al (2016). They estimated the effect of secondary collision on injury severity levels using SEM. Injury severity is ordinal in nature and therefore had to be specified in the model. SEM results were compared with those from ordered probit model. The ordered probit model tends to overestimate the safety effect of confounding variables by lumping their direct and indirect effects. By using SEM, it was possible to separate direct and indirect effects of confounding variables that were related directly to crash severity and occurrences of secondary collisions respectively.

2.5.2 Modeling of Road User Travelling Behavior and Mobility

A study conducted by Kim (2003) used SEM to determine factors that were significantly associated with elderly mobility. Urban form was used as the latent construct estimated by retail employment density, population density, age, gender and household size. Likewise, mobility was measured by non-home activity time, travel time and travel distance of elderly persons. SEM was used to unveil how urban form affect mobility of elder drivers. With the use of SEM age and gender showed to have significant effect on older driver mobility. Whereby older women had less mobility than older men and it's more likely for a person to refrain his or her desire for travelling as the age increases.

Ranaiefar et al (2016) estimated bicycle ridership using SEM as a function of different demographic and environmental characteristics surrounding the bike sharing stations. By using SEM, it was possible to forecast origin-destination bike share ridership.

3 Site selection

3.1 Sampling Strategy and Preliminary Data Collection

Many sampling strategies can be used in selecting a sample size from a population. They range from crude sampling procedures such as random sampling which doesn't take into account the sampling error, to more sophisticated sampling techniques such as stratified random sampling.

For this study, a sampling strategy designed to generate a true representation of statewide conditions was designed and implemented. Details of the factors that were used to obtain the sample size and selection technique were adopted from the procedure developed by Aggarwal (1988). Figure 1 below summarize the sampling strategy process.

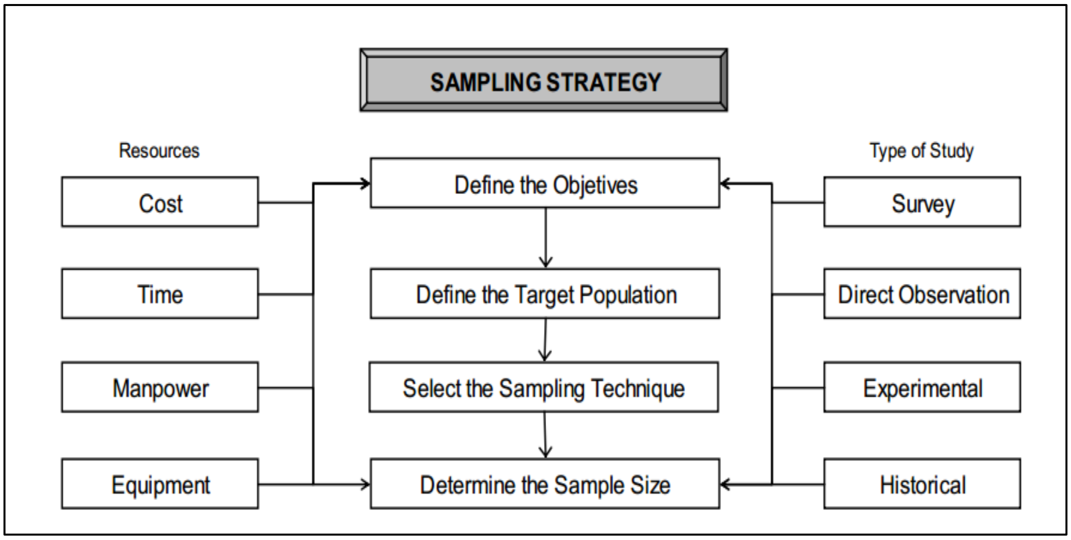


Figure 1 Sampling Strategy Process

In order to develop sample size and sampling technique, available resources in terms of cost, time, manpower and equipment have to be evaluated. This should go concurrently with the proper understanding of the type of study that will be carried out to achieve the project objectives. Upon consideration of all factors, stratified random sampling was selected as sampling technique for the study. The following section explains how stratified random sampling technique was utilized, selection of sample size and finally descriptive statistics of crashes that occurred at urban intersection are discussed.

The choice of urban intersections as the target group was driven by data availability such as average annual daily traffic (AADT) and high number of non-motorized crashes in urban areas as compared to rural areas. Also there was no statewide non-motorized safety performance function developed for urban intersections in Michigan. The urban intersections that were included involve those joining arterial and/or collector road segments. ArcGIS was used as the tool for identifying all urban intersections in Michigan so that the sample could be drawn from it. Sampling procedure, using ArcGIS is summarized below in a concise manner.

3.1.1 Identifying Collector and Arterial Road Intersections

Michigan road shapefile, which provide the statewide road network was used in ArcGIS to identify all road intersection points. As mentioned earlier, only intersections connecting arterial and collector road segments were selected. Therefore based on Road Functional Classification (NFC), three groups of intersections were identified which were Arterial-Arterial intersections, Arterial-Collector intersections and Collector-Collector intersections. The Figure 2 below provide an example of these intersection types as identified in Michigan road shapefile. Over eleven thousands urban intersections were identified.

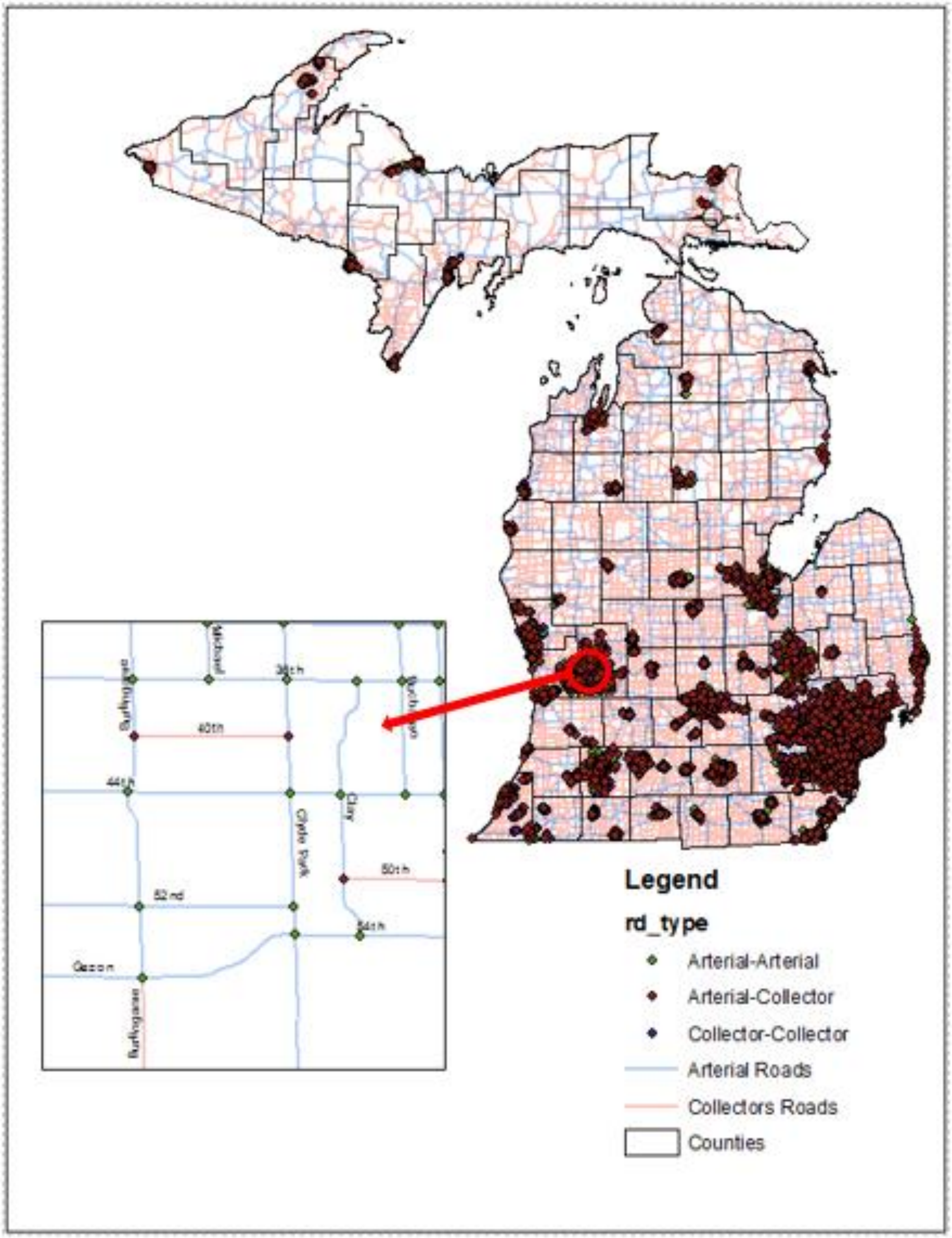


Figure 2 Distribution of Urban Intersections in Michigan

3.1.2 *Subdividing the Target Population into Subgroup*

Stratified random sampling require the target population to be subdivided into groups each having similar characteristics. To achieve this goal, parameters that were available at statewide level were used as shown in Table 2 below.

Table 2 List of Parameters and Subcategories.

Parameters	**Subcategory**
Road function	Intersection connecting arterial roads
	Intersection connecting arterial road and collector road
	Intersection connecting collector roads.
Intersection type	Three leg intersection
	Four leg intersection
Urban population	5000-49,999
	50,000-199,9999
	200,000-more
Non-motorized crashes: Pedestrians and Bicyclists crashes(2010-2014)	No crash observed
	1-5 crashes
	6-10 crashes
	11-16 crashes

Based on subcategory for each parameter, seventy-two groups were created and each of the eleven thousand intersections was placed to its corresponding group.

3.1.3 *Sample Size Computation*

The decision on the total sample size was based on the available resources such as time frame and manpower for data collection and cost associated with obtaining the data. In order to determine the number of intersections each of strata will contribute to the total sample size, a weighting factor was used. Formula to computing weighting factor and sample size for each of strata is shown below.

$$w_i = \frac{N_i}{N_{tot}}$$

$$S_i = w_i * N$$

Whereby

w_i = Weighted factor for intersections in group i

N_i = Number of intersections in group i

N_{tot} = Total number of intersections for all groups

S_i = Number of intersections drawn from group i

N = Required total sample size from all groups

Table 3 to Table 5 provide the output of the sampling process using stratified random sampling. Weighted factors were computed based on all urban intersections in Michigan joining collector and arterial roads.

Table 3 Sampling Process for Arterial-Arterial Intersections

Intersection type	no/Wi	Urban population	no/Wi	no	Non-motorized Crashes	No.	Weight (Wi)	Sample size (N)	Sample size(Si) Wi xN
3 leg	1890 0.169	5000-49,999	262 0.023	1	0	228	0.0203	500	11
				2	1-5	34	0.0030	500	2
				3	6-10	0	0.0000	500	0
				4	11-16	0	0.0000	500	0
		50,000-199,9999	337 0.030	1	0	285	0.0254	500	13
				2	1-5	51	0.0045	500	3
				3	6-10	1	0.0001	500	1
				4	11-16	0	0.0000	500	0
		200,000-more	1291 0.115	1	0	1072	0.0956	500	48
				2	1-5	219	0.0195	500	10
				3	6-10	0	0.0000	500	0
				4	11-16	0	0.0000	500	0
4 leg	3273 0.292	5000-49,999	448 0.0400	1	0	293	0.0261	500	14
				2	1-5	149	0.0133	500	7
				3	6-10	6	0.0005	500	1
				4	11-16	0	0.0000	500	0
		50,000-199,9999	702 0.063	1	0	401	0.0358	500	18
				2	1-5	293	0.0261	500	14
				3	6-10	8	0.0007	500	1
				4	11-16	0	0.0000	500	0
		200,000-more	2123 0.189	1	0	1034	0.0922	500	47
				2	1-5	1019	0.0909	500	46
				3	6-10	64	0.0057	500	3
				4	11-16	6	0.0005	500	1

Table 4 Sampling Process for Arterial-Collector Intersection.

Intersection type	no/Wi	Urban population	no/Wi	no	Non-motorized Crashes	No.	Weight (Wi)	Sample size (N)	Sample size(Si) Wi xN
3 leg	1740 0.168	5000-49,999	504 0.049	1	0	430	0.0415	500	21
				2	1-5	74	0.0071	500	4
				3	6-10	0	0.0000	500	0
				4	11-16	0	0.0000	500	0
		50,000-199,9999	340 0.033	1	0	286	0.0276	500	14
				2	1-5	54	0.0052	500	3
				3	6-10	0	0.0000	500	0
				4	11-16	0	0.0000	500	0
		200,000-more	896 0.086	1	0	648	0.0625	500	32
				2	1-5	239	0.0231	500	12
				3	6-10	9	0.0009	500	1
				4	11-16	0	0.0000	500	0
4 leg	2576 0.249	5000-49,999	685 0.0661	1	0	516	0.0498	500	25
				2	1-5	169	0.0163	500	9
				3	6-10	0	0.0000	500	0
				4	11-16	0	0.0000	500	0
		50,000-199,9999	522 0.0503	1	0	365	0.0352	500	18
				2	1-5	155	0.0150	500	8
				3	6-10	2	0.0002	500	1
				4	11-16	0	0.0000	500	0
		200,000-more	1369 0.132	1	0	840	0.0810	500	41
				2	1-5	509	0.0491	500	25
				3	6-10	18	0.0017	500	1
				4	11-16	2	0.0002	500	1

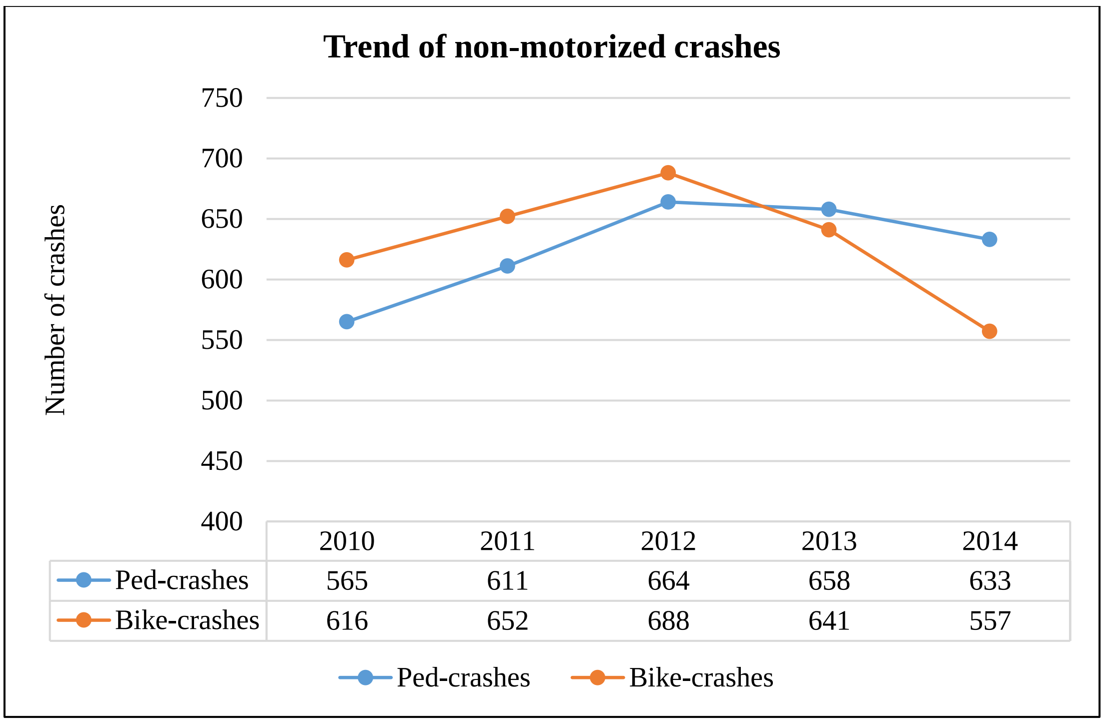
Table 5 Sampling Process for Collector-Collector Intersections

Intersection type	no/Wi	Urban population	no/Wi	no	Non-motorized Crashes	No.	Weight (Wi)	Sample size (N)	Sample size(Si) Wi xN
3 leg	764 0.074	5000-49,999	240 0.0232	1	0	229	0.0221	500	12
				2	1-5	11	0.0011	500	1
				3	6-10	0	0.0000	500	0
				4	11-16	0	0.0000	500	0
		50,000-199,9999	111 0.0107	1	0	99	0.0096	500	5
				2	1-5	12	0.0012	500	1
				3	6-10	0	0.0000	500	0
				4	11-16	0	0.0000	500	0
		200,000-more	413 0.040	1	0	370	0.0357	500	18
				2	1-5	43	0.0041	500	3
				3	6-10	0	0.0000	500	0
				4	11-16	0	0.0000	500	0
4 leg	970 0.094	5000-49,999	382 0.037	1	0	333	0.0321	500	17
				2	1-5	49	0.0047	500	3
				3	6-10	0	0.0000	500	0
				4	11-16	0	0.0000	500	0
		50,000-199,9999	149 0.014	1	0	132	0.0127	500	7
				2	1-5	17	0.0016	500	1
				3	6-10	0	0.0000	500	0
				4	11-16	0	0.0000	500	0
		200,000-more	439 0.042	1	0	352	0.0340	500	17
				2	1-5	85	0.0082	500	5
				3	6-10	2	0.0002	500	1
				4	11-16	0	0.0000	500	0

3.2 Descriptive Statistics of Crash Data Collected During Sampling Process

3.2.1 Trend of Pedestrian and Bicycle Crashes at Urban Intersections 2010-2014

Figure 3 below depicts the distribution of pedestrian and bicyclist involved crashes at urban intersections from 2010 through 2014. Pedestrian involved crashes had an upward trend up to 2012 and there was a drop in number of crashes up to 2014. Similar trend can be observed on bicyclist involved crashes. A more realistic trend could have been observed if non-motorized crashes at those intersections were normalized by number of non-motorized volume. However, such data was not available at statewide level.



	2010	2011	2012	2013	2014
Ped-crashes	565	611	664	658	633
Bike-crashes	616	652	688	641	557

Figure 3 Trend Of non-motorized Crashes from 2010-2014 at intersections

3.2.2 Distribution of Non-Motorized Crashes by Injury Severity Level

In total, there were 106 fatal pedestrian-involved crashes and 17 bicyclist-involved crashes from 2010 to 2014 at all urban intersections in Michigan connecting collector and arterial roads as indicated in Figure 4 and Figure 5. For pedestrians, it represented 3.4 percent of all pedestrian crashes occurred at urban intersections, while for bicyclist it represented 0.5 percent of all bicyclist crashes occurred at urban intersections. Based on these statistics, it is evident that pedestrians are

more likely to be involved in fatal crashes as compared to bicyclist in such locations. Upon looking on the fatal crashes distribution by intersection roadway functional type, it was found that most of these fatal crashes occurred at intersection joining two arterial roads as shown in
Figure 6 and Figure 7. High speed associated with such arterial roads is likely to exacerbate the severity of crash once it happens.

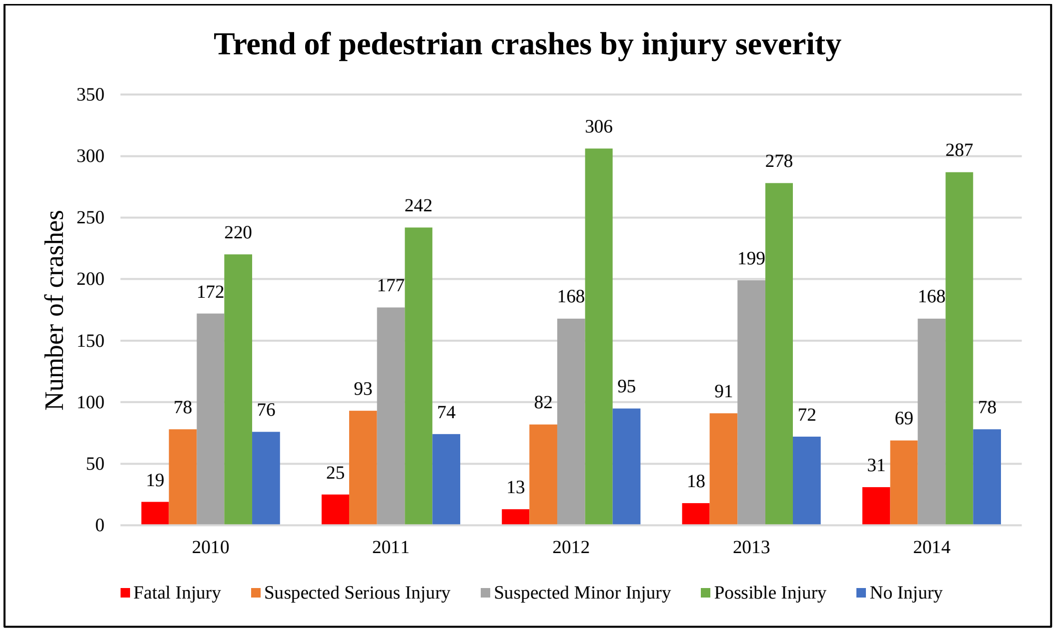


Figure 4 Trend of Pedestrian Crashes by Severity: 2010-2014

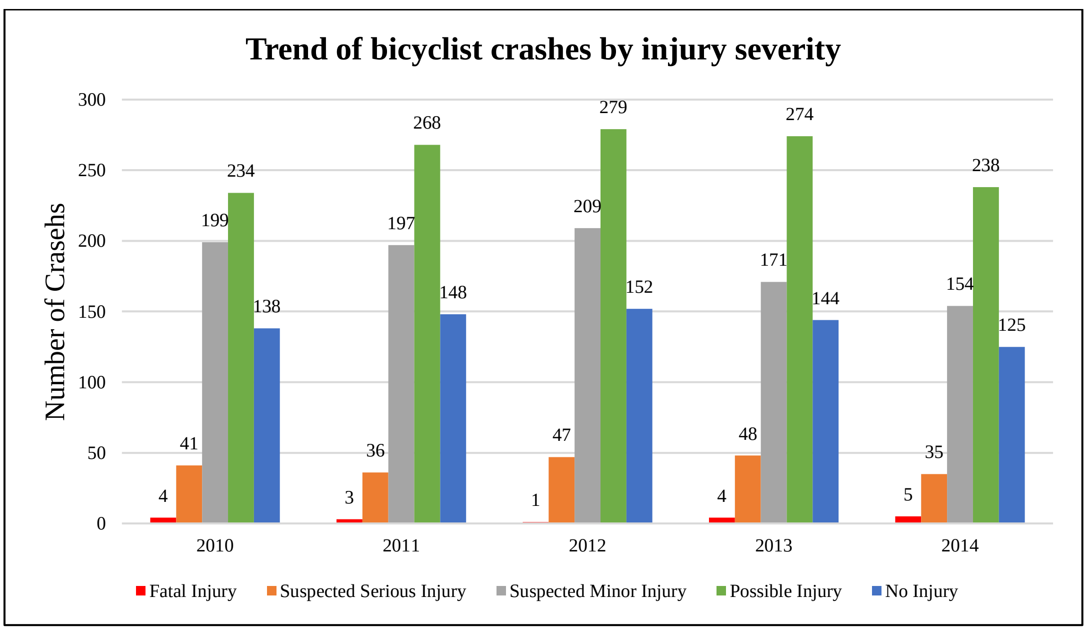


Figure 5 Trend of Bicycle Crashes by Severity: 2010-2014

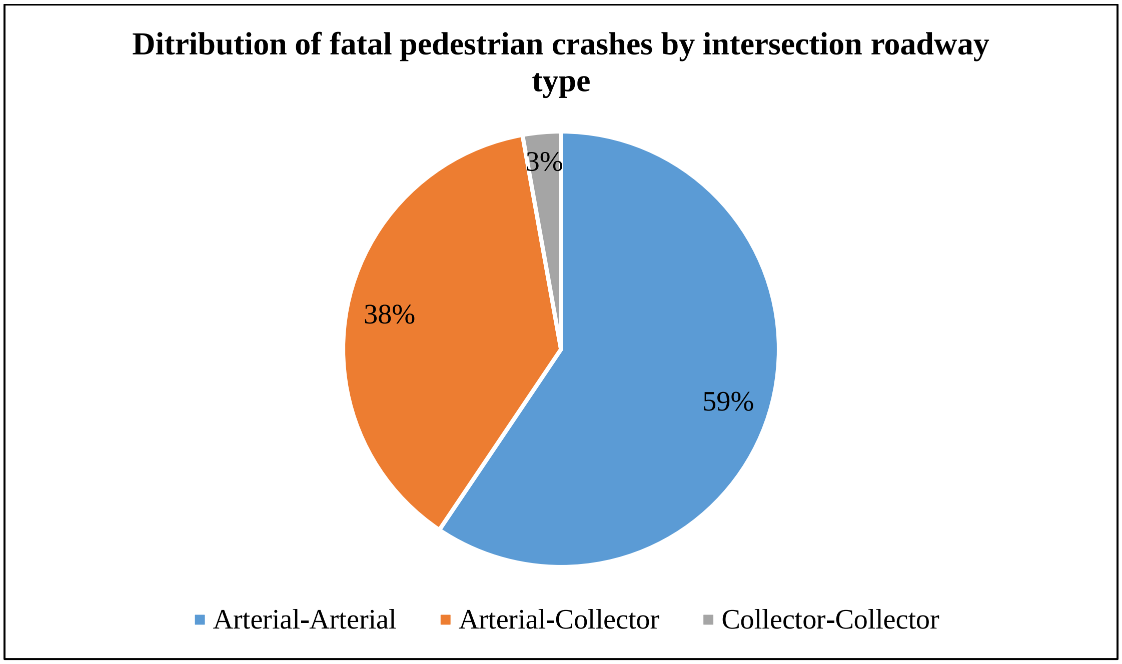


Figure 6 Distribution of Pedestrian Fatal Crashes by Intersection Roadway Types

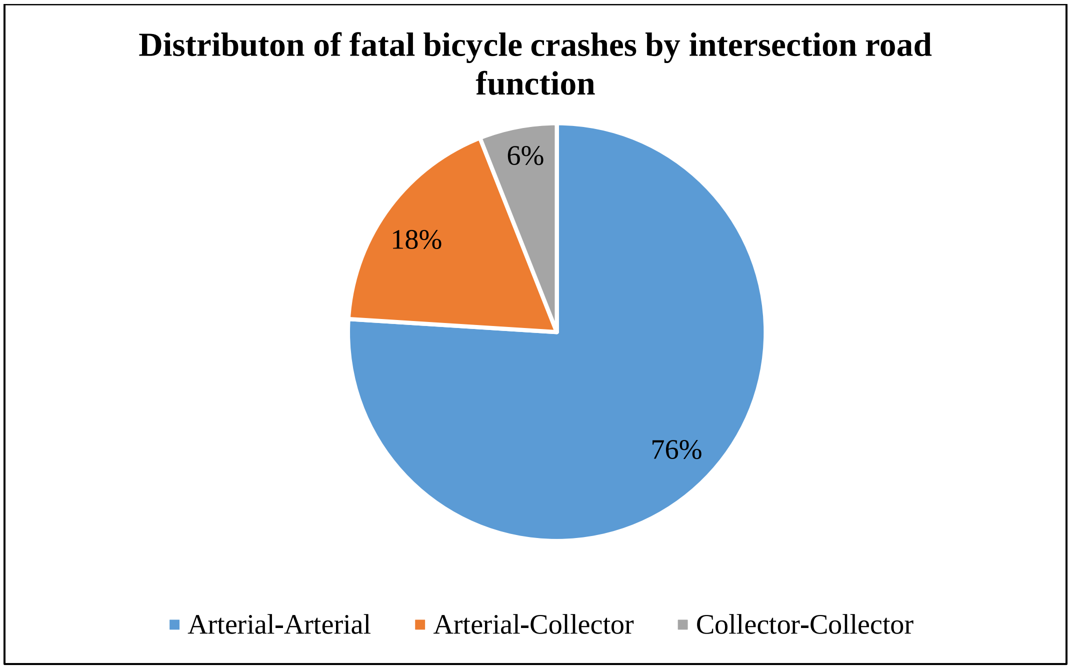


Figure 7 Distribution of Bicyclist Fatal Crashes at Urban Intersection

3.2.3 Distribution of Non-Motorized Crashes at Urban Intersection: Comparison with Other Locations

In five years, pedestrian crashes at urban intersections constituted 27.8 percent of all pedestrian related crashes. For bicyclist-involved crashes, the percentage was 33.5 percent of all bicyclist-involved crashes. More number of bicyclist were involved in crashes as compared to pedestrians at the urban intersection as shown in Figure 8 and Figure 9.

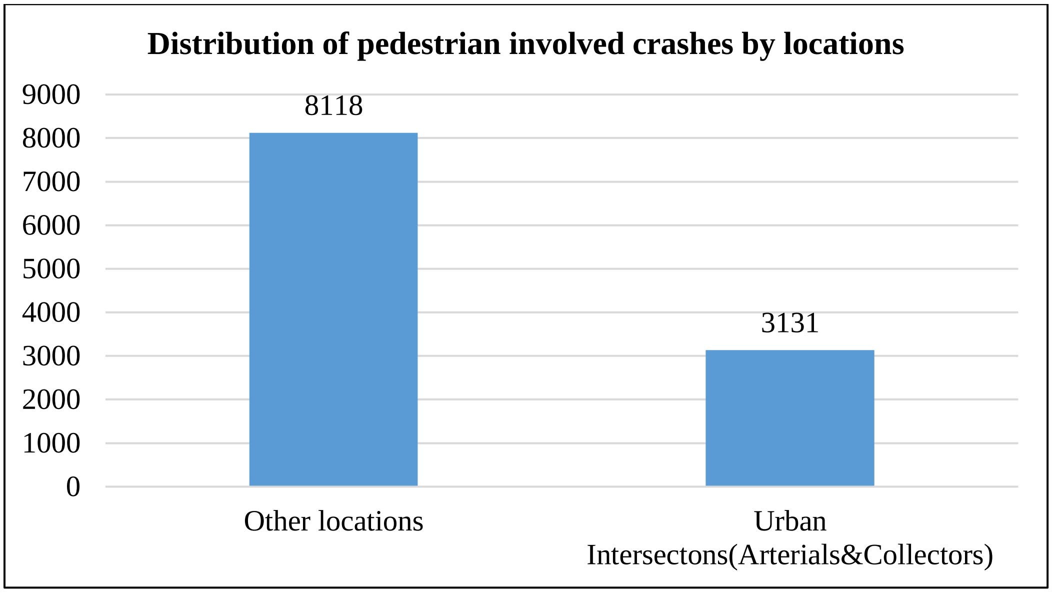


Figure 8 Distribution of Pedestrian Crashes by Location

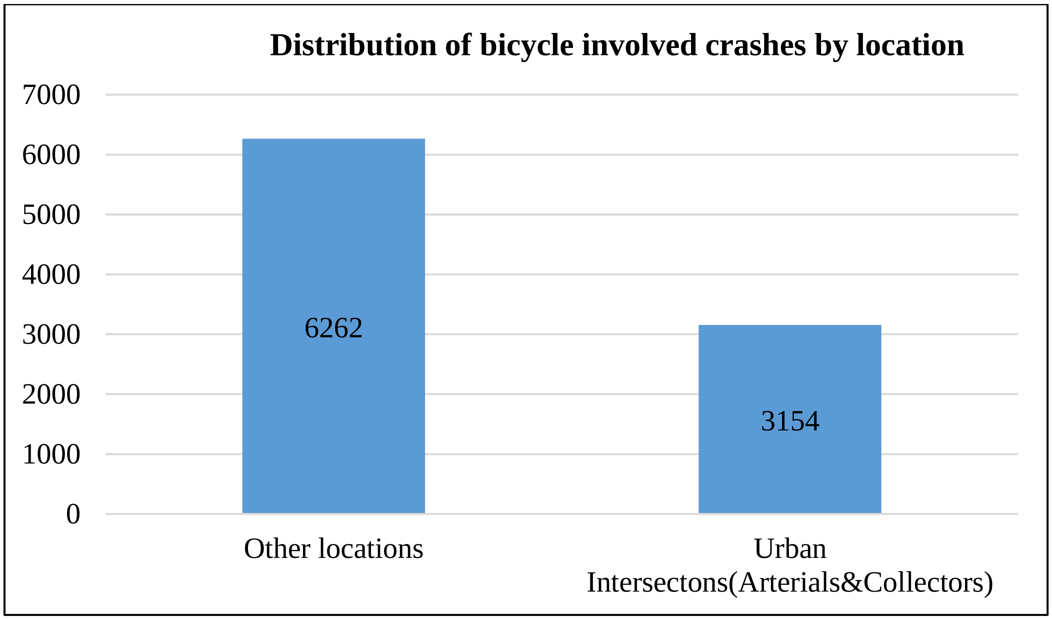


Figure 9 Distribution of Bicycle Crashes by Location

Figure 10 and Figure 11 show the distribution of pedestrian and bicycle involved crashes by roadway type and urban population. For both cases, nearly half of all crashes occurred at intersections joining two arterial roads located in areas with urban population greater than 200,000 people. Densely populated areas are more likely to have high pedestrian and bicyclist movements. The presence of arterial roads in such locations, which are usually characterized by high volume of traffic, likely increases the chances of non-motorized crashes occurrence.

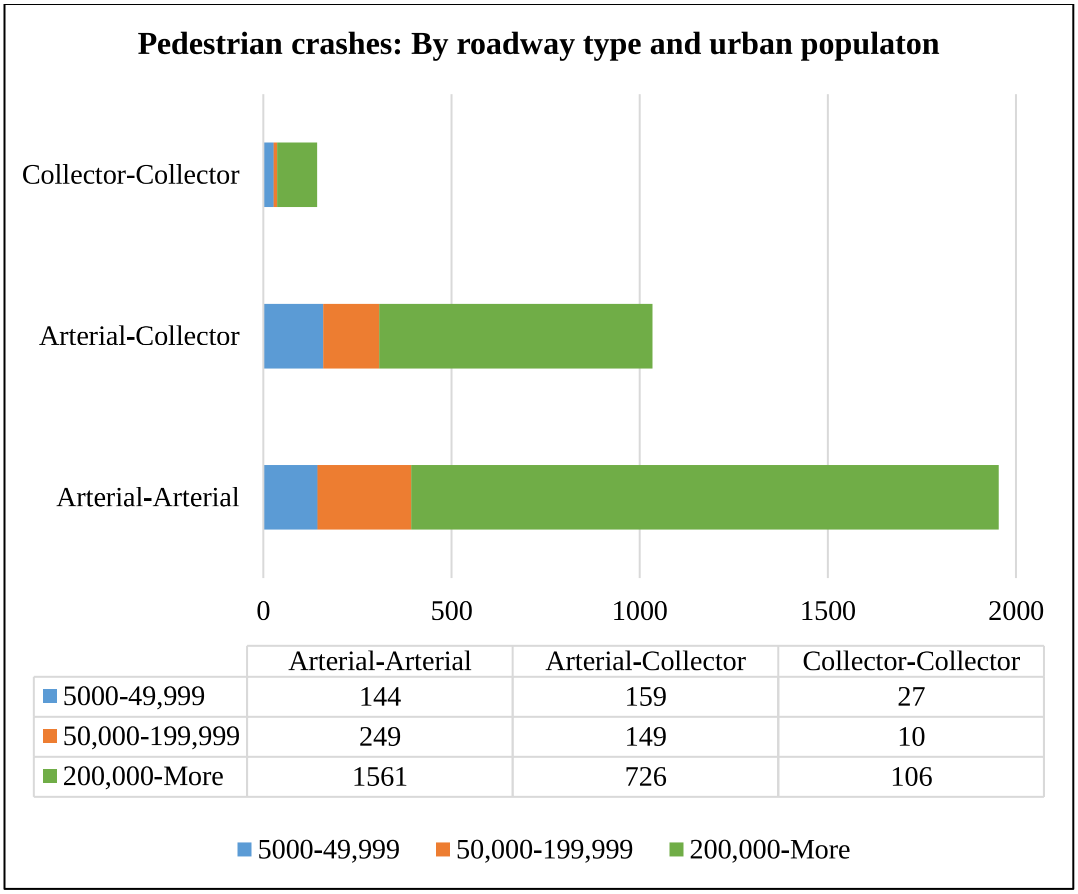
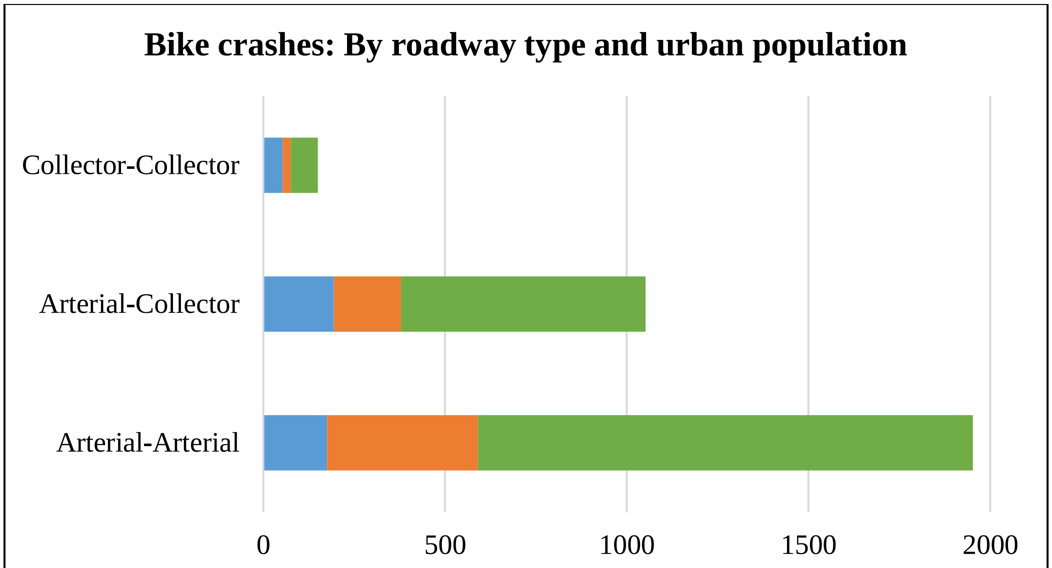


Figure 10 Pedestrian Involved Crashes: By Roadway Type and Urban Population



	Arterial-Arterial	Arterial-Collector	Collector-Collector
5000-49,999	176	194	53
50,000-199,999	414	185	22
200,000-More	1362	673	75

Figure 11 Bicycle Crashes by Roadway Type and Urban Population

4 Data Collection

This chapter covers methods and challenges encountered when gathering data that were used to develop safety performance functions for pedestrians and bicyclists. The data collected can be subdivided into six major groups:

- Non-motorized crash data
- Demographic data
- Land use data
- Road Geometry data
- Walk score index

4.1 Crash Data

Pedestrian and Bicyclist crash data for five years (2010-2014) were acquired from the Michigan State Police (MSP) in the office of Highways Safety and Planning (OHSP). Only crash data attribute that were considered relevant for this research were kept in order to facilitate efficient handling and processing of the data in tools like ArcGIS. ArcGIS was used to depict spatial trend and patterns of non-motorized crashes. A buffer of 150ft, established from previous study (Dolatsara, 2014) for aggregating non-motorized intersection crashes, was used. ArcGIS provides spatial join option which is the convenient means of aggregating crashes to each intersection. Figure 12 depicts how the buffer were created in ArcGIS.

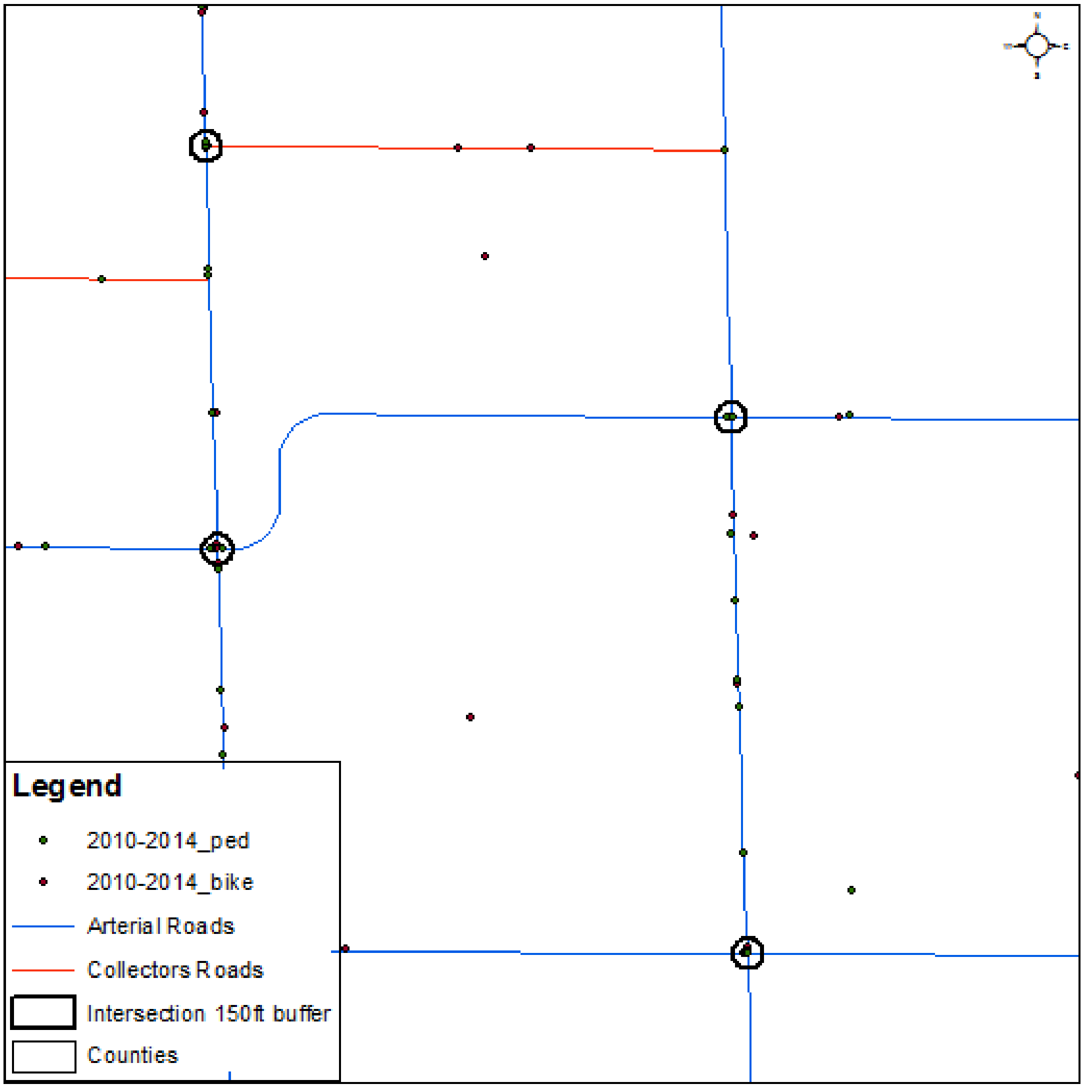


Figure 12 Collecting Intersection Crashes Falling Within 150ft Buffer

4.2 Land Use Data

Michigan land use shapefile was used to obtain the land use data for given urban intersections. Four major categories of urban land use data were considered for the analysis. These were commercial, residential, industrial, institutional, and outdoor recreation as shown in Figure 13. Commercial areas include Central Business District (CBD) and neighborhood business. In order to capture the dominant land use for a given intersection, weighted factors by area were used instead of dummy variables. Each land use area at the intersection was divided by the total area of blocks joining that intersection to obtain the weighted factors. Summation of weighted factors for all land use type in a given intersection will then be equal to one.

In previous studies, intersection with more than one land use type was considered as having mixed land use, not considering the fact that the proportions of each land use adjoining to intersection are different. Therefore area proportion were used so as to come up with unbiased description of land use characteristics surrounding a given intersection.

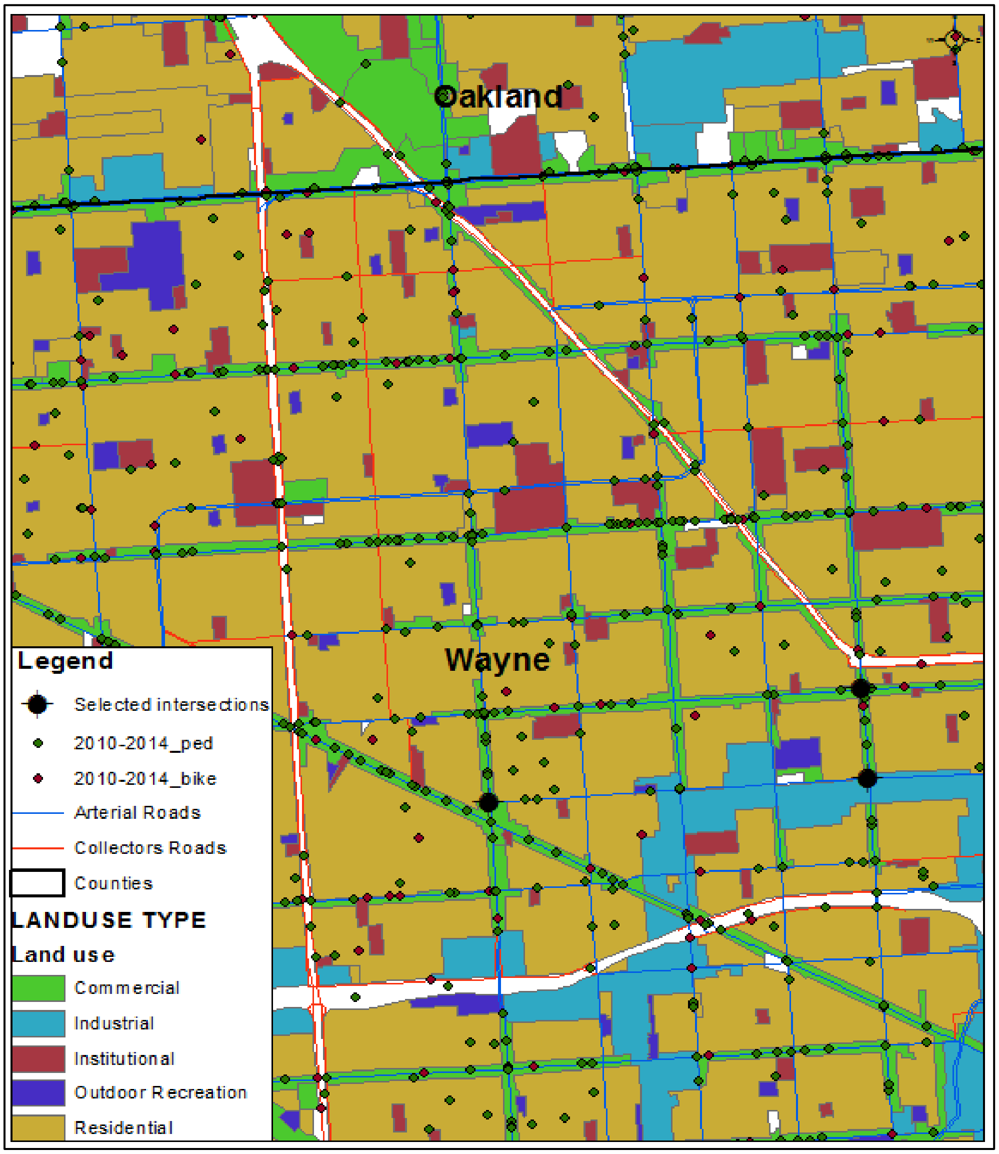


Figure 13 Example of Land Use Distribution

4.3 Average Annual Daily Traffic (AADT)

AADT is one of the essential parameter when evaluating risks that road users are experiencing when using road infrastructure. Most of the AADT data were collected from the Road Commission Transportation Count Database System (TCDS) of each county. The database acts like the central hub for storing and disseminating AADT data. Since the data are coming from different agencies within the same county, the data are first cleaned and validated before being available to the public. The level of details such as time of the day, hourly count differ across counties that have adopted this system. Below are some of the counties that have adopted this system in Michigan.

- Counties under SEMCOG (Wayne, Washtenaw, Macomb, Oakland, Monroe, St. Clair and Livingston)
- Counties under Grand Valley Metropolitan Counsel (Kent and Ottawa)
- Genesee County
- Kalamazoo County
- Eaton County
- Ingham County

With good cooperation from South Eastern Michigan Counsel of Governments (SEMCOG), it was possible to obtain AADT shapefiles for the counties under SEMCOG. This helped to automate the process of assigning AADT data to intersection segments. For other counties the data were recorded manually from their TCDS database.

4.4 Geometric Data

A list of all geometric factors of the roads that have been established from past studies to have an influence on non-motorized crashes was created. Google Earth was used as the main tool for obtaining all the road geometric characteristics. Below is the summary of main categories of roadway characteristics.

Signal information: Consist of the attributes such as signal control type, signal configuration (box or diagonal), left turn protection and no turn on red.

Intersection type: This provide information of whether the intersection was three leg or four leg intersection

Lane uses information: This group consists of attributes that described the lane use for each approach. Lane use information such as number of exclusive through lane, number of shared through-right lane, number of shared through-left lane, number of exclusive right lane, number of exclusive through lanes and total number of outgoing lanes were recorded.

Pedestrian facility: For each approach, information about the presence of pedestrian facility was collected. To be more precise, the pedestrian facilities were subdivided into four categories as shown below.

- Pedestrian sidewalk on one side separated from traffic
- Pedestrian sidewalk on one side not separated from traffic
- Pedestrian sidewalk on two sides separated from traffic
- Pedestrian sidewalk on two side not separated from traffic

The reason for subdivision of non-motorized facility information was to capture different level of risk that each category of pedestrian facility will have. For example presence of sidewalk which is not separated from the main road is more dangerous that the separated sidewalk. Also providing sidewalk only on one side of the road might have an implication on non-motorized movements and the way they interact with traffic as compared to providing pedestrian facility at both sides of the road.

Bicycle facility: This include information about presence of bike lane and the position of bike lane for each approach at the intersection. For example, the bike lane can be in-between lanes or the far right side of the approach.

Figure 14 and Figure 15 provide plan view and the street view in one of the intersection included in the study. The plan view provided geometric characteristics of the intersections, lane designation and facility information while the street view provided the signal information.

Figure 14 Plan View of Intersection as Seen From Google Earth

Figure 15 Google Earth Street View of an Intersection with Signal Information

4.5 Walk Score Index

This parameter has been used mostly in the field of urban planning, real estate and public health. Walk Score Index measures walkability of a given point or area on a scale of one to one hundred. The points are given after analyzing different walking routes to the amenities that are nearby. Distance decay function is used to model score index. Amenity that have 5min walk get the maximum points and the point keep on diminishing up to zero after 30 min walk. Also walk score captures pedestrian friendliness of a given location by considering population density, block length and intersection density.

Table 6 provides a description for different ranges of walk score. It ranges from car dependent areas to what is referred as walker's paradise. Figure 16 and Figure 17 shows two examples of intersections, one situated in a car dependent community while the other one situated in walker's paradise community.

Table 6 Definition of Walk Score Index

Score	Definition
90-100	Walkers' Paradise Daily trips do not require a car
70-89	Very Walkable Most trip can be accomplished on foot
50-69	Somewhat Walkable Some trips can be accomplished on foot.
25-49	Car Dependent Most trips require a car
0-24	Car Dependent almost all trips require a car

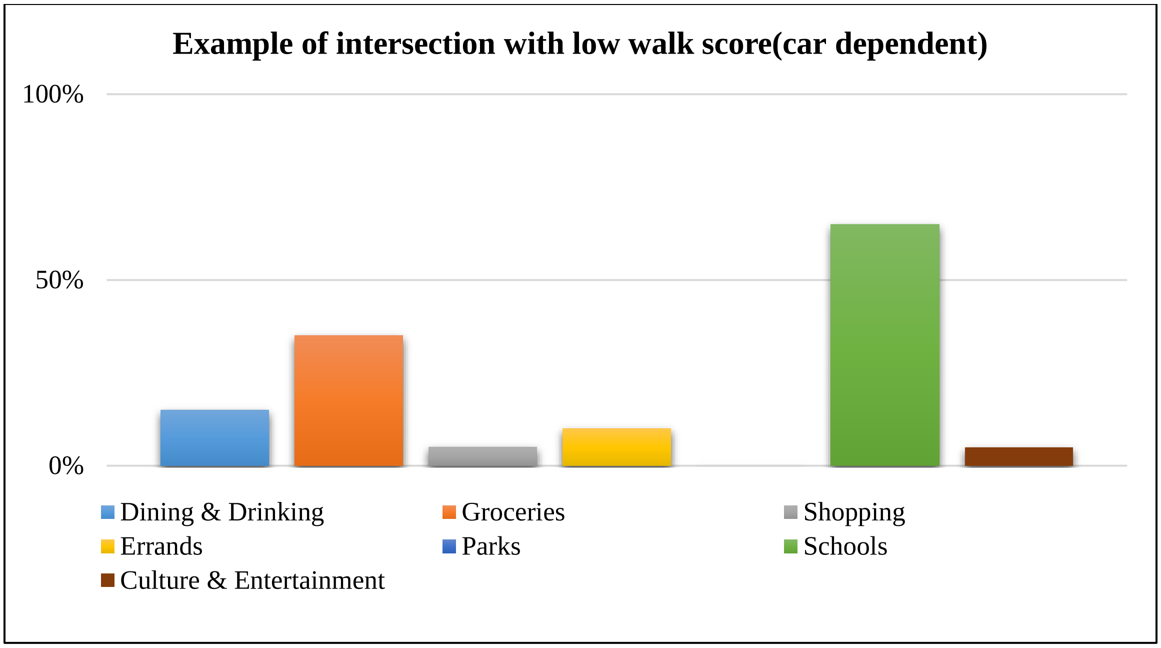


Figure 16 Eastern Ave SE @ 60th St SE Intersection with Walk Score of 17

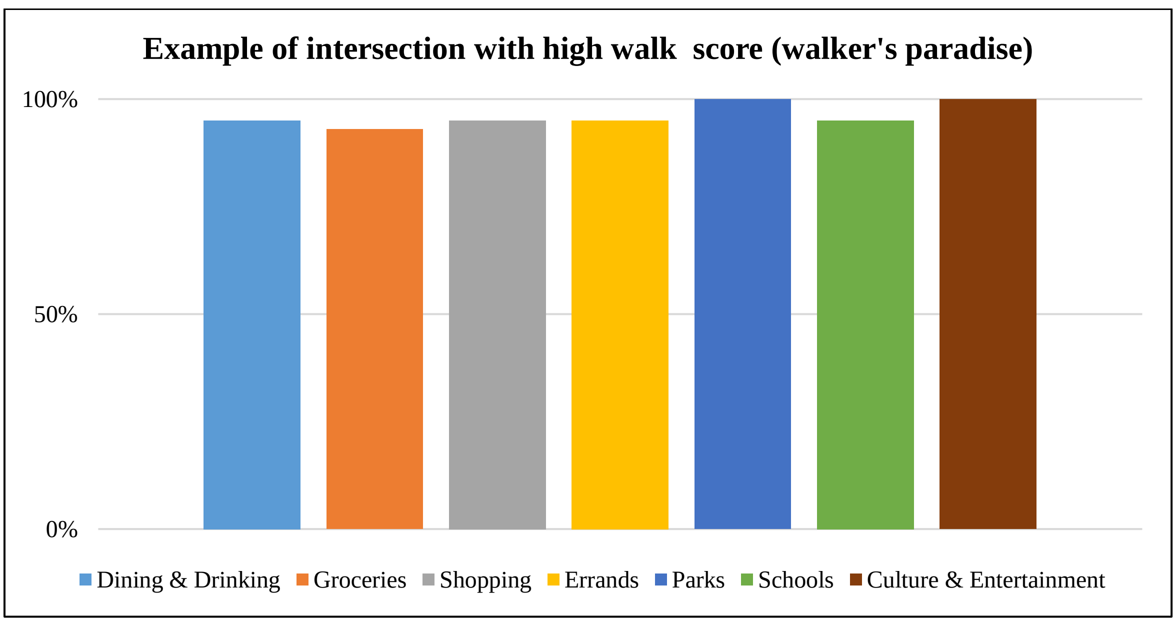


Figure 17: E Fulton St @ Lafayette Ave NE with the walk score of 91

4.6 Demographic Data

Demographic information at census block level using census shapefile were obtained for all selected urban intersections. Information that was extracted include population by age, educational status, poverty level, means of transportation to work and household income. Figure 18 provides part of Kent County were some of the urban intersections were included in the sample size. It can be observed that non-motorized crashes clustered in areas with high population density, below poverty level area and in areas with relatively high percentage of people who are walking and biking.

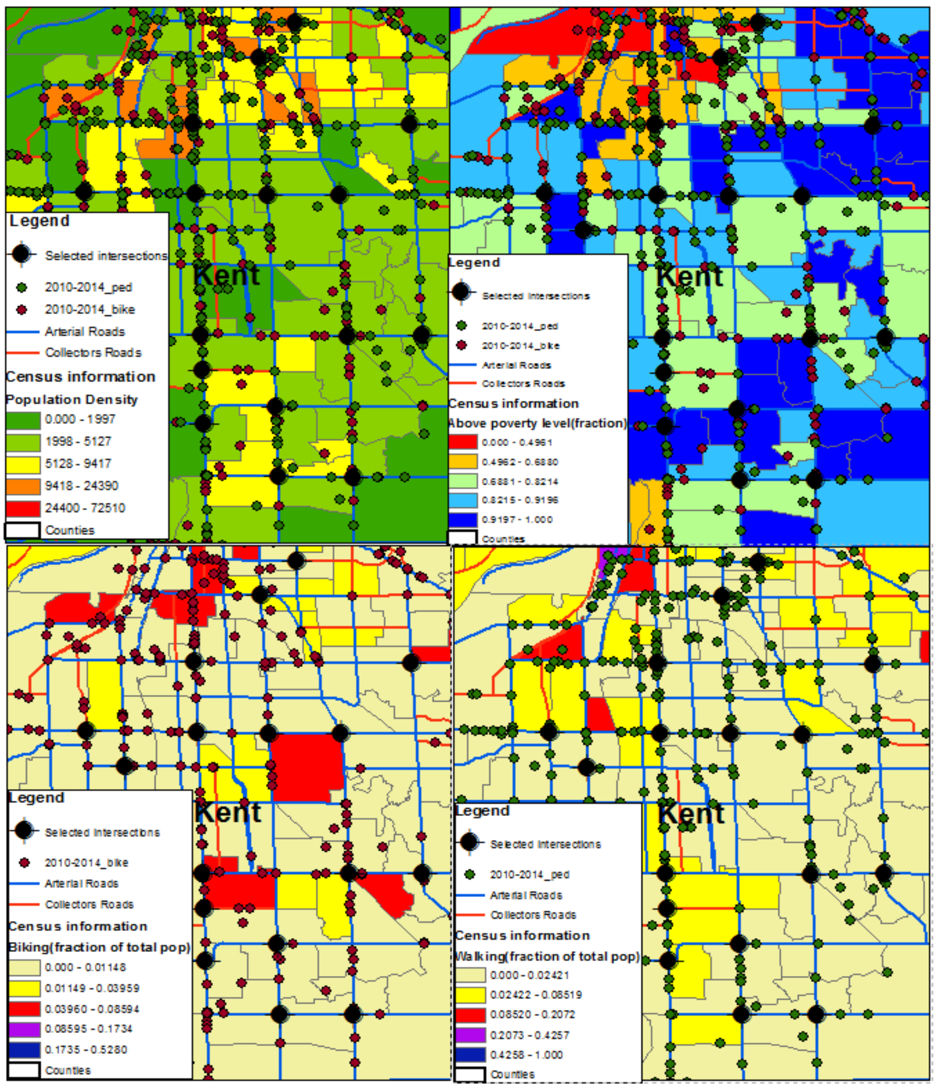


Figure 18 Census Information Extracted from Michigan Census Shapefile

5 Development of Safety Performance Function for Pedestrian and Bicyclist

This section describe how the safety performance functions (SPFs) for pedestrians and bicyclists were developed. The SPFs can be used to provide estimates of expected number of pedestrian and bicycle involved crashes given changes in traffic, geometric characteristics of the road, demographic and land use. In nutshell the procedure for developing SPFs is summarized below

- Structural equation modeling (SEM) to unveil the relationships between variables.
- Factor analysis for estimating proxy measure of pedestrians and bicyclists volume.
- Development of SPFs using different counts model.
- Comparison of models outcome using goodness of fit measures.
- Cross-validation of the SPFs
- Recommendation of final model based on within sample and cross validation results.

5.1 Structural Equation Modeling

In this project structural equation modeling (SEM) was used to unveil complex relationship between dependent and independent variables. With the use of SEM it was possible to quantify direct and indirect effects of observable variables on non-motorized crashes. The SEM has mainly two parts: (1) the structural model, which shows the potential causal dependencies between endogenous and exogenous variables, and (2) the measurement model, which shows the associations between latent variables and their indicators. Figure 19 shows the topology for pedestrian SEM while Figure 20 shows the topology for bike SEM. Variables that were significant at 95% confidence level were retained in the model.

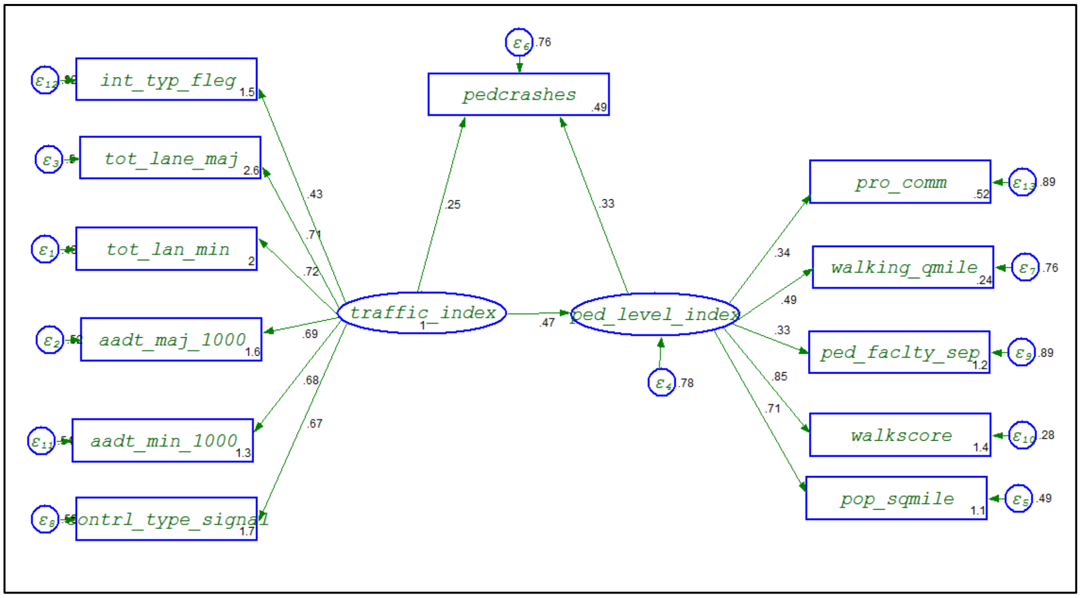


Figure 19. Pedestrian Crashes SEM

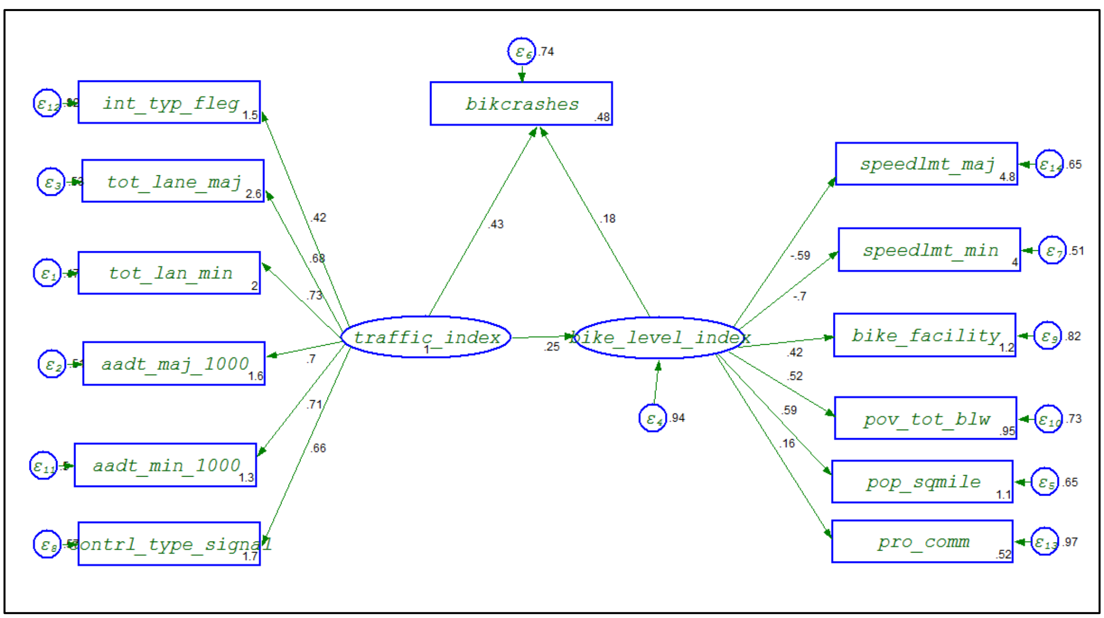


Figure 20. Bicycle Crashes SEM

The structural model results show that bike level and pedestrian level indices at intersection increase with the increase of vehicular activities (traffic index) which in turn increases pedestrian and bicycle crashes, respectively.

The measurement model indicated that high traffic index at intersections was manifested by type of intersection (i.e. four leg intersection), number of lanes in major and minor approaches, Average Annual Daily Traffic (AADT) in major and minor approach, and type of intersection control (i.e. signal control). The model also showed that high bike level score is associated with speed limit on both major and minor approaches, presence of bicycle facility, percentage of people below the poverty level in a given census block group where the intersection is located, population density around the intersection, and proportional of commercial land use in a given block where the intersection is situated. Furthermore, the measurement model showed that high pedestrian level score is manifested by proportion of commercial land use in a given block where the intersection is situated, increase in population density at a given census block group where the intersection is situated, and high walk score index. Other factors associated with pedestrian level score include number of people walking per square mile in a given census block group where the intersection is situated, presence of pedestrian facility separated from the motorized traffic, and percent of people using public transport in a given census block.

5.2 Factor Analysis

This is a multivariate technique which aims at explaining the joint variation and covariation of observed variables using less number of unobserved constructs which are called factors. It is a means of reducing dimensionality of correlated data as it tends to clusters variables into homogeneous sets. These set of unobserved constructs are unmeasured since we don't have a single perfect measure to represent them. In some instances, they are difficult to measure because of data insufficiency and other practical reasons.

Factor analysis accounts for measurement error when relating observed variables with the factors as opposed to methods which use Ordinary Least Square (OLS) approach. The error term express the percentage of observed variable variance that could not be explained by a factor. The estimation procedure utilize maximum likelihood approach which estimate model parameters that will minimize the discrepancy between the observed and predicted variance-covariance matrix. The parameters estimated from the factors analysis include factor loadings, observed variable error variances, factor variances and covariance. Factor loadings inform how each observed variable is related with the factor. It's a slope of regression coefficient between observed variable and a factor when presented in unstandardized form. When factor loading are standardized they represent correlation between a factor and an observed variable. Preference on which format of factor loading to be used depends on the type of study and intended outcome of the analysis.

Using matrix notation, factor analysis can be presented as

$$y_{nx1} = \Sigma_{nxm} F_{mx1} + e_{nx1}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}_{nx1} = \begin{bmatrix} \lambda_{11} & \cdots & \cdots & \lambda_{1n} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \lambda_{n1} & \cdots & \cdots & \lambda_{nm} \end{bmatrix}_{nxm} \begin{bmatrix} F_1 \\ \vdots \\ F_m \end{bmatrix}_{mx1} + \begin{bmatrix} e_1 \\ \vdots \\ \vdots \\ e_n \end{bmatrix}_{nx1}$$

Where

y_{nx1} = Observed variables matrix

Σ_{nxm}= variance-covariance matrix which comprises of factor loadings, λ_{nm}

F_{mx1} = Factor Matrix

e_{nx1} = Error term

Estimation procedure of unknown parameters such as factor loading and error term utilize Maximum Likelihood (ML) approach, which aims at minimizing the following function.

$$\Gamma_{ml} = ln|\Sigma| - ln|\text{S}| + trace|(S)(\Sigma^{-1})| - p$$

Where

Γ_{ml} = Log likelihood function

$|\Sigma|$ = Determinant of predicted covariance-variance matrix

$|S|$ = Determinant of observed covariance-variance matrix

p = Number of input indicators/observed variables

Trace= Sum of the diagonal values in the covariance-variance matrix

In ideal case where $|\Sigma|$ =$|S|$, $(S)(\Sigma^{-1})$ will turn out to be an identity matrix in which its trace value will be equal to p. Hence the log likelihood function, Γ_{ml} will be equal to zero (Jaccard et al, 1996)

5.2.1 *Model Specification*

Due to unavailability of non-motorized volume count, factor analysis was used in this research to estimate proxy measure of pedestrians and bicyclists volume at urban intersections. Observed variables that were used to form proxy measure of pedestrians and bicyclists exposure were selected based on prior research knowledge and results of the SEM model presented in Section 5.1. Variables that were significant at 95 percent confidence level were retained in the final factor analysis model. Table 7 and table 8 provide the descriptive summary of the variables that were significant for pedestrians and bicyclists factor analysis respectively. Figure 21 and Figure 22 provide a schematic diagram of significant observed variables for pedestrians and bicyclists factor analysis. The error terms ε for each observed variable was estimated in the process.

Table 7 Variable Description: Proxy Measure of Pedestrians Exposure

Variable	**Description**	**Mean**	**Std. Dev.**	**Min**	**Max**
Percent using public transport	Percentage of people using public transit in a census block where the intersection is located	0.97	2.39	0	22.15
Population per square mile	Population density for a census block	420.18	370.61	12.87	2384.90
Percent of poverty below	Percentage of people below poverty level in a census block	13.47	14.20	0	83.72
Walking per square mile	Walking commuters density in a census block	36.02	148.45	0	1671.94
Pedestrian facility	Dummy variable for the presence of pedestrian facility separated from roadway	0.59	0.49	0	1
Walk score	Walk score index estimated using distance decay function	35.77	24.98	0	94
Proportion of commercial land use	Proportion of commercial land use by area	0.15	0.28	0	1

Table 8 Variables That Description: Proxy Measure of Bicyclists Exposure

Variable	**Description**	**Mean**	**Std. Dev.**	**Min**	**Max**
Bike facility	Presence of bike facility (side path/bike lane)	0.60	0.49	0	1
Poverty level below	Percentage of population below poverty level in a given census block group	13.44	14.19	0	83.72
Population per square mile	Population density for a census block	419.05	370.75	0	2384.90
Speed limit major	Speed limit in the major approach	42.83	8.98	25	70
Speed limit minor	Speed limit in the minor approach	34.89	8.65	20	55
Proportion of commercial land use	Proportion of commercial land use by area	0.15	0.28	0	1

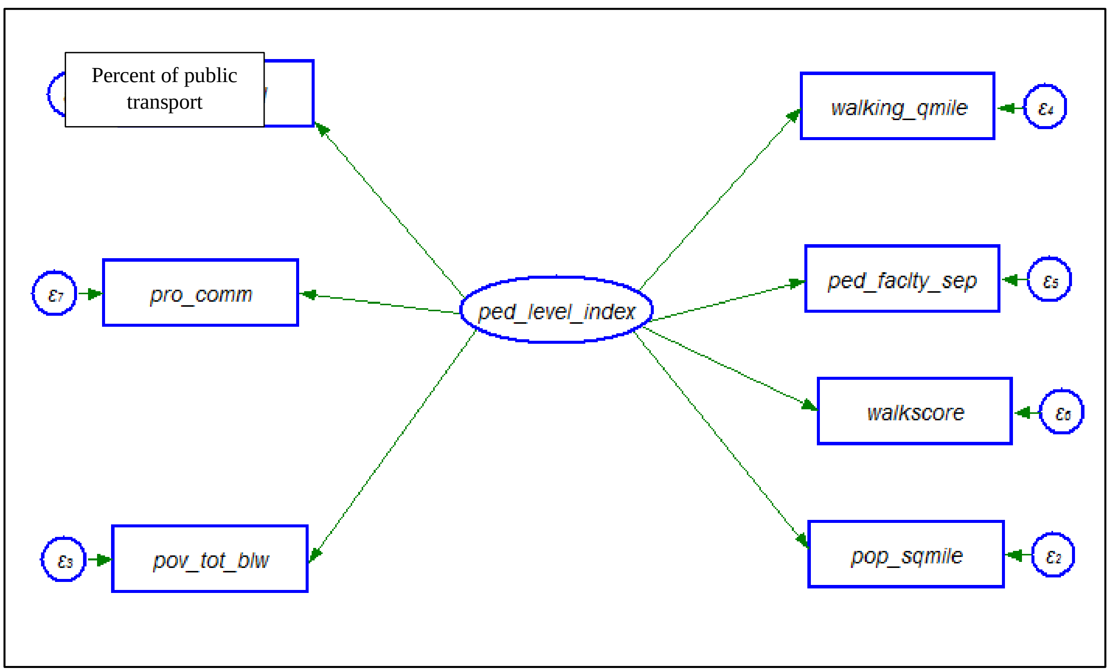


Figure 21 Schematic Diagram for Pedestrians Factor Analysis

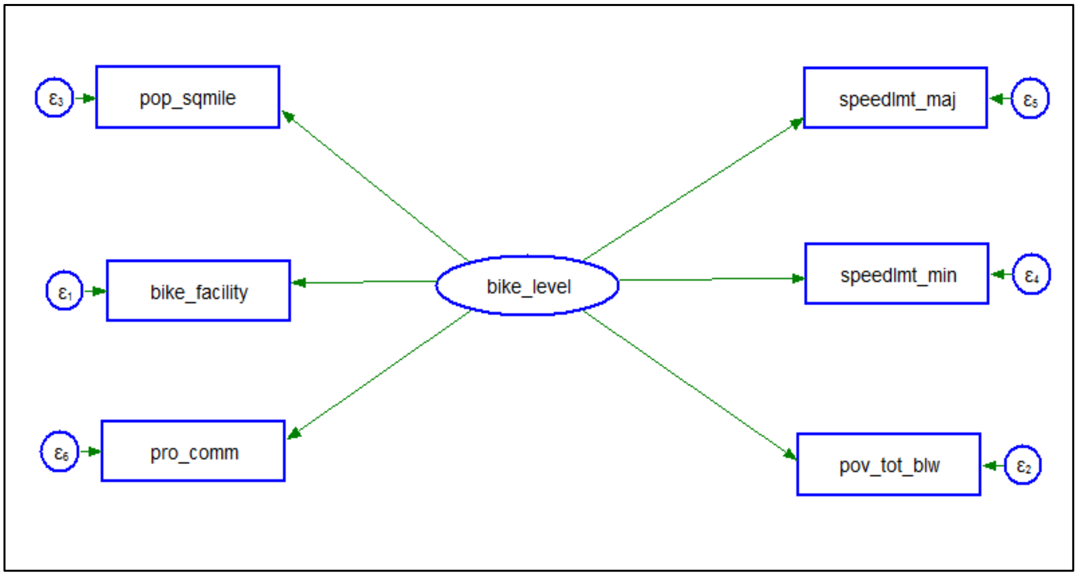


Figure 22 Schematic Diagram for Bicyclist Factor Analysis

5.2.2 Model Estimation

As summarized in Table 9, the increase in pedestrians level score index at a given intersection was manifested by the increase in percentage of people using the public transit in a given block group where the intersection was located, population density, percentage of household below poverty level, number of workers commuting to their working places by foot per square mile, walk score index, proportion of commercial land use and presence of pedestrian facility separated from the roadway.

Table 9 Standardized Factor Loadings for Pedestrians Level Score

Variable	**Standardized Coef.**	**Std. Err.**	**z**	**P>z**
Percent using public transport	0.5397	0.0440	12.26	0
Population per square mile	0.6959	0.0345	20.17	0
Percent of poverty below	0.6131	0.0392	15.65	0
Walking per square mile	0.5299	0.0448	11.82	0
Pedestrian facility	0.2568	0.0545	4.72	0
Walk score	0.8347	0.0288	29.01	0
Proportion of commercial land use	0.3244	0.0518	6.26	0

Table 10 summarized the significant factor loading for bicyclist factor analysis. Bicyclist level score, a proxy measure of bicyclist volume was found to increase with the following factors; the presence of bicycle facility which includes bike lanes and sidewalks, increase in percentage of people below poverty level, increase population density, lower speed limit in major and minor approach and increase in proportion of commercial land use by area in a given census block group were the intersection is situated.

Table 10 Standardized Factor Loadings for Bicyclist Level Score

Variable	**Standardized Coef.**	**Std. Err.**	**z**	**P>z**
Bike facility	0.3713	0.0547	6.79	0
Poverty level below	0.4860	0.0507	9.59	0
Population per square mile	0.5454	0.0496	11.01	0
Speed limit major	-0.7318	0.0415	-17.61	0
Speed limit minor	-0.6646	0.0423	-15.7	0
Proportion of commercial land use	0.1358	0.0601	2.26	0.024

5.2.3 Estimation of Bicyclists and Pedestrians Level Score

Pedestrians and bicyclists level score were then estimated from their respective significant observed variables. There are different methods in which the factor score can be estimated such as sum score by factor, weighted sum scores, regression scores, Bartlett Scores and Anderson-Rubin Scores. Distefano et al (2009) provides a good description of these factors giving applicability, pros and cons of each. For this study, the estimation procedure adopted was least squares regression approach a procedure similar to regression score developed by Thomson (1935).

Stata, which was the statistical package used in data analysis for this project, utilizes this approach in computing factor scores. In this method, the observed variables are centered to their mean. The final factor score is the sum of the product between factor score weights and their respective observed variables. The factor score weight are obtained by multiplying the inverse of observed variable covariance matrix by factor-observed variables covariance matrix.

Mathematically estimation of factor score can be expressed as follows

$$f_i = (\Sigma^{-1} * \Lambda)$$

$$Factor\ score = \sum_{i=1}^{i=n} fi * (x_i - \bar{x}_i)$$

Whereby

f_i = the factor score weight for observed variable i

Σ^{-1} = Inverse of observed variable covariance matrix

Λ = Factor-observed variable covariance matrix

x_i = the observed variable i

$\bar{x}_i$ = the mean of observed variable i

Latent pedestrian level score can be computed as

$$\boldsymbol{Pedlevel} = 0.0707\left(perc_{publ} - 0.974\right) + 0.0008\left(pop_{sqmile} - 420.178\right) + 0.0153\left(pov_{tot_{blw}} - 13.473\right) + 0.0011\left(walking_{qmile} - 36.32\right) + 0.1233\left(ped_{faclty} - 0.586\right) + 0.0244(walkscore - 35.772) + 0.2828(pro_{comm} - 0.146)$$

Latent bicyclist level score can be computed as

$$\boldsymbol{Bikelevel} = 0.0415\left(bike_{facility} - 0.598\right) + 0.0021\left(pov_{tot_{blw}} - 13.44\right) + 0.0001(pop\_sqmile - 419.052) - 0.0086(speedlmt\_min - 34.893) - 0.0063(speedlmt\_maj - 42.828) + 0.0231(pro\_comm - 0.146)$$

5.3 Development of SPFs

Before considering any potential count model to be used for formulating safety performance function, distribution of pedestrian and bicyclist-involved crashes at selected urban intersections were studied. 85 percent of the selected intersections were used for model calibration process and the remaining 15 percent were used for cross-validation. Figure 23 and Figure 24 below provide the summary of intersections that were used for calibration process. Non-motorized crashes as described in the data collection section comprised of four year from 2010 through 2014. About 70 percent of selected intersections had zero pedestrian crashes in five years while for bicyclist 73 percent of the intersections had zero crashes. The presence of excess zero in the sample size provided a clue on which type of count models that were to be considered in the analysis.

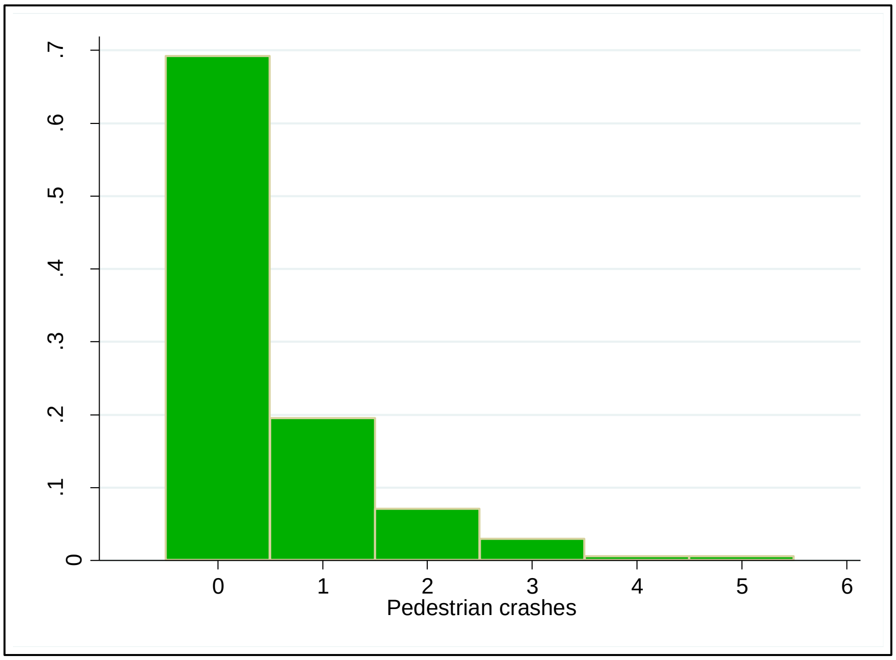


Figure 23 Distribution of Pedestrian Crashes(2010-2014) at Selected Intersections

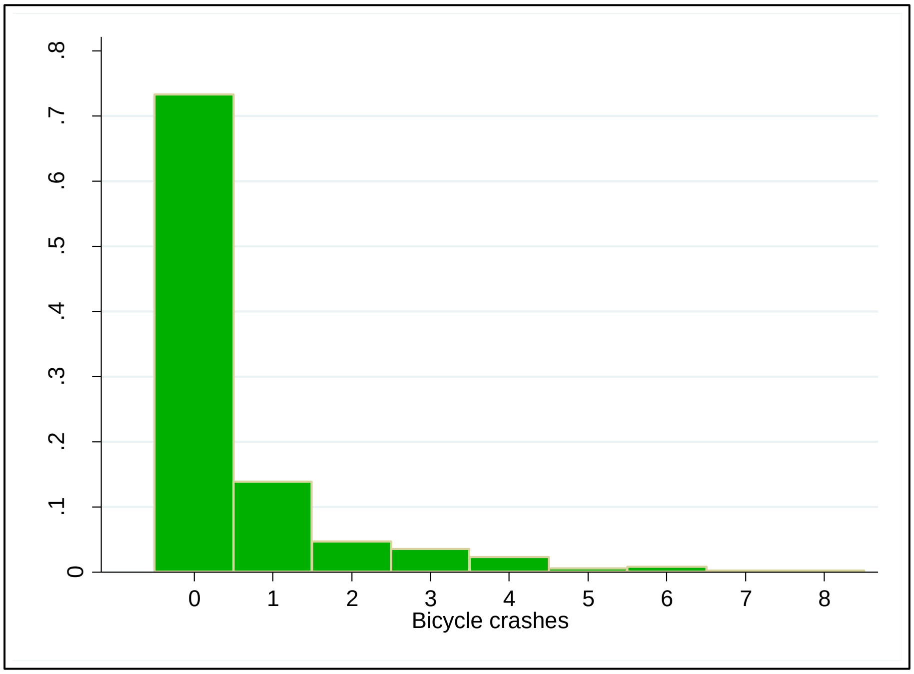


Figure 24 Distribution of Bicyclist Crashes(2010-2014) at Selected Intersections

5.3.1 *Model Comparison*

After studing the distribution of non-motorized crashes at selected intersections, four count model were considered for the analysis as listed below:

- Poisson Regression Model (NRM)
- Negative Binomial Regression Model (NBRM)
- Zero Inflated Poisson Regression Model (ZIP)
- Zero Inflated Negative Binomial Model (ZINB)

Poisson model and negative binomial regression model have been used widely in most of the researches that analyze count data. Equidispersion assumption of Poisson regression model that the mean and variance are identical is often violated. That's why Negative Binomial have become most commonly used count model because it account for over-dispersion.

Zero inflated models were also considered as the potential fit to the data due to presence of excess zero at the selected intersections. For zero inflated models, there are two type of zero counts. The first type of zero is predicted by the binary component of the model, whereby it shows locations that will always have zero count. The second type of zero are predicted by the count model component whereby it shows location that are most likely but not always have zero counts. Kwigizile et.al (2014) provided a good and simple formulation of four count models that were compared in this analysis as show below.

5.3.1.1 *Poisson and Negative Binomial Regression*

The probability of intersection i having pedestrian/bicycle crashes in a given time period can be written as:

$$P(y_i) = \frac{EXP(-\lambda) \cdot \lambda_i^{y_i}}{y_i!}$$

Whereby

λ_i is the Poisson parameter for urban intersection i, which for this study it can be defined as the expected number of pedestrian/bicyclist crashes in five years period. This parameter is a function of predictor variables given as

$$\lambda_i = EXP(\boldsymbol{\beta X}_i)$$

Where $\boldsymbol{\beta}$ is the vector of estimable parameter

Estimation of parameters deploy maximum likelihood method given as

$$LL(\beta) = \sum_{i=1}^{N} [-EXP(\boldsymbol{\beta X}_i) + y_i \boldsymbol{\beta X}_i - \ln(y_i!)]$$

Negative binomial regression which handle cases where mean and variance of the count data are not equal can be derived from the Poisson model as follows;

Generalizing Poisson model by introducing unobserved effect ε_i, whereby the expected Poisson parameter becomes

$$\lambda_i = EXP(\boldsymbol{\beta X}_i + \varepsilon_i)$$

With $\lambda_i = EXP(\varepsilon_i)$ known as gamma distributed error term with mean of one and variance of α^2.

Upon modification mean-variance relationship for expected number of pedestrian/bicycle crashes y_i becomes:

$$Var[y_i] = E(y_i) \cdot [1 + \alpha E(y_i)] = E[y_i] + \alpha E(y_i)^2$$

If α is significantly different from zero then the bicyclist/pedestrian involved crash data are said to be overdispersed for positive α values and underdispersed for negative α values. For overdispersion case, the resulting Negative binomial probability distribution becomes

$$P(y_i) = \frac{\Gamma\left(\left(1/\alpha\right) + y_i\right)}{\Gamma\left(1/\alpha\right) y_i!} \left(\frac{1/\alpha}{\left(1/\alpha\right) + \lambda_i}\right)^{1/\alpha} \left(\frac{\lambda_i}{\left(1/\alpha\right) + \lambda_i}\right)^{y_i}$$

Whereby

$\Gamma(x)$ is a value of the gamma function.

α is an overdispersion parameter

y_i is the number of pedestrian/bicyclist involved crashes for intersection *i*

5.3.1.2 Zero Inflated Models

As it has been mentioned earlier, zero inflated models are used when there is excess number of zero in the count that tends to violate assumptions used in Poisson or Negative binomial model formulation. For ZIP model the probability for the two component (binary logistic and Poisson regression) can be estimated as follows (Lord et al, 2005)

$$\Pr(y_i = 0) = p_i + (1 - p_i)e^{y}$$

$$\Pr(y_i > 0) = (1 - p_i)\frac{e^{-y}y^{n}}{n!}$$

The probability of zero pedestrian/bicyclist intersection crashes for the binary component of the ZINB model can be computed as:

$$Pr(y_i = 0) = p_i + (1 - p_i)\left[\frac{1/\alpha}{1/\alpha + \lambda_i}\right]^{1/\alpha}$$

As for count component of the model with the probability of $y_i > 0$ it can be computed as

$$\Pr(y_i = y) = (1 - p_i)\left[\frac{\Gamma\left(\left(1/\alpha\right) + y\right)\psi_i^{1/\alpha}(1 - \psi_i)^{y}}{\Gamma\left(1/\alpha\right)y!}\right]$$

With $\psi_i = \frac{1/\alpha}{1/\alpha + \lambda_i}$.

5.3.2 Goodness of Fit Tests

Goodness of fit test that were used to analyze how well the model fits the data are summarized below

- Akaike's Information Criterion (AIC)
- Bayesian Information Criterion (BIC)
- Vuong test
- Residual probability plot
- Mean Standard Error (MSE)

The selection of the best model was based on collective assessment of all goodness of fit measures. AIC, BIC and Vuong test were used to test within sample goodness of fit. Residual probability was used for within sample and for cross validation. Hilbe (2011) provide good description and application of BIC, AIC, Vuong test and Residual probability plot. Mathematical formula for the given goodness of fit measures are provided below:

Akaike's Information Criterion (AIC)

$$AIC = \frac{-2L + 2k}{n}$$

Bayesian Information Criterion (BIC)

$$BIC = -2L + klog(n)$$

With

k=number of predictors including the intercept

n= number of observation

L= model log-likelihood.

Vuong Test

Test whether the zero inflated models are preferred over non-inflated models. It's the most commonly used test despite invention of other test serving the similar purpose. It is conservative and therefore reduces the chances of making incorrect decision (Clarke 2007).

It is given as the log ration of the sum of probability for each observation computed as

$$\phi_i = \ln\left(\frac{\sum_i P_1(y_i/x_i)}{\sum_i P_2(y_i/x_i)}\right)$$

With Vuong test statistics is a calculated as

$$V = \frac{\sqrt{N}(\bar{\phi})}{SD(\phi_i)}$$

Where

$P_1(y_i/x)$= Probability of observing pedestrian/bicyclist involved y crashes on the basis of variable x for model i (inflated model)

$P_1(y_j/x)$ = Probability of observing pedestrian/bicyclist involved y crashes on the basis of variable x for model j (Non-inflated model)

$\bar{\phi}$ = Average of the log ratios

$SD(\phi_i)$ = Standard deviation of the log ratios

If V is greater than 1.96, model i is favored while if V is less than -1.96, model j is favored

Residual Probabilities

It is computed as the difference between the average observed probability and average predicted probability for each pedestrian/bicyclist observed crash count at intersection. The model with the best performance has residual probabilities close to zero for all the pedestrian or bicyclist observed crash count

Root Mean Square Error (RMSE)

RMSE is the square of the difference between observed values and the values predicted by a model. Individual differences between observed and predicted values are normally called residuals. It can be computed as

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}$$

With

$\hat{y}_i$ = predicted pedestrian/bicyclist crashes for intersection i

y_i = observed pedestrian/bicyclist crashes for intersection i

N= total number of intersections

5.4 Pedestrian Safety Performance Function

Table 11 provides the descriptive statistics of significant variable that were retained in the final pedestrian SPF. Significant predictor variables were Average Annual Daily Traffic in major approach, Average Annual Daily Traffic in minor approach and pedestrians level score.

Table 11 Descriptive Statistics of Significant Variables Used in Pedestrians SPF

Variable	**Description**	**Mean**	**Std. Dev.**	**Min**	**Max**
Pedestrian crashes	Pedestrian Crashes (2010-2014)	0.479	0.872	0	5
AADT major approach	AADT in the major approach in thousands	14.887	8.826	1.388	57.285
AADT minor approach	AADT in the minor approach in thousands	6.807	5.107	0.346	24.716
Pedestrian level score	Standardized Latent Pedestrians Level	-0.016	1.175	-1.580	4.562

The coefficient for significant predictor variables and their respective z-score values in bracket are shown in Table 12. Both AADT in the major approach, AADT in the minor approach and pedestrians level score increase the likelihood of expected number of pedestrian crashes. Both were significant at 95 percent confidence level. For Zero-inflated model, predictor variable in the binary component of the model was pedestrian level score.

Table 12 Summary of Significant Predictor Variables Used in Pedestrian SPFs

Variable	**PRM**	**NBRM**	**ZIP**	**ZINB**
AADT major approach	0.0352 (3.84)	0.0361 (3.3)	0.0234 (2.14)	0.0234 (2.13)
AADT minor approach	0.0433 (3.02)	0.0454 (2.6)	0.0405 (2.51)	0.0409 (2.49)
Pedestrian level score	0.5204 (9.19)	0.5627 (7.81)	0.2329 (2.69)	0.2392 (2.59)
Constant term	-1.912 (10.36)	-1.965 (-9.45)	-1.094 (-4.46)	-1.1117 (-4.27)
Over dispersion parameter				
alpha		0.319		0.027
Inflate(For zero-inflated models)				
Pedestrian level score			-2.375 (-4.31)	-2.403 (-4.18)
Constant			-0.9181 (-1.88)	-0.972 (-1.73)

*Note: Values in brackets indicate z-score

5.4.1 Comparison of the Models Based on Goodness of Fit Tests

Post estimation results such as expected number of counts and predicted probability of each count were obtained. Goodness of fit tests were performed so as to select the model with the best fit. Table 13 provides the BIC and AIC values for each model. The model that fits best the data, is the one with the lowest BIC or AIC value. Zero Inflated Poisson (ZIP) model had the lowest value for both BIC and AIC. The Vuong test was performed to test whether the zero inflated models are preferable to the non-inflated count models. In other words, the test shows if the number of zeros in the data exceed Poisson or negative binomial distributional assumption. The test favored zero-inflated models as shown in Table 14 with the p-value that was below 0.05.

Table 13 BIC and AIC Values for Pedestrian SPFs

Variable	**PRM**	**NBRM**	**ZIP**	**ZINB**
BIC	563.6	564.3	550.7	556.5
AIC	548.3	545.2	527.8	529.7

Table 14 Vuong Test for Pedestrian SPFs

Comparison	**z**	**p**	**Comment**
ZIP over PRM	2.443	0.007	Significant(<0.05)
ZINB over NBRM	2.206	0.014	Significant(<0.05)

Comparison based on Residual probabilities was performed for both models as shown in Figure 25 . Theoretically, if the model fit the data well then residual probabilities will be zero for each crash count. Poisson regression showed to have high deviation from zero residual probability when predicting zero up to two crashes per intersection. The remaining three models had better predictive performance for all crash counts per intersection found in the data.

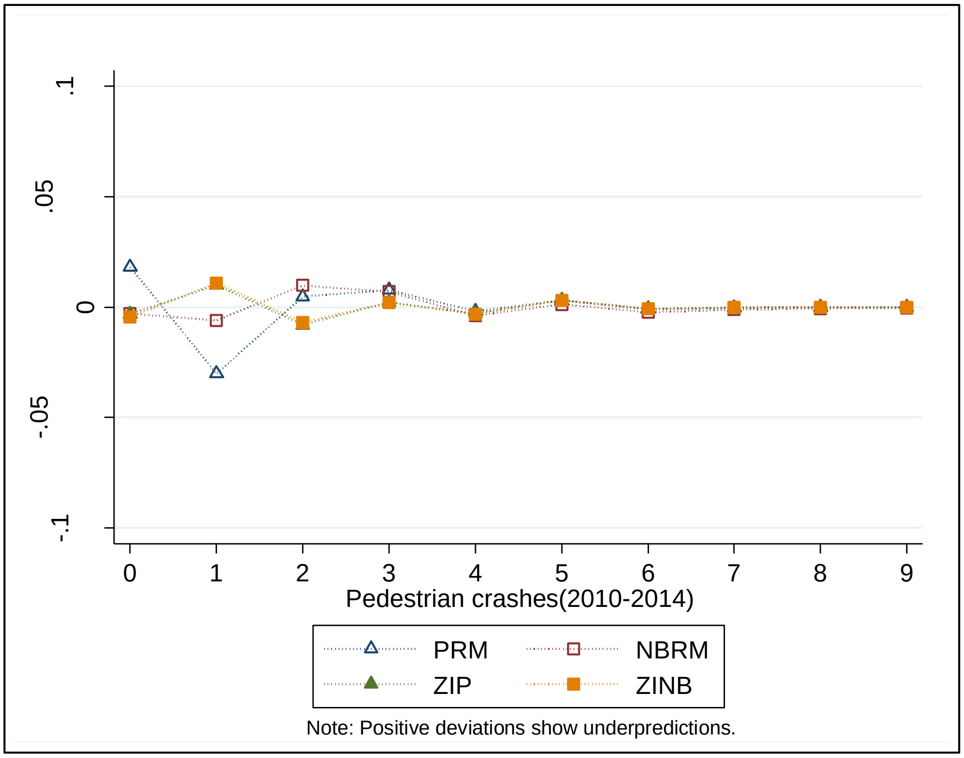


Figure 25 Residual Probabilities for Each Crash Counts per Intersection

5.4.2 Cross Validation

Figure 26 and Figure 27 display residual probability and root mean square error for out-of-sample estimation respectively. Residual probability plots from cross validation analysis using the remaining 15 percent of the data were less precise in prediction for all four models as compared to within sample prediction. Inflated models had a better performance as compared to NBRM and PRM. Similar outcome was observed when comparing RMSE values for each model as shown in Figure 27.

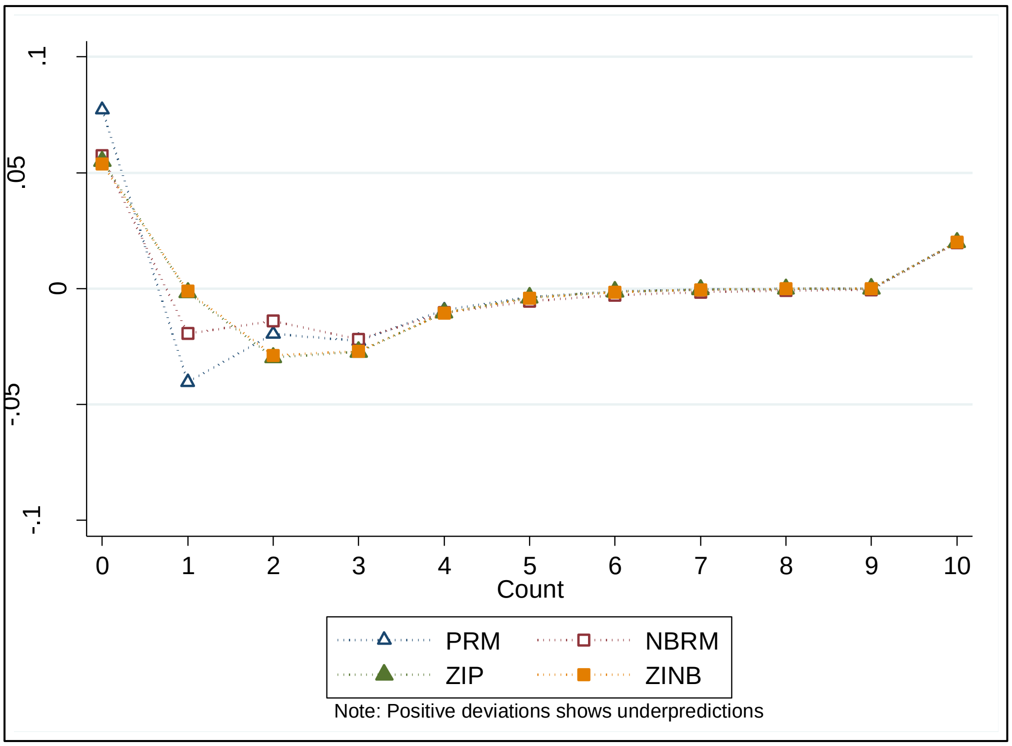


Figure 26 Probability Residual Plots for Out-Of-Sample Prediction

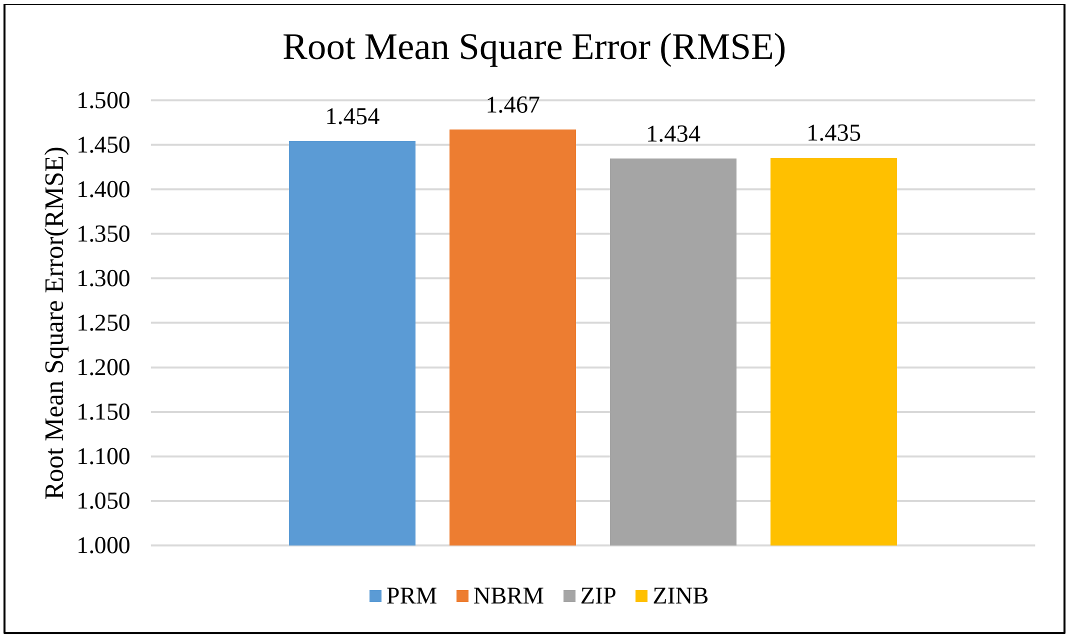


Figure 27 Mean Square Error from Cross Validation Data Set

5.4.3 *Final Proposed Pedestrian SPF*

Based on the assessment of goodness of fit measures, Zero Inflated Poisson Regression Model was selected as the final model for pedestrians SPF. Mathematically it can be expressed as:

$$Number\ of\ pedestrian\ crashes\ per\ five\ years =$$

$$\left[1 - \frac{1}{1 + e^{(0.918+2.375pedlevel)}}\right] \left[e^{-1.094+0.0234aadt_{maj}+0.0405aadt_{min}+0.2392pedlevel}\right]$$

Where

$aadt_{maj}$ = AADT in the major approach in thousands

$aadt_{min}$ = AADT in the minor approach in thousands

$pedlevel$ = Pedestrian level score

5.5 Bicycle Safety Performance Function

Table 15 below provides the descriptive statistics of significant variable that were kept in the final bicyclist SPFs for the count models that were compared. Significant predictor variables were Average Annual Daily Traffic (AADT) in major approach, Average Annual Daily Traffic (AADT) in minor approach and bicycle level score.

Table 15 Descriptive Statistics of Variables Used in Bicyclist SPFs.

Variable	**Description**	**Mean**	**Std. Dev.**	**Min**	**Max**
Bicycle-involved crashes	Bicyclist Crashes (2010-2014)	0.5621	1.2316	0	8
AADT major approach	AADT in the major approach in thousands	14.628	8.9760	1.388	57.285
AADT minor approach	AADT in the minor approach in thousands	6.647	5.1600	0.346	24.716
Bicycle level score	Standardized latent Bicyclist Level	-0.012	0.159	-0.340	0.411

5.5.1 Modeling Results

Both of the predictor variables had a positive coefficient as listed in Table 16. As their magnitude increases they tend to increase the likelihood for the crash to occur. For example using PRM model, a unit increase in bike level score will tend to increase the log of expected bicyclist crashes by 3.487. For other predictor variables, their coefficient can be interpreted in similar fashion. For the inflated model, the predictor variable for binary component was bike level score. For ZIP it was significant at 95 percent confidence while for ZINB it was significant at 90 percent confidence level.

Table 16 Summary of Significant Predictor Variables Used in Bicycle SPFs

Variable	**PRM**	**NBRM**	**ZIP**	**ZINB**
AADT major approach	0.0334 (3.72)	0.0406 (2.69)	0.0217 (2.15)	0.0347 (2.37)
AADT minor approach	0.0801 (6.29)	0.0870 (3.63)	0.0730 (5.15)	0.087 (3.91)
Bicycle level score	3.487 (6.75)	3.561 (4.64)	1.593 (2.07)	2.007 (1.94)
Constant term	-1.949 (-11.29)	-2.144 (-8.39)	-0.895 (-3.56)	-1.637 (-4.29)
Over dispersion parameter				
alpha		1.561		0.801
Inflate(For zero inflated models)				
Bicycle level score			-3.677 (-2.34)	-5.839 (-1.73)
Constant			0.177 (0.77)	-0.903 (-1.12)

*Note: The values inside the brackets are z-score

5.5.2 Comparison of Models

Table 17 and Table 18 summarize goodness of fit measure for within-sample prediction. Negative Binomial regression model (NBRM) had a lowest BIC value while Zero inflated negative binomial (ZINB) had AIC value slightly below NBRM. Vuong test show insignificant preference of ZINB over NBRM with the p-value>0.05. Therefore based on the three test NBRM outperform other three models.

Table 17 Information Criteria Goodness of Fit Tests for Bicycle SPFs

Variable	**PRM**	**NBRM**	**ZIP**	**ZINB**
BIC	677.453	618.335	633.446	626.686
AIC	662.161	599.22	610.507	599.925

Table 18 Vuong Test for Bicycle SPFs

Comparison	**z**	**p**	**Comment**
ZIP over PRM	2.71	0.003	Significant
ZINB over NBRM	0.893	0.186	Not significant

Figure 28 illustrate the residual probability for within the sample bicyclist crashes per intersection prediction. Within-sample residual probability depicts fairly similar trend for NBRM and ZINB. ZINB had a slight better prediction for intersection with zero, one and three bicycle crashes per intersection in five years period. Poisson regression model had a least performance.

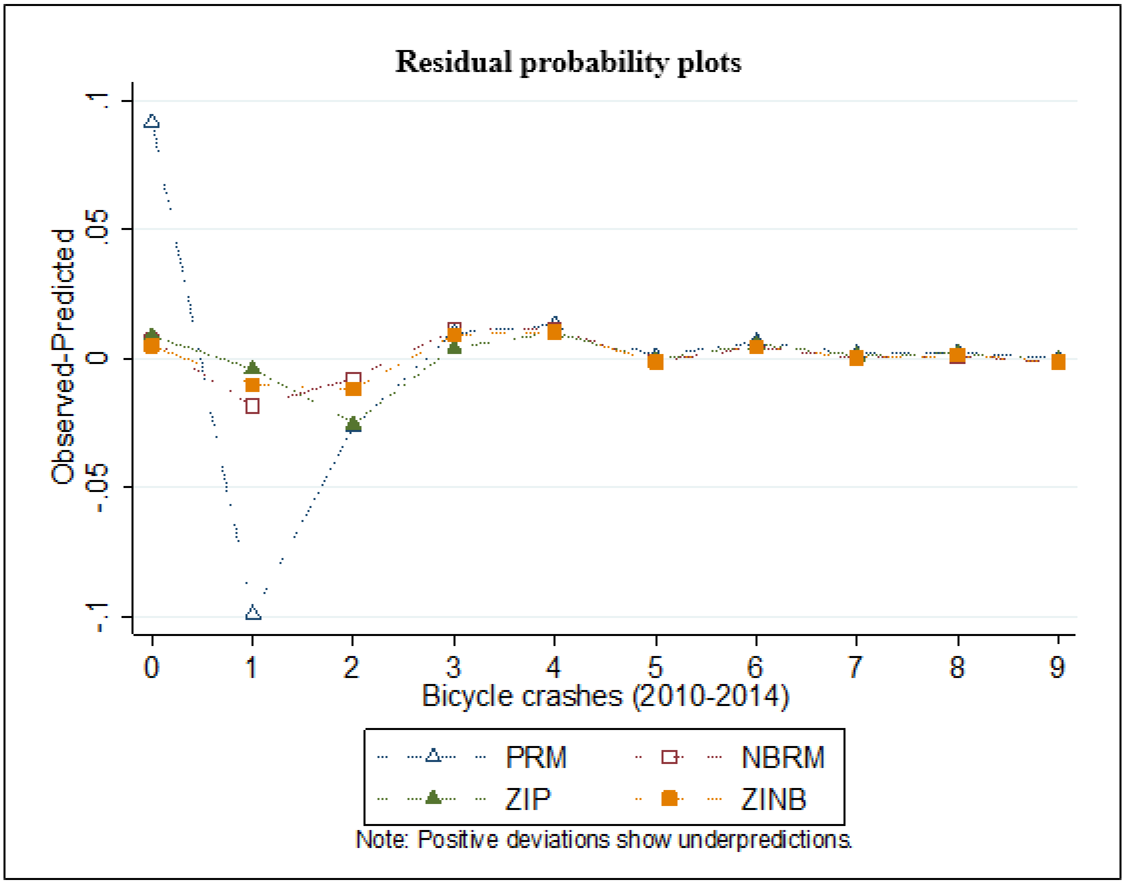


Figure 28 Within-Sample Residual Probability Plot Comparing Bicyclist SPFs

5.5.3 Cross Validation

Residual probability plots in Figure 29 indicated less precision of both models in predicting out-of-sample bicyclist-related crashes. ZIP had a better predictive performance as compared to NBRM and ZIP. Upon comparing Root Mean Square Error (RMSE) as shown in Figure 30, zero-inflated models had a better performance than non-inflated models.

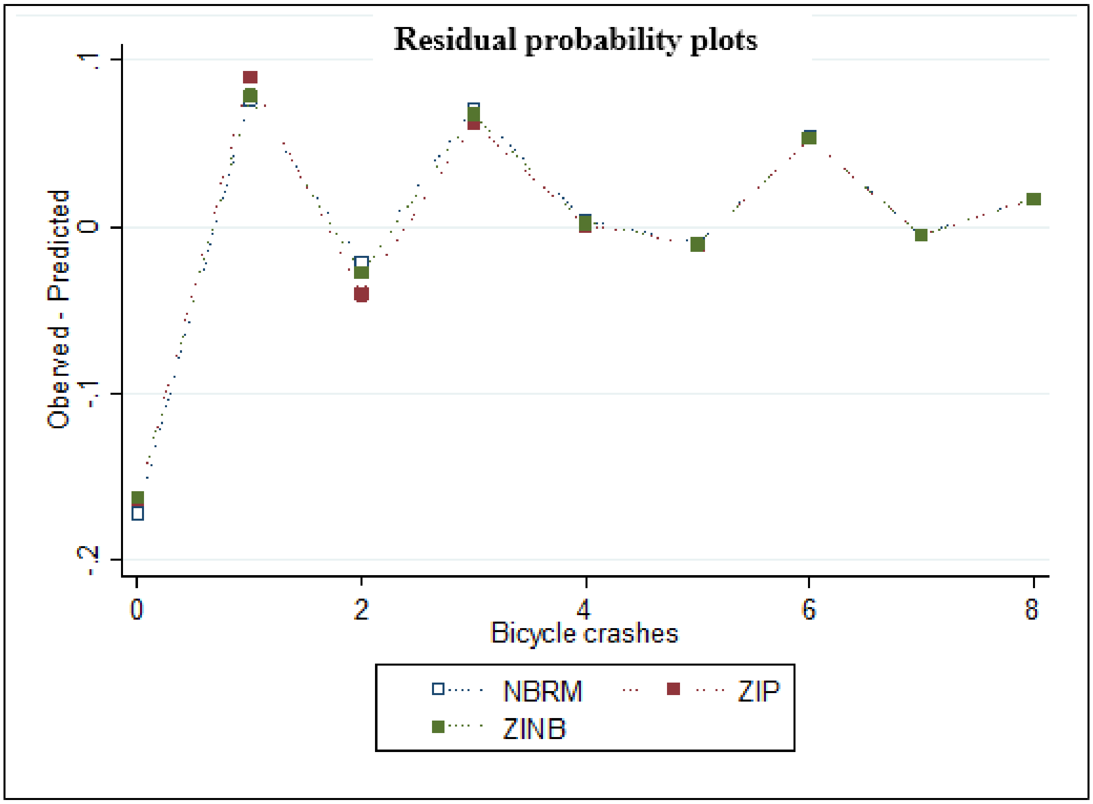


Figure 29 Cross Validation Residual Probability Plot Comparing Bicyclist SPFs

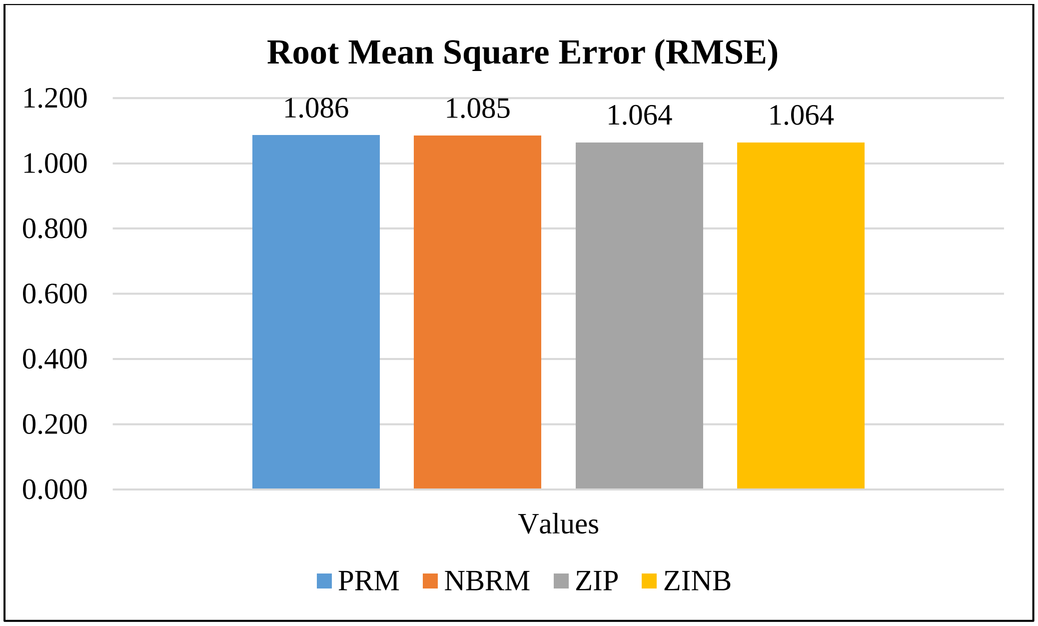


Figure 30 Mean Square Error from Bicyclist Cross Validation Data

5.5.4 *Final Proposed Bicyclist Safety Performance Function*

Therefore based on the assessment of goodness of fit measures it was concluded that Negative Binomial model seems to slightly outperform other count models. The final equation for Bicyclist SPF is provided below:

$$Number\ of\ Bike\ crashes\ per\ five\ years\ = e^{(-2.114+0.087 aadt_{maj}+3.561 aadt_{min}-2.144 bikelevel)}$$

Where

$aadt_{maj}$ = AADT in the major approach in thousands

$aadt_{min}$ = AADT in the minor approach in thousands

$bikelevel$ = Bicycle level score

6 Conclusions

This project established the methodology for developing statewide safety performance function for bicyclist and pedestrian. Specific focus was on urban intersections in Michigan connecting collector and arterial roads. Proper sampling procedure was needed so as to come up with the unbiased sample representative of all urban intersections in Michigan. Stratified random sampling was selected as the sampling strategy. All urban intersections in Michigan that were put into strata of similar characteristics. These characteristics were established by the parameter that were available at statewide level. These were National Function Classification (NFC) of the roadway forming an intersection, intersection type (three legged or four legged intersection), Urban population and number of non-motorized crashes per intersection in five years. A total of 72 strata were created from which a sample intersections were selected for developing SPFs.

Due to lack of pedestrian and bicycle volume counts at intersections it was necessary to develop a reliable proxy exposure measure. Factor analysis was used to develop pedestrian and bicycle level score using variable that are readily available at statewide level. Latent bicyclist level score, a proxy measure of bicyclist volume was found to increase with the presence of bicycle facility which includes bike lanes and sidewalks, increase in percentage of people below poverty level, increase population density, lower speed limit in major and minor approach and increase in proportion of commercial land use by area in a given census block group were the intersection is situated. Pedestrian level score was manifested by the increase in percentage of people using the public transit in a given block group where the intersection was situated, population density, percentage of household below poverty level, number of workers commuting to their working places by foot per square mile, walk score index, proportion of commercial land use and presence of pedestrian facility separated from the roadway.

During model development comparison was made across all potential count models that could fit the data. Appropriate goodness of fit tests and cross validation techniques were used in selecting the model with the best fit. Zero Inflated Poisson Model (ZIP) and Zero Inflated Negative Binomial Model (ZINB) were used as the final count model for pedestrian and bicycle SPFs respectively.

7 Bibliography

Aggarwal, Y. P. (1988). Better sampling: concepts, techniques, and evaluation. Stosius Inc/Advent Books Division.

Beck, L. F., Dellinger, A. M., & O'neil, M. E. (2007). Motor vehicle crash injury rates by mode of travel, United States: using exposure-based methods to quantify differences. American Journal of Epidemiology, 166(2), 212-218.

Chu, X. (2003, February). The fatality risk of walking in America: A time-based comparative approach. In Walk 21 (IV) Conference Proceedings.

Davis, S., King, E., Robertson, D., Mingo, R., & Washington, J. (1987). Measuring pedestrian volumes and conflicts. Volume I. Pedestrian volume sampling. Final report (No. FHWA/RD-88/036).

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. Practical Assessment, Research & Evaluation, 14(20), 1-11.

Dolatsara, H. A. (2014). Development of Safety Performance Functions for Non-Motorized Traffic Safety. Master’s Thesis, Western Michigan University.

Greene-Roesel, R., Diogenes, M. C., & Ragland, D. R. (2010). Estimating Pedestrian Accident Exposure (No. UCB-ITS-PRR-2010-32).

Hedlund, J., & North, H. S. (2000). NHTSA/FHWA Pedestrian and Bicycle Strategic Planning Research Workshops. Highway Safety North, Ithaca, NY.

Hilbe, J. M. (2014). Modeling count data. Cambridge University Press.

Jonsson, T. (2013). Safety performance models for pedestrians and bicyclists. In 16th International Conference Road Safety on Four Continents. Beijing, China (RS4C 2013). 15-17 May 2013.

Kim, S. (2003). Analysis of elderly mobility by structural equation modeling. Transportation Research Record: Journal of the Transportation Research Board, (1854), 81-89.

Knoblauch, R., Pietrucha, M., & Nitzburg, M. (1996). Field studies of pedestrian walking speed and start-up time. Transportation Research Record: Journal of the Transportation Research Board, (1538), 27-38.

Kwigizile, V., Mulokozi, E., Xu, X., Teng, H. H., & Ma, C. (2014). Investigation of the impact of corner clearance on urban intersection crash occurrence. Journal of transportation and statistics, 10(1), 35-48.

Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis & Prevention, 37(1), 35-46.

McArthur, A., Savolainen, P., & Gates, T. (2014). Spatial Analysis of Child Pedestrian and Bicycle Crashes: Development of Safety Performance Function for Areas Adjacent to Schools. Transportation Research Record: Journal of the Transportation Research Board, (2465), 57-63.

Minikel, E. (2012). Cyclist safety on bicycle boulevards and parallel arterial routes in Berkeley, California. Accident Analysis & Prevention, 45, 241-247.

Molino, J. A., Kennedy, J. F., Inge, P. J., Bertola, M. A., Beuse, P. A., Fowler, N. L., ... & Do, A. (2012). A Distance-Based Method to Estimate Annual Pedestrian and Bicyclist Exposure in an Urban Environment (No. FHWA-HRT-11-043).

Nordback, K., Marshall, W. E., & Janson, B. N. (2014). Bicyclist safety performance functions for a US city. Accident Analysis & Prevention, 65, 114-122.

Oh, J. S., Kwigizile, V., Van Houten, R., McKean, J., Abasahl, F., Dolatsara, H., ... & Clark, M. (2013). Development of Performance Measures for Non-Motorized Dynamics (No. RC-1603). Michigan Department of Transportation, Lansing, MI.

Qin, X., & Ivan, J. (2001). Estimating pedestrian exposure prediction model in rural areas. Transportation Research Record: Journal of the Transportation Research Board, (1773), 89-96.

Raford, N., & Ragland, D. (2004). Space syntax: Innovative pedestrian volume modeling tool for pedestrian safety. Transportation Research Record: Journal of the Transportation Research Board, (1878), 66-74.

Ranaiefar, F., & Rixey, R. A. (2016). Bike Sharing Ridership Forecast using Structural Equation Modeling. In Transportation Research Board 95th Annual Meeting (No. 16-6573).

Santos, A., McGuckin, N., Nakamoto, H. Y., Gray, D., & Liss, S. (2011). Summary of travel trends: 2009 national household travel survey (No. FHWA-PL-ll-022).

Schneider, R., Diogenes, M., Arnold, L., Attaset, V., Griswold, J., & Ragland, D. (2010). Association between roadway intersection characteristics and pedestrian crash risk in Alameda County, California. Transportation Research Record: Journal of the Transportation Research Board, (2198), 41-51.

Schumacker, R. E., & Lomax, R. G. (2004). A beginner's guide to structural equation modeling. Psychology Press.

Schwartz, W., & Porter, C. (2000). Bicycle and pedestrian data: Sources, needs, and gaps. U.S. Department of Transportation, Bureau of Transportation Statistics.

Thomson, G. H. (1935). The definition and measurement of" g"(general intelligence). Journal of Educational Psychology, 26(4), 241.

Tobey, H. N., Knoblauch, R. L., & Shunamen, E. M. (1983). Pedestrian Trip Making Characteristics and Exposure Measures. Final Report. Federal Highway Administration, Office of Safety and Traffic Operations.

Turner, S., Wood, G., Hughes, T., & Singh, R. (2011). Safety performance functions for bicycle crashes in New Zealand and Australia. Transportation Research Record: Journal of the Transportation Research Board, (2236), 66-73.

Wang, K., & Qin, X. (2014). Use of structural equation modeling to measure severity of single-vehicle crashes. Transportation Research Record: Journal of the Transportation Research Board, (2432), 17-25.

Xie, K., Ozbay, K., & Yang, H. (2016). A Joint Analysis of Secondary Collisions and Injury Severity Levels Using Structural Equation Models. In Transportation Research Board 95th Annual Meeting (No. 16-0206).

8 Appendix

Table 19: List of all the data that were collected for modeling purpose

Description	Mean	Std. Dev.	Min	Max
Total number of bicycle crashes(2010-2014)	0.647	1.353	0	8
Total number of pedestrian crashes(2010-2014)	0.474	0.968	0	10
Average pedestrian crashes(crashes/year)	0.299	0.464	0	2
Average Bicycle Crashes(crashes/year)	0.317	0.513	0	2
Intersection type	3.701	0.458	3	4
Intersection type: Three leg	0.299	0.458	0	1
Intersection type: Four leg	0.701	0.458	0	1
AADT of the major approach	14664.0	8879.0	1388	57285
AADT of the minor approach	6696.3	5207.7	346	24716
Number of exclusive through lane in the major approach	1.534	1.564	0	8
Number of shared through-right turn lane in the major approach	0.876	0.848	0	2
Number of share through-left turn lane in the major approach	0.126	0.403	0	2
Number of shared through-right-left turn lane in the major approach	0.209	0.584	0	2
Number of shared left-right turn lane in the major approach	0.039	0.193	0	1
Number of exclusive right turn lane in the major approach	0.570	0.739	0	3
Number of exclusive left turn lane in the major approach	1.253	0.849	0	4
Number of lane for leaving traffic in the major approach	2.925	1.367	0	9
Presence of crosswalk in the major approach	0.611	0.488	0	1
Presence of median in the major approach	0.088	0.283	0	1
Presence of pedestrian facility in the major approach	0.773	0.419	0	1
Presence of pedestrian facility separated from traffic in the major approach	0.446	0.498	0	1
Presence of bike lane in the major approach	0.018	0.133	0	1
Presence of on-street parking in the major approach	0.031	0.173	0	1
One way indicator for the major approach	0.008	0.088	0	1
Number of exclusive through lane in the minor approach	0.853	1.314	0	8
Number of shared through-right turn lane in the minor approach	0.784	0.841	0	2

Number of share through-left turn lane in the minor approach	0.147	0.427	0	2
Number of shared through-right-left turn lane in the minor approach	0.312	0.688	0	2
Number of shared left-right turn lane in the minor approach	0.088	0.283	0	1
Number of exclusive right turn lane in the minor approach	0.515	0.713	0	3
Number of exclusive left turn lane in the minor approach	1.080	0.924	0	4
Number of lane for leaving traffic in the minor approach	2.356	1.140	1	8
Presence of crosswalk in the minor approach	0.580	0.494	0	1
Presence of median in the minor approach	0.064	0.246	0	1
Presence of pedestrian facility in the minor approach	0.742	0.438	0	1
Presence of pedestrian facility in the minor approach separated from roadway	0.392	0.489	0	1
Presence of bike lane in the minor approach	0.015	0.124	0	1
Presence of on-street parking in the minor approach	0.039	0.193	0	1
One way indicator for the minor approach	0.018	0.133	0	1
Control type: Traffic signal	0.742	0.438	0	1
Control type: Two way stop sign	0.216	0.412	0	1
Control type: All way stop sign	0.041	0.199	0	1
Control type: Stop sign(Two way and all way combined)	0.258	0.438	0	1
Signal configuration/arrangement: Diagonal	0.466	0.500	0	1
Signal configuration/arrangement: Box	0.276	0.447	0	1
No turn on red on the major approach	0.015	0.124	0	1
Protected left turn on the major approach	0.302	0.460	0	1
No turn on red on the minor approach	0.018	0.133	0	1
Protected left turn on the minor approach	0.291	0.455	0	1
Presence of crosswalk	0.647	0.479	0	1
Presence of median	0.131	0.338	0	1
Presence of pedestrian facility	0.794	0.405	0	1
Presence of pedestrian facility separated from traffic	0.577	0.495	0	1
Presence of bike lane	0.034	0.180	0	1
Presence of on-street parking	0.590	0.492	0	1
One way indicator for the major approach	0.049	0.216	0	1
Presence of pedestrian facility	0.026	0.159	0	1
National functional classification: Arterial: Arterial	0.521	0.500	0	1
National functional classification: Collector-Arterial	0.358	0.480	0	1
National functional classification: Collector-Collector	0.121	0.327	0	1
Speed limit on the major approach	43.144	9.038	25	70
Speed limit on the minor approach	35.180	8.781	20	55
Walk score index	35.188	24.828	0	94

Proportion of land use by census block : Commercial	0.252	0.281	0	1
Proportion of land use by census block: Industrial	0.063	0.171	0	1
Proportion of land use by census block: Institutional	0.078	0.159	0	1
Proportion of land use by census block: Outdoor recreation	0.036	0.136	0	1
Proportion of land use by census block: Residential	0.570	0.316	0	1
Proportion of land use by area: Commercial	0.146	0.283	0	1
Proportion of land use by area: Industrial	0.054	0.184	0	1
Proportion of land use by area: Institutional	0.030	0.117	0	1
Proportion of land use by area: Outdoor recreation	0.025	0.132	0	1
Proportional of land use by area: Residential	0.746	0.357	0	1
Means of transportation: Percentage of worker using cars in a given census block	94.216	8.341	34.96162	100
Means of transportation: Percentage of worker using public transport in a given census block	0.940	2.348	0	22.2
Means of transportation: Percentage of worker using bus in a given census block	0.931	2.321	0	22.2
Means of transportation: Percentage of worker using taxi in a given census block	0.055	0.512	0	9.2
Means of transportation: Percentage of worker using motorcycle in a given census block	0.217	0.615	0	5.9
Means of transportation: Percentage of worker biking in a given census block	0.044	0.187	0	2.4
Means of transportation: Percentage of worker walking in a given census block	1.453	4.595	0	40.8
Bicyclist commuter density for a census block	0.596	2.716	0	29.8
Walking commuter density for a census block	34.638	145.510	0	1671.9
Percentage of household above poverty level in given census block	79.934	25.587	0.468543	100.0
Percentage of household below the poverty level in a given census block	13.128	14.032	0	83.7
Percentage of whites in a census block	79.409	26.327	0	100.0
Percentage of blacks in in a census block	13.446	25.636	0	99.7
Percentage of Indian Alaska in a census block	0.385	1.591	0	28.1
Percentage of Asian in 0.25 mile in a census block	3.167	5.081	0	35.6
Population density in a census block	412.574	366.545	0	2384.9