Developing Michigan Pedestrian and Bicycle Safety Models

Robert C. Hampshire
Lisa J. Molnar
Alex Cao
Yiming Cai
Xiao Li
Tayo Fabusuyi
The University of Michigan Transportation Research Institute

May 7, 2018

Final Report

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.
SPR - 1651

2. Government Accession No.
N/A

3. Recipient's Catalog No.

4. Title and Subtitle
UTC - Developing Michigan Pedestrian and Bicycle Safety Models

5. Report Date
May 7, 2018

6. Performing Organization Code
N/A

7. Author(s)
Robert C. Hampshire , Lisa J. Molnar, Alex Cao, Yiming Cai
Xiao Li, Tayo Fabusuyi

8. Performing Organization Report No.
N/A

9. Performing Organization Name and Address
University of Michigan

10. Work Unit No.
N/A

11. Contract or Grant No.
Contract 2016-0068 Z1

12. Sponsoring Agency Name and Address
Michigan Department of Transportation (MDOT)
Research Administration
8885 Ricks Road
P.O. Box 33049
Lansing, Michigan 48909

13. Type of Report and Period Covered
Final Report, 3/1/2016 to 12/31/2017

14. Sponsoring Agency Code
N/A

15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration. MDOT research reports are available at www.michigan.gov/mdotresearch.

16. Abstract
Reducing the number of pedestrian and bicycle-involved crashes, injuries, and fatalities is a priority nationwide, including Michigan. It is important to identify those locations and attributes associated with locations that are most prone to such crashes. Relying on observed crashes or hotspot analysis can be misleading due to statistically anomalies or not properly accounting for pedestrian exposure and other risk factors. This project created a pedestrian and bicycle risk score, based on mapping crashes and the risk characteristics, for a defined area or network for the entire state of Michigan. This report describes how the project team developed the risk scores and corresponding non-motorized exposure estimates.

17. Key Words
non-motorized users, bicyclist, pedestrian, risk models, risk score, safety performance functions, highway safety manual, Empirical Bayes, geographic information systems

18. Distribution Statement
No restrictions. This document is also available to the public through the Michigan Department of Transportation.

19. Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
Unclassified

21. No. of Pages
65

22. Price
N/A

Abstract Page

Despite existing research and improvements, pedestrian and bicycle-involved crashes, injuries, and fatalities have remained relatively stable in Michigan. Because of this, it is important to better identify behaviors and those locations and attributes associated with locations that are most prone to such crashes.

However, many state agencies are not able to systematically identify and compare areas of high risk for pedestrians. Relying on observed crashes or hotspot analysis can be misleading due to statistically anomalies or not properly accounting for pedestrian exposure and other risk factors.

A key objective of this project was to create a risk score, based on mapping crashes and the risk characteristics, for a defined area or network for the entire state of Michigan. This report describes how the project team developed the risk scores and corresponding non-motorized exposure estimates.

Acknowledgements and disclaimer

This report was developed through the support of the Center for Advancing Transportation Leadership and Safety (ATLAS-center.org), a University Transportation Center that is a collaboration between the University of Michigan Transportation Research Institute (UMTRI) and the Texas A&M Transportation Institute (TTI). The Center is sponsored by the U.S. Department of Transportation's Office of the Assistant Secretary for Research and Technology (OST-R) through Grant Number DTRT13-G-UTC54.

The authors would like to thank Carissa McQuiston and Dean Kanitz from the Michigan Department of Transportation for their input and support for this project. Alex Cao was the GIS consultant for this project. Tian Tian and Yu-Hung Kuo provided support for the literature review. Brian Hilbrands developed an earlier version of the pedestrian risk score. Thank you all.

Contents

List of Tables

List of Figures

Executive Summary

This reports describes efforts to develop both a pedestrian and bicycle risk assessment model for the entire State of Michigan. This report addresses non-motorized[1] safety risk by introducing a practice ready crash risk assessment methodology that incorporates a model of exposure.

Our objective was the development of a robust, statewide pedestrian and bicycle crash risk score. To accomplish this, we developed statistical and data science methods to help identify the places in need of countermeasure and the factors that contribute to pedestrian and bicycle crashes.

However, pedestrian and bicycle crashes occur relatively infrequently over time and space. Therefore, safety analysis based on observed crashes is an inadequate and unsatisfactory way to proceed. Many believe that they are nearly impossible to predict. On the contrary, our hypothesis is that it is possible to accurately predict the time and place of non-motorized crashes.

A key hypothesis of this report is that it is possible to identify locations in need of countermeasures for pedestrian and bicycle crashes using a "risk score" that leverages information beyond observed crashes.

We leveraged the empirical Bayes (EB) methods from the Highway Safety Manual in combination with a model of non-motorized exposure to create the risk score. With this framework, we developed fine-scale risk score and exposure estimates for the entire State of Michigan. The unit of analysis is 400m by 400m. Through a limited validation study, we show that the risk score can reliability predict the most dangerous areas for pedestrians and bicyclist.

We results are available in 2 output formats: GIS files and the webtool. The GIS tool allows users to interactively view the results in ArcGIS. The webtool allows the user to display and interact with the results, without any special software, from any web browser including on mobile devices.

[1] Throughout this report, we use the term "non-motorized" to include both pedestrian and bicyclist.

1 Introduction

This reports describes our efforts to develop a both a pedestrian and bicycle risk assessment model for the entire State of Michigan. This report addresses non-motorized[2] safety risk by introducing a practice ready crash risk assessment methodology that incorporates a model of exposure.

Walking and bicycling promote healthy and socially connected neighborhoods and positively impact the local economy. Despite all of these benefits, bicyclists and pedestrians are at risk and lack protection in a crash with motor vehicles. In the State of Michigan, 38 bicyclists were killed and an additional 1,526 bicyclists were injured in motor vehicle crashes in 2016. The 38 bicyclist fatalities in 2016 and 33 bicyclist fatalities in the previous year 2015 were the highest numbers since 2000. For pedestrians, 165 pedestrians were killed and 1,852 injured on Michigan roads in 2016 (Michigan State Police, 2017). Bicyclists in the United States are 12 times more likely than car occupants to get killed (72 vs 6 fatalities per billion kilometers) in a crash, and bicyclists in the United States are twice as likely to get killed as bicyclists in Germany and over three times as likely as bicyclists in the Netherlands (Pucher and Dijkstra, 2003).

The aim of this project is to reduce the number of bicycle and pedestrian crashes in Michigan. To accomplish this, we developed statistical and data science methods to help identify the places in need of countermeasure and the factors that contribute to pedestrian and bicycle crashes. The Federal government and other states are currently undertaking similar efforts. They are adopting and advocating data driven safety analysis (DDSA) *(8)*. The stated purpose of DDSA is

“to analyze crash and roadway data to predict the safety impacts of highway projects allows agencies to target investments with more confidence and reduce severe crashes on the roadways”(p.2)**.**

There are two main DDSA approaches: predictive and systemic. Predictive approaches seek to develop analytic tools to estimate the number of crashes in a specific location for certain types of roads and circumstances. Systemic approaches seek to identify the common set of underlying risk factors across a large swath of crash types. This project used the predictive approach for safety analysis.

We embarked on a research project to assess the crash risk to pedestrians and bicyclists. Specifically, the objective was the development of a robust, statewide pedestrian and bicycle crash risk score. The department’s non-motorized safety engineers wanted a measure of risk that goes beyond merely counting the

[2] Throughout this report, we use the term “non-motorized” to include both pedestrian and bicyclist.

number of vehicle-pedestrian or vehicle-bicycle crashes in a given area. They wanted to understand the risk factors underlying pedestrian crashes. It is in this context that the risk score methodology presented in this report was developed. Unfortunately, the Highway Safety Manual (HSM)[3], intended to offer guidance to states, provides an unsatisfactory solution to measure pedestrian risk. In the HSM, the safety performance functions (SPFs) for vehicle-pedestrian have only vehicle and road geometry factors, but no pedestrian factors. Even SPFs specifically calibrated for Michigan *(34)* do not consider pedestrian factors such as exposure. These factors provided the rationale for this present study.

The HSM provides the set of best practices for predictive safety analysis. Specifically, section 2c of the HSM presents predictive methods to estimate the crash frequency for many road types and circumstances. The prescribed method to predict vehicle-pedestrian and vehicle-bicycle crashes utilize information about the vehicle and roadway characteristics but does not include information about pedestrian and bicycle exposure. However, exposure is a critical element in prediction, given that it is difficult to have a vehicle-pedestrian crash in the absence of a pedestrian.

Recognizing this limitation, we developed a pedestrian and bicycle crash risk score that combines empirical Bayes methodology used by the HSM and the model of demand (MoPED) developed by Clifton et al. *(1)*. A key contribution of this report is demonstrating how to scale MoPED and the risk score to a statewide level. We accomplish this by populating MoPED with imputed statewide multi-way tables of the household variables of interest using both the US Census Public Use Microdata Sample (PUMS) and the American Community Survey (ACS).

1.1 Background

Pedestrian and bicycle-involved crashes are an emphasis area of the traffic safety community. Given the emphasis of non-motorized transportation at the local, state and national level, there has been additional interest in further understanding pedestrian and bicyclist safety issues. Despite existing research and improvements, pedestrian and bicycle-involved crashes, injuries, and fatalities have remained relatively stable in Michigan. Because of this, it is important to better identify behaviors and those locations and attributes associated with locations that are most prone to such crashes.

1.1.1 Objectives

Given the overarching importance of non-motorized safety particularly in urban areas, many cities and metropolitan planning organizations are devoting considerable resources towards addressing it. However, many state agencies

[3] Manual, H. S. (2010). American association of state highway and transportation officials (AASHTO). *Washington, DC*, *10*.

are not able to systematically identify and compare areas of high risk for people using non-motorized modes. Relying on observed crashes or hotspot analysis can be misleading due to statistical anomalies or not properly accounting for the intensity of travel. From the outset, this research project had six primary objectives:

- Document and learn from existing research on modeling/mapping pedestrian and bicycling safety areas.
- Gather new data on how to model/map pedestrian and bicycling crashes in Michigan.
- Analyze these data in order to produce a model/mapping tool that best determines locations in Michigan that could benefit from pedestrian and bicycling crash countermeasure installations.
- The methodologies/tool(s) will be able:
    a. Visualize both pedestrian and bicycle crashes
    b. Provide a risk score (based on mapping crashes and the risk characteristics mentioned above) for a defined area or network (with crash summaries)
    c. Provide risk scores across the entire state (with crash summaries)
    d. Provide a process that results in viable output formats (GIS oriented: .kml, .dbf, .csv) and the process to add data and update regularly.
- Report out methodology and results of this analysis.
- Produce a dataset for use in a GIS tool.

1.1.2 Scope

This is a statewide analysis.

1.2 Statement of hypotheses

This research is intended to determine locations in Michigan that could benefit from pedestrian and/or bicycle crash countermeasure installations. To accomplish this, we developed a method to produce a model/mapping tool that best determines locations that could benefit from pedestrian and/or bicycle crash countermeasures.

Pedestrian and bicycle crashes occur relatively infrequently over time and space. However, these crashes are often fatal when they occur. Therefore, safety analysis based on observed crashes is an inadequate and unsatisfactory way to proceed.

A key hypothesis of this report is that it is possible to identify locations in need of countermeasures for pedestrian and bicycle crashes using a "risk score" that leverages information beyond observed crashes.

2 Literature Review

To put this report into context, we review the relevant research literature on pedestrian and bicycle crash risk assessment.

2.1 Review of previous research

Current practice defines risk as the probability of a crash occurring given exposure to potential crash events. This definition is intuitive; however, measuring exposure to potential crash events is difficult. We first survey the various approaches to risk, and then we discuss ways to measure and incorporate exposure into a pedestrian risk score.

An approach suggested by Raford and Ragland *(9)* is to define pedestrian risk as the number of observed crashes in a given geographical area divided by the daily pedestrian volume. The authors make the case that the level of pedestrian risk can be defined as the annual number of pedestrian-involved crashes divided by exposure, which is represented by the annual estimated pedestrian volume *(9-12)*.

Given that pedestrian crashes occur infrequently, the aforementioned methods assign a risk of 0 to many areas. Traffic safety engineers have known for some time that evaluating risk solely by counting observed number of crashes simultaneously leads to an overestimation of risk in high risk areas and an underestimation in low risk areas. To address this issue, Hauer et al. *(13)* proposed the Empirical Bayes (EB) method for the estimation of the level of risk that is capable of correcting for regression-to-the mean bias.

The EB method combines a model of predicted crash frequency and observed crash frequency to obtain a pedestrian risk score. The predicted number of crashes is determined by a parametric regression function referred to as SPF. SPF predicts the crash frequency for roads that are similar to the one under investigation. Traditionally, SPF was developed using a negative binomial regression model to predict the number of crashes for a particular site based on known information, such as annual average daily traffic (AADT) and road geometry *(14-15)*. Unfortunately, SPFs for pedestrian-vehicle crashes do not include a measure of pedestrian volume. This report contributes to the literature by incorporating a measure of pedestrian exposure into an SPF model.

Pedestrian exposure is defined as the measure of the number of potential opportunities for a pedestrian-involved crash to occur *(16)*. However, in contrast to vehicular behavior, pedestrian trips are of a different variety in terms of trip purposes, and their route choices are less well defined *(17)*. Therefore, new methods that adequately reflect the context specific nature of pedestrian crashes are required for estimating the volume of pedestrian trips.

There are many studies regarding pedestrian exposure models using built environment, socioeconomic characteristics, demographic characteristics, and other factors to predict pedestrian volume *(18-19)*. Notably, Clifton et al. *(1,20)* introduce an innovative model called the model of pedestrian demand (MoPED) that modifies the conventional four-step modelling (FSM) to better represent pedestrian walking behavior at a uniform 80-meter-by-80-meter raster grid cell, called pedestrian analysis zone (PAZ). The redrawn boundary at a more granular level enables the prediction of finer pedestrian behavior at a microscopic level than traffic analysis zone (TAZ). Mode choices for travelers are highly related to the built environment and socioeconomic factors *(21)*. To estimate the amount of walking trips generated by a household in each PAZ, MoPED uses a binary logit model based on socio-demographic characteristics of travelers and built environment to split walking trips from all person trips estimated by conventional travel forecasting model.

Clifton's study further applies a multinomial logit destination choice model to distribute those walking trips to destinations within the aggregations of 25 (5x5) PAZs. It is known that distance to destinations plays a determinant role in destination choice models *(22)*. Additionally, attractiveness, pedestrian supports, pedestrian barriers and traveler characteristics also play key roles in the destination choice model. The destination choice model provides a linkage between built environment and pedestrian destination choices *(23)*.

Finally, the trip generation and destination models are combined to predict the probability of a walking trip traveling from one zone to another. The output of MoPED is a predicted measure of daily pedestrian volumes at a very fine geographic resolution (i.e. 80 meters by 80 meters). The key inputs to the method are socio-demographic variables, trip generation tables, detailed employment data, built environment data, and a travel survey.

Nevertheless, MoPED requires trip generation tables typically created by a region's metropolitan planning organization (MPO), which limits the statewide application of MoPED because trip generation tables are often not available in many places (i.e. small to medium sized cities without membership with an MPO). One of contributions of this report is the modification of MoPED by estimating walking trip generation directly from a statewide travel survey without the need for a general trip generation tables. We generate walk trip generation tables by creating synthetic households using iterative proportional fitting.

The MoPED regression models for predicting walking trips require disaggregate household and traveler demographics data as input. Since disaggregated data of household and traveler characteristics are typically not available at more granular geographic levels, a method for generating synthetic population data is needed. The Iterative Proportional Fitting (IPF) approach is applied to construct a population synthesizer for the purpose of obtaining population estimates in a generic way *(24)*. The IPF process requires both aggregate data and disaggregate data as inputs. Specifically, the process starts by assigning initial values retrieved from disaggregate data, and then proceeds by iteratively updating the estimates based on the marginal totals of the same list of characteristics obtained from aggregate data. The disaggregate data represent a sample of households or individuals with values for a list of characteristics, and the most popular data source is the Public Use Microdata Sample (PUMS) *(25-26)*. The PUMS consists of a 5 percent representative sample of people and housing units from contiguous geographic units containing no fewer than 100,000 people each, named Public Use Micro Area (PUMA) *(27)*. As for aggregate data, it is normally drawn from Census summary tables at different aggregate level (e.g., Census Block Group or Census Tract) based on the specific needs of the researcher.

2.2 Summary of state-of-the-art

We conducted a thorough review of the research literature to identify the evidence base for including factors into the risk score assessment. A detailed summary and categorization of these factors are available in the Appendix. We also investigated the potential of using naturalistic driving study (NDS) data to estimate non-motorized travel demand. The conclusions from this investigation are also in the Appendix.

3 Methodology

The project team created a methodology that facilitates the identification of locations in Michigan that could benefit from pedestrian and/or bicycle crash countermeasure installations. To accomplish this, we developed statistical regression models of both pedestrian and bicycling crashes by combining a variety of data sources, as noted earlier and elaborated on in Section 3.1.

The methodology is general, and applies equally to both pedestrian and bicycle risk assessment. In order to reduce redundant statements that apply to both the pedestrian and bicycle model, we first present the method in the context of the pedestrian model (Sections 3.2-3.3). Section 3.4 delineates the deviations from the general method as it applies to bicycle risk assessment model.

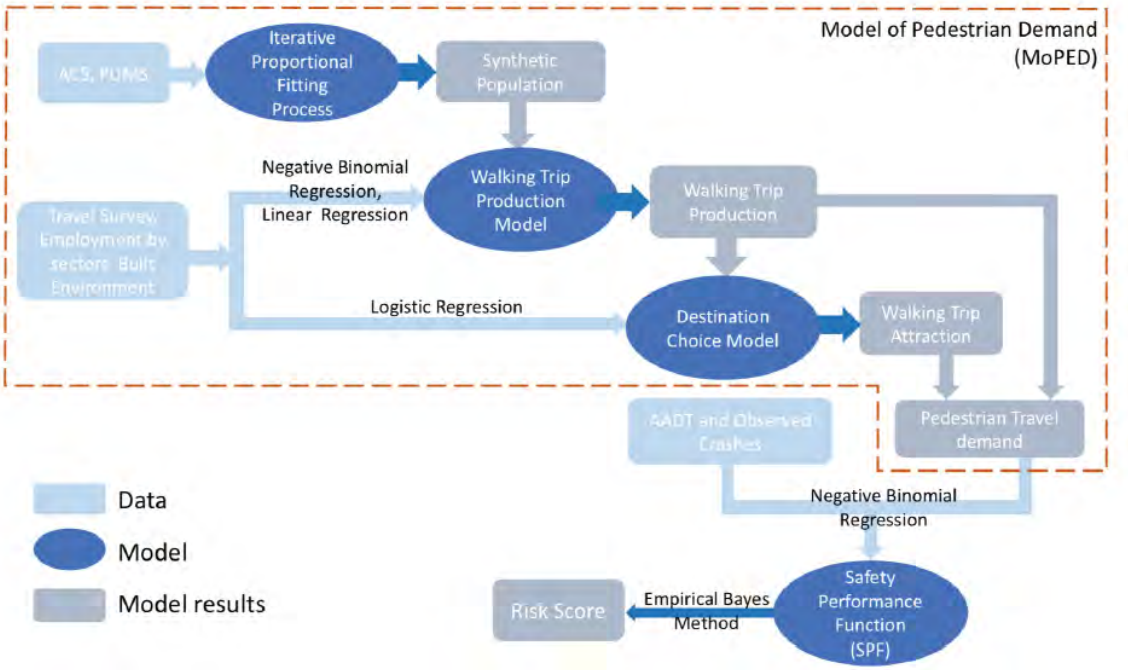

Figure 1 Project methodology for non-motorized crash risk assessment

Figure 1 provides a panoramic view of the study methodology that applies for both estimating pedestrian and bicyclist risk scores. It shows the essence our methodology, which is to combine the EB methods from the Highway Safety Manual with a model of non-motorized exposure. As noted earlier, we believe the number of crashes involving non-motorized users does not solely determine the risk level. We argue that non-motorized exposure should also play an important role in influencing the risk level. Given the importance of exposure in estimating risk, we first provide a detailed discussion of our exposure model (Section 3.2) and then discuss how exposure is integrated into the overall risk scores (section 3.3). The bulk of Section 3 focuses on pedestrian exposure and risk scores. We develop the bicycle risk scores using the same methodology with some notable modifications that we describe in Sections 3.4

3.1 Data Sources

One of our objectives in deciding the input data sources for the risk assessment was to select data that are 1) widely available, and 2) updated frequently.

The key inputs to the framework are: a statewide travel survey called MI Travel Counts, the American Community Survey (ACS), the public use microdata sample (PUMS), roadway geometry data as well as statewide crash records.

ACS is the model's main source of block-group level socio-demographic and household variables including household size, number of vehicles, and number of workers. The framework in Figure 1 requires disaggregate household characteristics in order to be applied statewide. We utilize the PUMS which contains the disaggregate responses of approximately 5% of households to the U.S. census. Vehicle-pedestrian crashes are derived from Michigan police crash reports from 2005-2015 *(30)*.

3.1.1 Michigan Traffic Crash Facts

Michigan Traffic Crash Facts (MTCF) is a web-based depository of official police-reported crash data for the state of Michigan (see https://www.michigantrafficcrashfacts.org/). The website contains a publication section with crash data statistics dating back to 1952, as well as a data query tool, which allows public users to search the dataset and display results in variety of formats (e.g., maps, tables, lists, charts).

3.1.2 American Community Survey (ACS)

ACS is a nationwide yearly census administered by the U.S. Census bureau[4]. It is the premier data source for socio-economic information about households and individuals. The ACS is the model's main source of block-group level socio-demographic and household variables including household size, number of vehicles and number of workers.

3.1.3 ACS Public Use Microdata Sample (PUMS)

We utilized the ACS PUMS which contains the disaggregate responses of approximately 5% of households to the U.S. census.

3.1.4 MI Travel Counts III (MTC III)

MI Travel Counts III (MTC III) is a travel survey conducted in 2015 by the Michigan Department of Transportation (MDOT)[5]. MTC III contains samples of 16,276 households across the state reporting their weekday trips. Each surveyed trip includes mode, origin and destination, travel time, activities at each destination, and also household characteristics.

Figure 2 MI Counts Traveler and Household Data

[4] https://www.census.gov/programs-surveys/acs/
[5] http://www.michigan.gov/mdot/0,1607,7-151-9615_51690---,00.html

3.1.5 Reference USA

We rely on ReferenceUSA® [6]for the location of businesses and employment data across the state *(28)*.

3.1.6 Highway Performance Monitoring System (HPMS)

The roadway geometry data are taken from the Highway Performance Monitoring System (HPMS) *(29)*[7]. These data include number of lanes, road width, and segment length.

3.1.7 Unit of Analysis

Our basic unit of analysis is called a Pedestrian Analysis Zone (PAZ)[8], which is 400 meters by 400 meters. The state of Michigan is comprised of nearly 95,000 PAZs. The following graph compares the scale between PAZs with typical urban streets. This unit of analysis more appropriate for the shorter travel distances of non-motorized travel compared to the longer travel distances vehicle travel. The term PAZ is coined in contrast to the well-known traffic analysis zone (TAZ).

Figure 3 Pedestrian Analysis Zone (PAZ)

---

[6] www.referenceusa.com
[7] https://www.fhwa.dot.gov/policyinformation/hpms.cfm
[8] Clifton et al., call this unit a Super PAZ. However, we simply use PAZ for ease of exposition.

3.2 Risk Score

Here we describe how to compute the risk score for each PAZ. The basic idea of the risk score is to make joint use of the observed number of crashes between non-motorized users and vehicles (historical data), and the predicted number of pedestrian-vehicle crashes for similar geographic areas. The pedestrian risk score for a PAZ is the expected number of crashes. The empirical Bayes estimate of the risk score for a PAZ is

$$R_{paz} = w \cdot \mu_{paz}^{spf} + (1-w) \cdot y_{paz}^{spf} \tag{1}$$

where y_{paz}^{obs} is the number of observed crashed in the target PAZ , μ_{paz}^{spf} is the predicted number of crashes, overdispersion parameter ϕ and $w$is

$$w = \frac{1}{1 + \frac{\mu_{paz}^{spf}}{\phi}} \tag{2}$$

is between 0 and 1.

We subsequently develop a SPF for the predicted number of crashes. Our SPF includes not only the AADT, but also the exposure and an interaction term between vehicle AADT and pedestrian exposure. The resulting negative binomial regression model specification is

$$\mu_{paz}^{spf} = \exp\left(a + b \cdot \ln\left(AADT_{paz}\right) + c \cdot \ln\left(\text{exposure}_{paz}\right) + d \cdot \ln\left(\text{exposure}_{paz}\right) \cdot \ln\left(AADT_{paz}\right)\right) \tag{3}$$

The regression coefficients a, c, b, d and the overdispersion parameter ϕ are calculated via maximum likelihood estimation. In the next couple of sections, we provide more details on the development of the exposure measure used in the SPF.

This risk assessment framework leverages data from several categories: built environment, characteristics of the people traveling, crash data and roadway features.

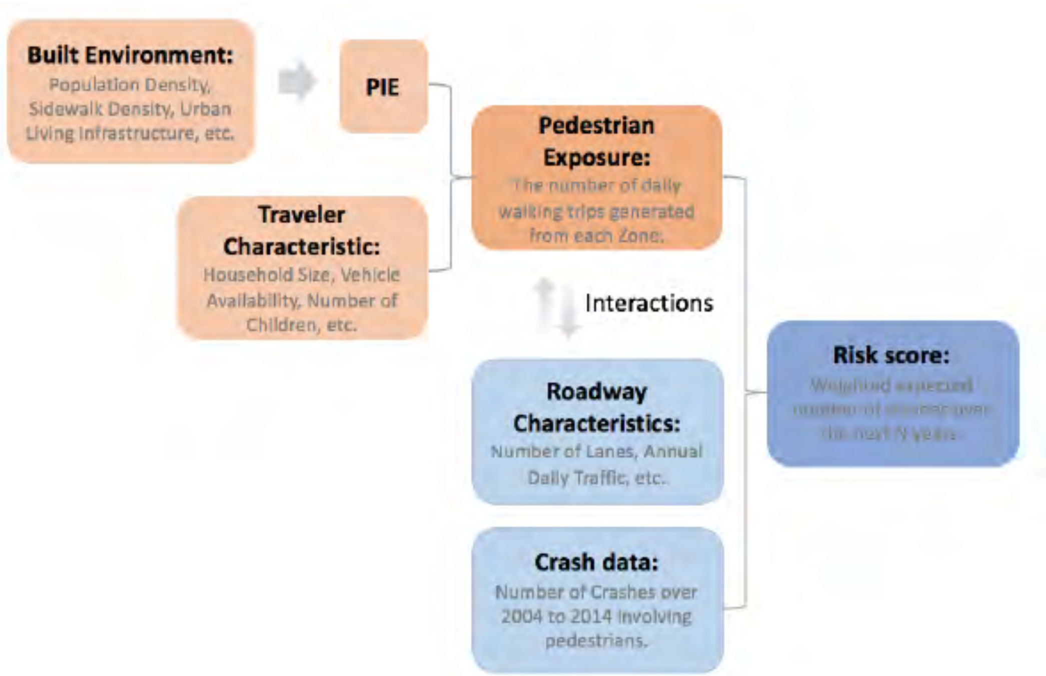

Figure 4 We compose data from diverse sources to develop the risk score.

3.3 Non-motorized Exposure Model

Our model of non-motorized exposure borrows from the MoPED framework of Clifton et al. *(1)* with a few notable exceptions discussed here. First, we deviate from MoPED and the traditional 4 step method (see Figure 2) in that we do not estimate mode split. Instead, we generate trips directly from the statewide travel survey (see Figure 3). Second, we extend MoPED so that it can be applied statewide by estimating disaggregate Block Group household characteristics via iterative proportional fitting. This allows us to generate trips statewide, even in areas that do not currently have an MPO that creates trip generation tables.

Below, we summarize the key aspects of our modified MoPED framework along the following dimensions: unit of analysis, pedestrian index of environment (PIE), trip generation, creating a synthetic population and pedestrian destination choice.

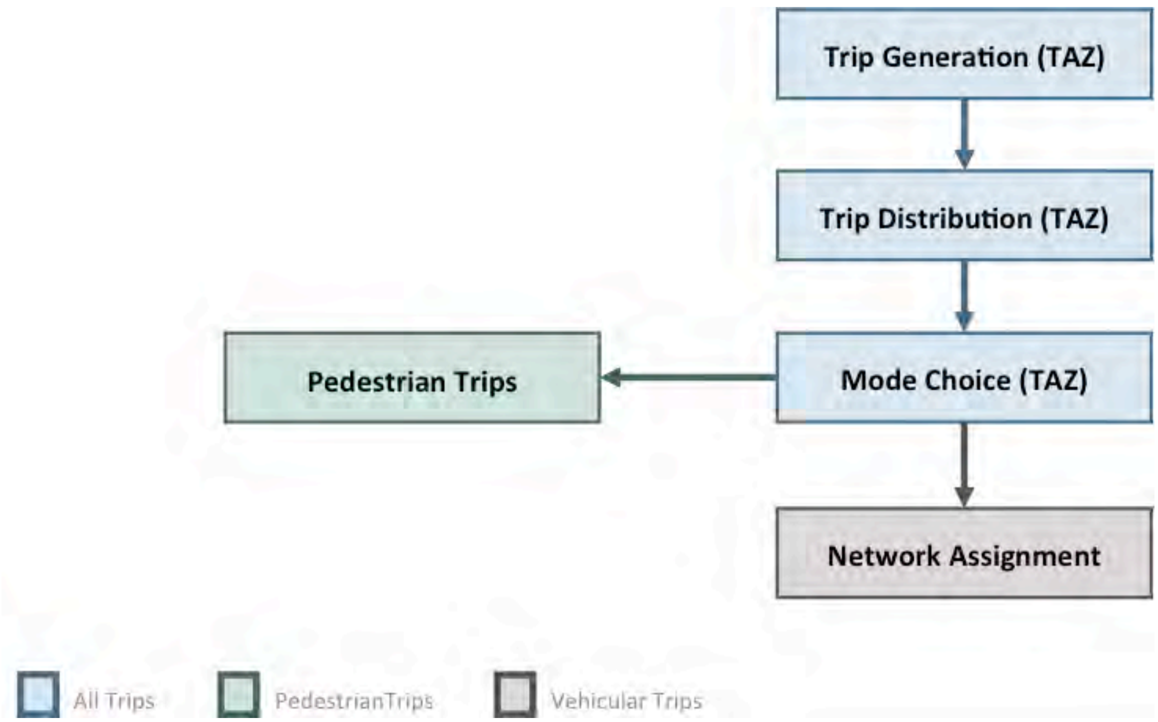

Figure 5 Traditional Four Step Method (FSM)

Figure 6 Modified Four Step Method

3.3.1 Built Environment Factors

To measure the impact of built environment on travel choices, our model employs a statewide travel survey data from MTC III and built environment data from various sources. Following Clifton et al.*(23),* we categorize the data into five groups as shown in Figure 6 for pedestrians.

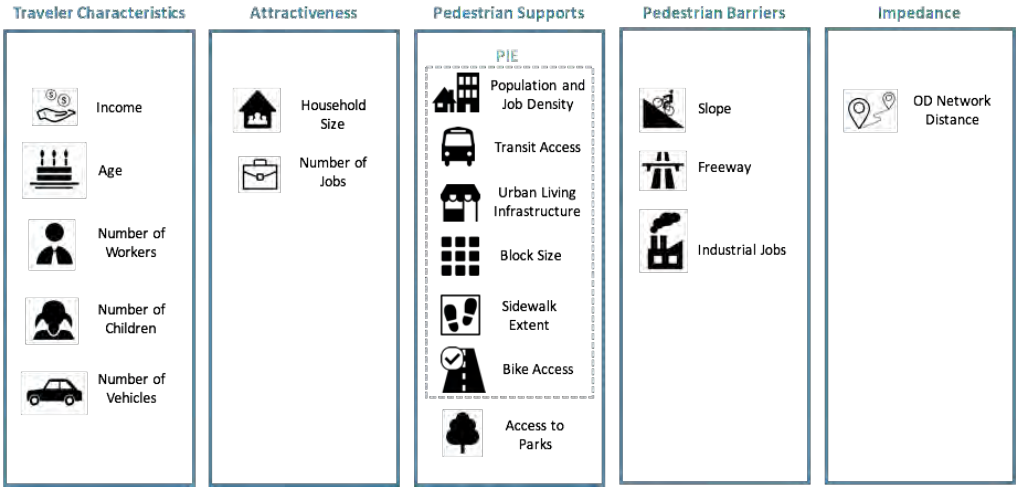

Figure 7 Model inputs including the Pedestrian Index of Environment (PIE)

We collect the index of environment variables above for each PAZ. In Figure 6, “Traveler characteristics” contain socio-demographic information from MTC III. The column “Attractiveness” corresponds to the potential number of trips as measured by the number of households and the number of jobs per PAZ.

“Pedestrian supports” encompasses seven dimensions shown in third column of Figure 6. Except for access to parks, the other six factors are grouped and called PIE. Access to parks was dropped from PIE because of its limited influence. Sidewalk extent was not available for the entire state. So we developed a proxy measure based on the road classification. Details of this proxy are in the Appendix. Factors in PIE are quantified on a scale from 0 to 5 for each individual PAZ cell, in which a score of 0 means no access to certain infrastructure. A subtotaled score of weighted PIE factors helps to illustrate geography specific to the most granular spatial unit; otherwise, users can aggregate the PIE into larger spatial units, for example, census block or tract level. We used ArcGIS to quantify all the factors in pedestrian supports category. In addition, we estimated coefficients for PIE index by running a logistic regression. Our approach is similar Clifton et al. *(23).* PIE is an index with values between 0 and 100. More information on the construction of PIE is in the Appendix.

“Pedestrian Barriers” include the mean slope in the destination zone, the presence of freeways and the proportion of industrial-type employment. We processed data in ArcGIS via the spatial join tool, so that every PAZ contains a value about slope, and information about whether freeway or industry appears.

As a measure of "Impedance," we directly used travel times reported in MTC III, which reflects time cost between the centroid of origin PAZ and destination PAZ along the network that includes the complete street network and major off-street paths.

3.3.2 Trip Production Model

The Conventional Four Step Method (FSM) for travel demand estimation (Figure 2) uses multiple techniques including cross-classification or linear regression to estimate total trip generation. However, the analysis unit of the FSM, which is generally a transportation analysis zone (TAZ), is way too large for people using non-motorized modes. Also, when pedestrians make decisions about where they walk, it is understandable that distance should matter. Again, TAZs are unable to capture the variation of travelers' walking distance since most of the walking trips would be intra TAZ trips rather than inter TAZ.

To overcome these limitations of TAZ in analyzing pedestrian behavior, we refer to the notion of PAZ proposed by Clifton et al. *(1)* as mentioned above. We found that about 63 percent of walking trips, in the travel survey MTC III, are from one PAZ to another, an improvement from the 38.5 percent found for inter Census Block Group. With built environment measured for PAZ, we further model household and employee's walking behavior for home-based and non-home based trips respectively. The exact walking behavior we model is the total number of walking trips produced from each PAZ, which is believed to be associated with household characteristics, built environment, and the number of employees.

Moreover, following the manner of FSM, trips are divided into five categories based on trip purposes *(31)*: home-based other (HBOther); home-based shopping (HBShopping); home-based school (HBSchool); non-home-based other (NHBO); and non-home-based work (NHBW). We perform regression analysis on each of them. Furthermore, based on the nature of the data, we choose a negative binomial regression and linear regression to fit the data for home-based trips and non-home based trips respectively. The regression can be represented by the following functions:

> Number of HB walking trips = f(number of households+household characteristics + built environment)

and

> *Number of NHB walking trips = f(number of employees + built environment).*

In Section 4.1.1, we present regression results including model coefficients and goodness of fit.

Synthetic Population
In order to generate trip production estimates, the composition of household characteristics must be known at the PAZ level. This is a very fine level of geographic scale not available in the American community survey (ACS).

Pedestrian exposure measures the intensity of pedestrian activities in each analysis zone. However, to measure exposure, it is essential for us to know the population for each PAZ first (i.e., the synthetic population). The synthetic population is comprised of households and individuals associated with list of characteristics -- e.g., household size, number of vehicles, gender or age *(33)*. In our case, due to the limited level of detail regarding individual characteristics recorded by the travel survey, we only chose household characteristics to generate the synthetic population.

Nevertheless, as noted earlier, such disaggregated data are not available at a small geographical scale like Census Block due to privacy concerns. To address this issue, our team generated the synthetic population for the state of Michigan at Census Block Group level based on following the iterative proportional fitting process (IPF)[9]. Specifically, based on the household characteristic needed for subsequent analysis and data availability, we took four household characteristics into account: household size; presence of children in the household; presence of workers in the household; and the number of vehicles owned by the household. Also note that for simplicity, we do not differentiate family households from non-family households. A section of the synthetic population results for Census Block Group 261614032001 is as follows:

Table 1 Sample synthetic population with GEOID 261614032001

Number of Persons	Presence of Children (binary)	Number of Vehicles	Presence of Workers (binary)	Count of this type of Households
1	0	0	1	11
1	0	1	0	13
2	1	1	1	5
3 and more	1	2 and more	1	25

The output of the IPF process is a cross-tabulation of household variables for every census block group in Michigan. Finally, the PAZs within each block group are assigned synthetic households uniformly at random. The details of synthetic population generation are in the Appendix.

[9] https://github.com/UDST/synthpop

3.3.3 Destination Choice Model

Now that we know how many trips are produced per day in each PAZ, we next ask: where do the trips terminate? The destination choice model answers this question.

The destination choice model utilizes a logistic regression, and given a set of variables of a possible destination PAZ, the model predicts the probability that a pedestrian would go to that PAZ. The explanatory variables are categorized into pedestrian impedance, attraction, support and barrier (see Figure 6).

The destination choice model is based on logistic regression, and the regression is built upon household travel survey (i.e., MI Travel Counts). To prepare training data for the regression model, for every actual walking trips by trip type, we create a 1.5 mile buffer zone around the origin PAZ, and randomly sample nine PAZs as possible destinations Figure 7 illustrates this process. The dependent variable for the travel records corresponding to the actual destination equal 1; while the records of the sampled destinations equal 0.

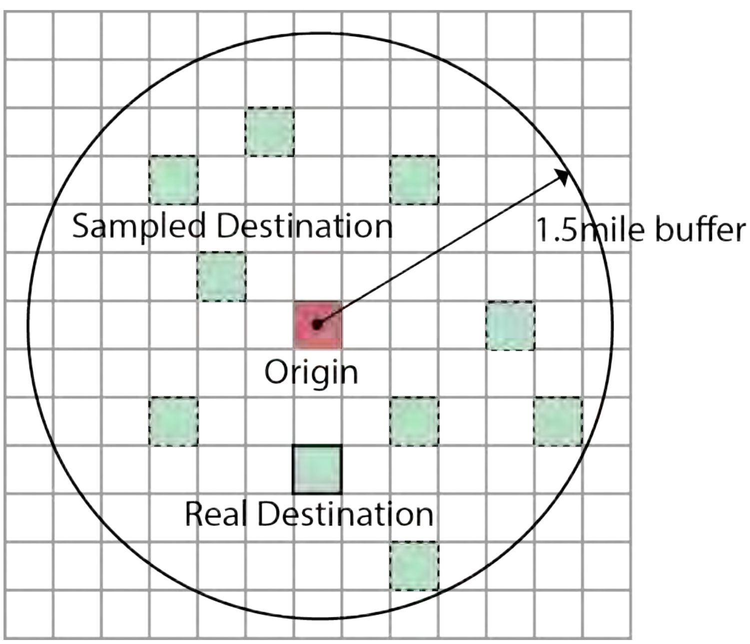

Figure 8 Destination model is based on sampling of real and possible destinations.

For each pair of origin and destination (O-D) PAZ, the destination choice model above helps us predict the probability that a pedestrian would walk between this O-D. After calculating the probability between all O-D within the buffer, we normalize these probabilities to make them add up to 1. The detailed destination choice regression results are in the Appendix.

Finally, for each origin PAZ, we multiply the total number of trips produced by the normalized destination choice probability to arrive at the total number of trips attracted to each possible destination PAZ.

3.3.4 Route choice (Network Assignment)

Now that we know where the trips are likely to begin and end; how do people travel from the origin to the destination? The route choice model answers this question. Once we know the route, we can aggregate over all of the trips to estimate the exposure for each network segment.

The following diagram illustrates how we calculate the exposure at the roadway segment level. To clarify, the destination PAZ below is one of all possible destinations within the 1.5 miles buffer zone of the origin PAZ. For convenience, we make them close to each other.

As mentioned above, we multiply the walking trip production by the normalized probability to estimate walking trips attraction. Because some PAZs would produce more walking trips given that they would attract more walking trips, certain streets are expected to have larger pedestrian exposure than others. To capture this, we assign walking trips that happen between each O-D along roadway network. Since the number of walking trips between each O-D is known from the last step, the network assignment is to find a route connecting each O-D. However, the O and D are not represented by the exact coordinates but an ID, thus we first try using the centroid of the PAZ as the O and D. Nevertheless, the results are not realistic because we observe that many streets have not been traveled.

To overcome this, as the following diagram suggests, we randomly choose one of the three nearest network nodes as O and D. With O and D represented by nodes on the roadway network, we presume each pedestrian would take the shortest route to reach their destination, and hence we utilize Dijkstra’s algorithm to find the shortest path (represented by the orange line). After performing the same procedure for each O-D, we end up getting the routes and their corresponding weights, i.e., the number of walking trips between the O-D.

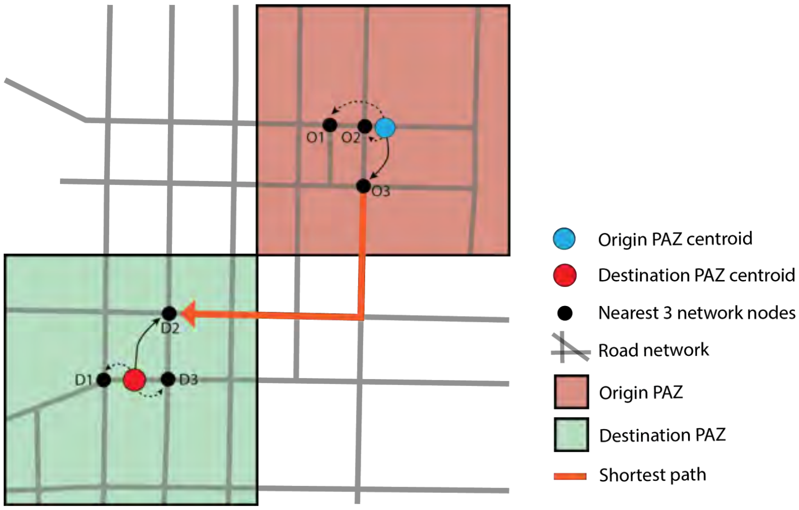

Figure 9 Route choice – Network Assignment Approach

This is a computational expensive task given that it is calculated for a very large number of origin and destination pairs of PAZs in the entire state of Michigan. In order to improve computational times, we developed several approximations to the shortest path problem. These included searching over a subnetwork (< 2 miles) for the shortest path instead of the entire road network. We were also able to use previously computed shortest paths in subsequent calculations. By using these approximations, the computational times decreased an order of magnitude.

3.4 Bicycle Risk Model

This section delineates the differences from the general framework as it relates to developing a bicycle risk model.

3.4.1 Data Sources

There are no changes in the data sources. Additionally, the unit of analysis is the same 400m by 400m area.

3.4.2 Risk Model

We use the same Empirical Bayes risk model as explained above in Section 3.2

3.4.3 Exposure Model

To develop the bicycle exposure model, we build the trip production, destination choice model using the reported bicycle trips in the MTC III travel survey. We replaced PIE with an analogous built environment measure called the Bicycle index of environment (BIE). In BIE, the sidewalk extent is replaced by the bike facilities extent.

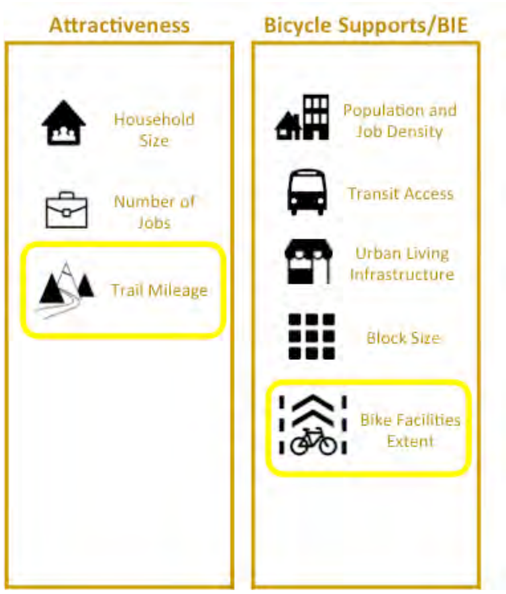

Figure 10 Bicycle specific built environment factors

The details of these factors and the construction of the index BIE are in the Appendix.

Bicycle Trip Production Model

The structure of the trip production is the same as the pedestrian model with BIE replacing PIE.

Bicycle Destination Choice Model

We increase the catchment and search radius of possible destination BAZs from 1.5 miles to 4 miles for the bicycle destination choice model. We justify this by the fact that bicycle trips are longer than pedestrian trips.

Bicycle Route choice (Network Assignment)

The bicycle route choice procedure is identical to the general approach described above in Section 3.3.4.

3.5 Experimental Design (Model Specification and Analysis)

This project did not contain an experiment. In order to test our key hypothesis the developed the risk assessment framework, and applied it to data derived from the state of Michigan. Finally, we tested the model predictions on data from Michigan to ascertain the accuracy of the risk scores.

3.6 Equipment

The statistical models and spatial analysis depended on a combination of ArcGIS Pro and Python. We relied heavily on the Python packages: pandas, geopandas and network x. All models were run on a PC with an Intel Xeon CPU at 3.50 GHz and 32.0 GB RAM.

3.7 Procedures

We sequentially developed the risk and corresponding exposure models for each county in the state.

4 Results

In this section, we describe the risk score results, as well as the exposure estimates for Wayne county. The results for all of counties in the state of Michigan are available in both the GIS files and the webtool. The GIS tool allows users to interactively view the results in ArcGIS. The webtool allows the user to display and interact with the results, without any special software, from any web browser including on mobile devices.

4.1 Method of analysis

We ran the risk assessment model on the entire state, which consists of 83 counties. A model for each of these counties was run separately to minimize memory requirements on our computer. Each county took approximately 45 minutes to run on a PC with an Intel Xeon CPU at 3.50 GHz and 32.0 GB RAM.

4.1.1 Trip Production

We used the Michigan household travel survey (MTC III) to fit our trip production model. As mentioned in the trip production model, we divided trips into five categories, namely home-based other (HBOther), home-based shopping (HBShopping), home-based school (HBSchool) and non-home-based other (NHBO) and non-home-based work (NHBW) and run the regression separately. In Tables 2 and 3, we highlight the results for home based other (HBO) and non-home based other (NHBO) trips.

Table 2 Pedestrian Trip production estimates for home based other trips.

Variable	Coef	Std err	P > |z|	[95.0% Conf. Int.]	
Constant	-0.7348	0.034	0.000	-0.801	-0.669
HHSIZE_1	-0.6565	0.071	0.000	-0.795	-0.518
HHSIZE_2	-0.1999	0.051	0.000	-0.299	-0.101
HHSIZE_3_or_more	-0.1217	0.073	0.0094	-0.021	0.264
HHVEH_0	0.8794	0.095	0.000	0.693	1.065
HHVEH_1	-0.2959	0.055	0.000	-0.404	-0.188
HHVEH_2	-0.5610	0.053	0.000	-0.664	-0.458
HHVEH_3_or_more	-0.7573	0.067	0.000	-0.888	-0.626
HHCHILD_0	-0.4110	0.051	0.000	-0.512	-0.310
HHCHILD_1_or_more	-0.3238	0.059	0.000	-0.439	-0.208
HHWORKER_0	-0.2714	0.040	0.000	-0.349	-0.194
HHWORKER_1_or_more	-0.4635	0.039	0.000	-0.540	-0.387
PIE	0.0272	0.002	0.000	0.022	0.032
Sample size	12062				
Log-Likelihood	-7457.5				

Table 3 Pedestrian Trip production estimates for non-home based other trips.

Variable	Coef	Std err	P > |z|	[0.025	0.0975]
Constant	-0.1958	0.015	0.000	-0.226	-0.166
Number_of_employees	-0.0675	0.002	0.000	-0.064	0.071
PIE	0.0049	0.000	0.000	0.004	0.006
Sample size	11650				
Log-Likelihood	-10927.0				

The results match our expectation for both home based trips and non-home based trips. The coefficients of PIE are always positive, which is reasonable because a larger PIE value stands for more pedestrian-friendly environment. For home based trips, the absence of vehicles is the most significant factors that make households choose to walk. However, owning more cars would make a household increasingly less likely to walk. Regarding non-home based trips the number of employees in a PAZ largely impacts the number of walking trips. Estimation results for the other trip purposes are in the Appendix.

4.1.2 Destination Choice

The table below summarizes the results of the destination choice model for pedestrians for all trip purposes.

Table 4 Pedestrian Destination choice model results

		Home-based work (HBW)			Home-based shopping (HBShopping)			Home-based school (HBSch)			Home-based other (HBO)			Non-home-based work (NHBW)			Non-home-based other (NHBO)		
	Variable	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P
Impedance	Distance (miles)	-1.859	0.154	0.000	-2.241	0.115	0.000	-2.208	0.004	0.000	-2.695	0.043	0.000	-3.096	0.123	0.000	-2.694	-0.065	0.000
	× Auto (yes)	-0.409	0.153	0.008	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Attraction	Retail jobs (#)	0.003	0.001	0.001	0.010	0.001	0.000	-0.005	0.002	0.011	--	--	--	0.003	0.001	0.000	0.003	0.000	0.000
	Service (#)	--	--	--	-0.001	0.000	0.028	--	--	--	--	--	--	--	--	--	--	--	--
	Government jobs (#)	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Finance jobs (#)	--	--	--	--	--	--	--	--	--	0.001	0.000	0.003	--	--	--	-0.002	0.000	0.000
	All other (#)	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Households (#)	--	--	--	0.003	0.001	0.014	--	--	--	0.002	0.000	0.001	-0.003	-0.001	0.000	-0.002	0.000	0.000
Support	Park (yes)	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	PIE	0.029	0.005	0.000	0.017	0.008	0.033	0.021	0.004	0.000	0.010	0.002	0.000	0.035	0.005	0.000	0.0235	0.000	0.000
Barrier	Slope (degree)	-0.325	0.074	0.000	-0.290	0.091	0.001	-0.350	0.060	0.000	-0.288	0.032	0.000	-0.467	0.071	0.00	-0.360	0.000	0.000
	Freeway (yes)	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	-0.404	0.161	0.012
	Industrial jobs (prop.)	--	--	--	--	--	--	-0.704	0.248	0.004	--	--	--	-0.868	0.299	0.004	-0.559	0.180	0.002
	Sample size	3996			3420			5769			31725			6768			15192		
	Log-likelihood	-766.40			-640.69			-1166.5			-5198			-988.59			-2438		
	Pseudo-R^2	0.406			0.419			0.387			0.495			0.5521			0.504		

In general, the model looks reasonable, distance, slope consistently have negative effect on making destination choice. Specifically, for a quarter mile (400 meters) increase in distance, the odds ratio in choosing a PAZ as destination varies from 0.461 to 0.628. In comparison, an increase of ten degree in the slope would lead to an odds ratio being as small as about 0.05, which is quite surprising. Moreover, an increase in PIE tends to increase the probability of choosing that PAZ as destination although its effect varies for different types trips. For example, for home-based work trips, a twenty-point increase in PIE would produce an odds ratio being 1.34. However, some values of the results could be counterintuitive. For example, proximity to park is not statistically significant for all trip purpose, and retail jobs would deter walking trips whose purpose is for school.

4.2 Presentation of results

4.2.1 Risk Score

The main step in generating the risk score is to estimate the safety performance function (SPF). This is accomplished by a negative binomial regression. The input variables to the SPF are: 1) log (AADT), 2) log of exposure, and an interaction term log(AADT)*log(exposure). Another input to the risk score is the number of observed crashes. For this case study, we aggregate the crash data from 2004-2015. We estimate the using a negative binomial distribution. Both the pedestrian SPF estimation results and bicycle SPF estimation results are below.

```
Optimization terminated successfully.
         Current function value: 0.576265
         Iterations: 28
         Function evaluations: 29
         Gradient evaluations: 29
                     NegativeBinomial Regression Results
==============================================================================
Dep. Variable:           Ped_Crash_Num   No. Observations:                29194
Model:                NegativeBinomial   Df Residuals:                    29191
Method:                            MLE   Df Model:                            2
Date:                 Mon, 02 Apr 2018   Pseudo R-squ.:                  0.1734
Time:                         19:50:31   Log-Likelihood:                -16823.
converged:                        True   LL-Null:                       -20353.
                                         LLR p-value:                     0.000
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -1.7638      0.019    -92.954      0.000      -1.801      -1.727
x1            -0.8991      0.109     -8.226      0.000      -1.113      -0.685
x2             2.3712      0.115     20.592      0.000       2.146       2.597
alpha          3.0400      0.078     38.994      0.000       2.887       3.193
==============================================================================
```

Figure 11 Estimation results for the pedestrian SPF negative binomial regression.

```
Optimization terminated successfully.
         Current function value: 0.517333
         Iterations: 29
         Function evaluations: 30
         Gradient evaluations: 30
                     NegativeBinomial Regression Results
==============================================================================
Dep. Variable:          Bike_Crash_Num   No. Observations:                29194
Model:                NegativeBinomial   Df Residuals:                    29191
Method:                            MLE   Df Model:                            2
Date:                 Mon, 02 Apr 2018   Pseudo R-squ.:                  0.1171
Time:                         19:50:32   Log-Likelihood:                -15103.
converged:                        True   LL-Null:                       -17107.
                                         LLR p-value:                     0.000
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -1.8350      0.020    -90.062      0.000      -1.875      -1.795
x1            -2.3961      0.153    -15.669      0.000      -2.696      -2.096
x2             3.8067      0.162     23.551      0.000       3.490       4.123
alpha          4.4984      0.122     36.967      0.000       4.260       4.737
==============================================================================
```

Figure 12 Estimation results for the bicycle SPF negative binomial regression.

Once the SPF is estimated and the observed crash data are compiled, the risk scores are generated from the empirical bayes Equation (1). Figure 4 displays the risk score for Wayne County over the same extent as the exposure map. The risk score has the interpretation of the expected number of crashes in the next 11 years. The results for both pedestrian and bicycle are presented below.

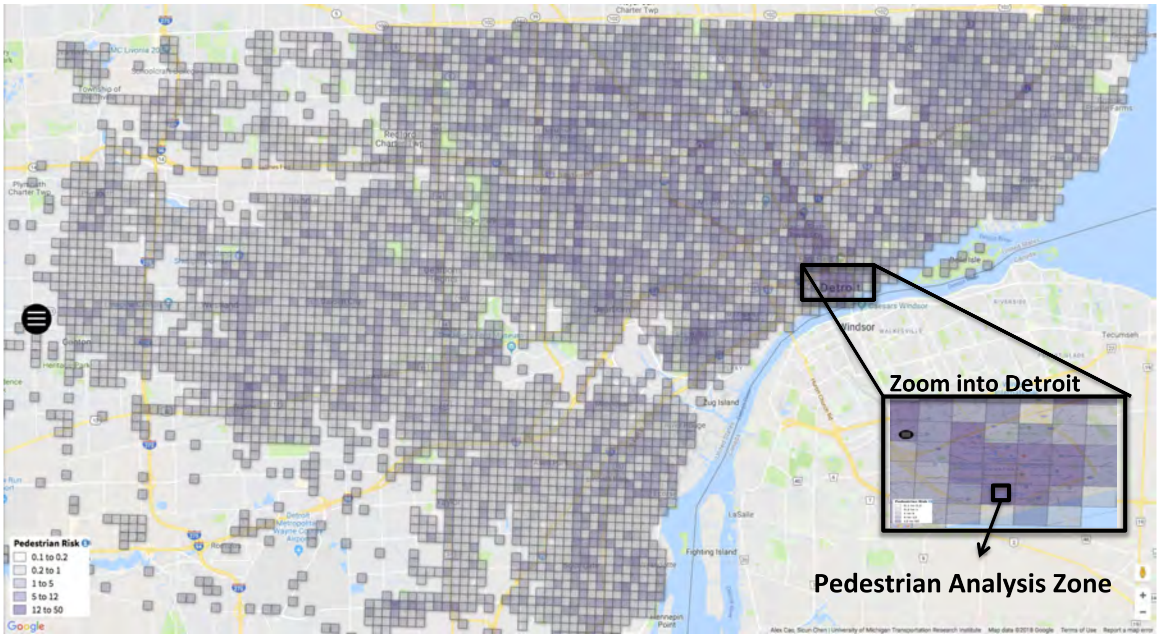

Figure 13 Map of Pedestrian risk score in Wayne County

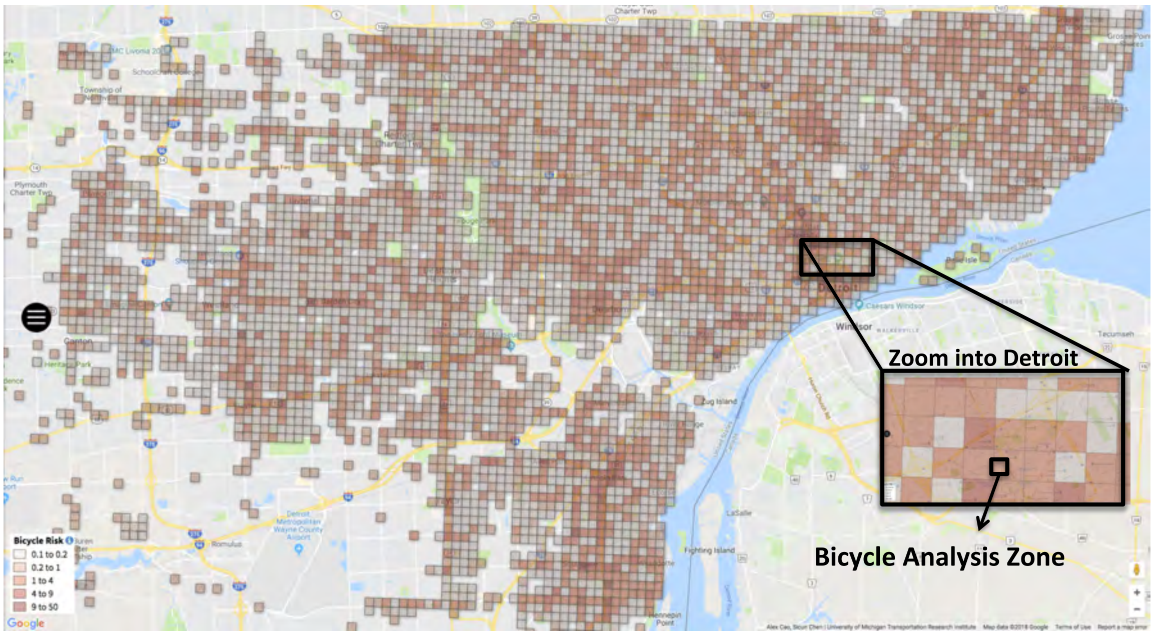

Figure 14 Map of Bicycle risk score in Wayne County

We computed the risk score for every county in the state using the same approach.

4.2.2 Exposure

Pedestrian Exposure

The resulting pedestrian exposure results are presented in Figures 15 and 16 for Wayne County. Exposure is measured as the number of daily walking trips that originate and terminate in a PAZ. Through the PIE index the exposure measure accounts for the population and job density, transit access, block size, and urban living infrastructure. The multi-way tables for household variables are created via IPF. The resulting household variables and PIE are the inputs to the trip generation model described in Section 3.3.2.

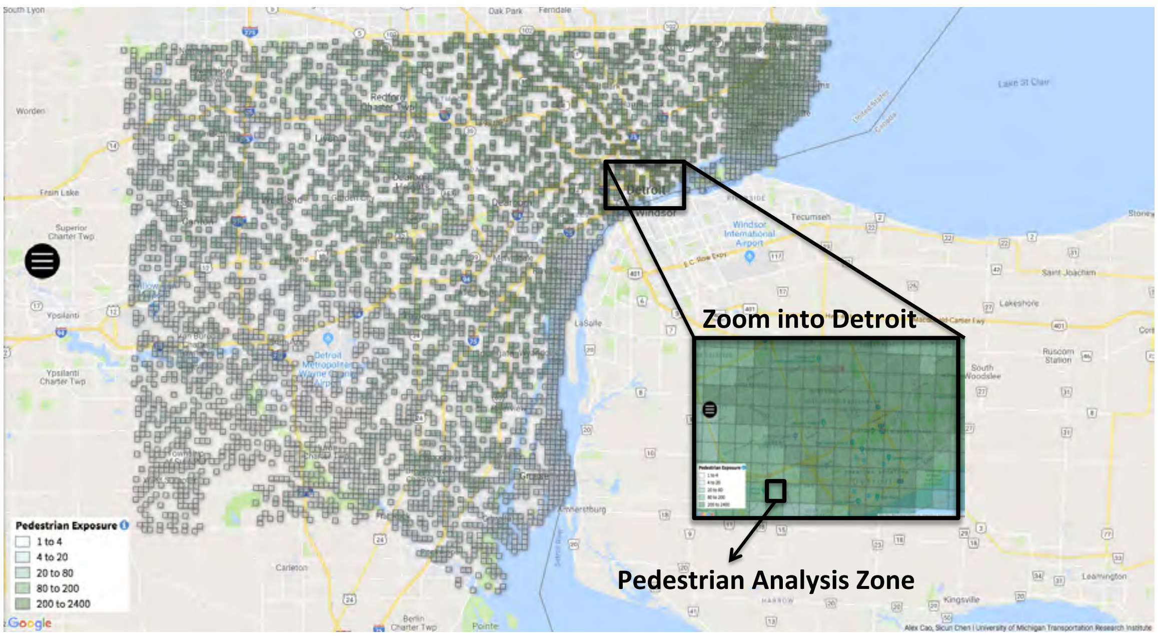

Figure 15 Map of daily pedestrian exposure in Wayne County at the PAZ level

Exposure for each road segment is also an output of the route choice and network assignment.

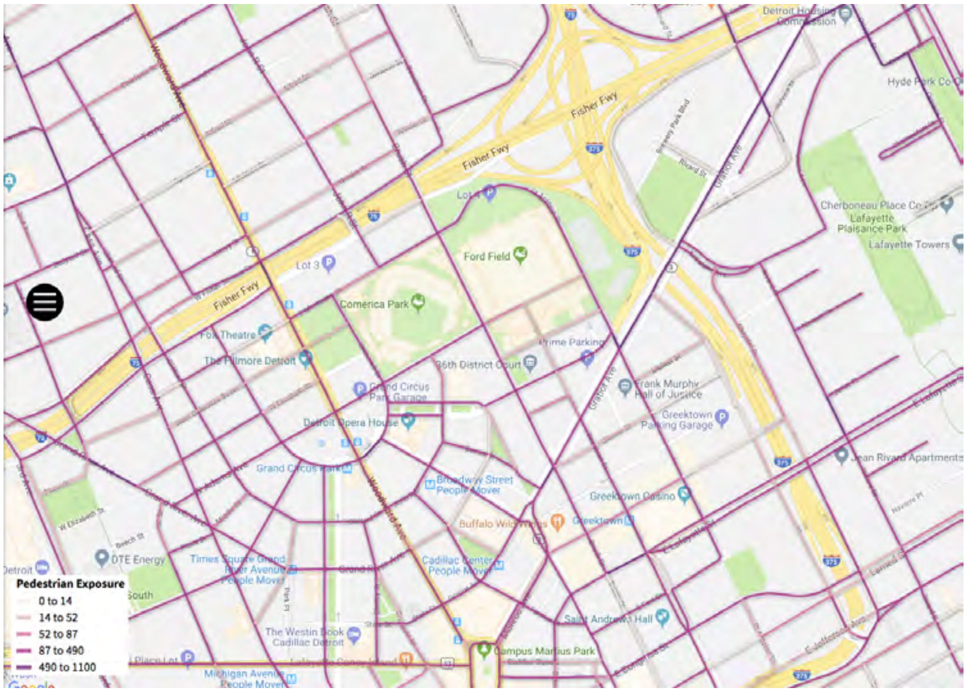

Figure 16 Map of daily pedestrian exposure in Wayne County at the road segment level.

Bicycle Exposure

The resulting bicycle exposure results are presented below for Wayne County. Exposure is measured as the number of daily bicycling trips that originate and terminate in a BAZ. Through the BIE index the exposure measure accounts for the population and job density, transit access, block size, and urban living infrastructure. The multi-way tables for household variables are created via IPF. The resulting household variables and BIE are the inputs to the trip production model.

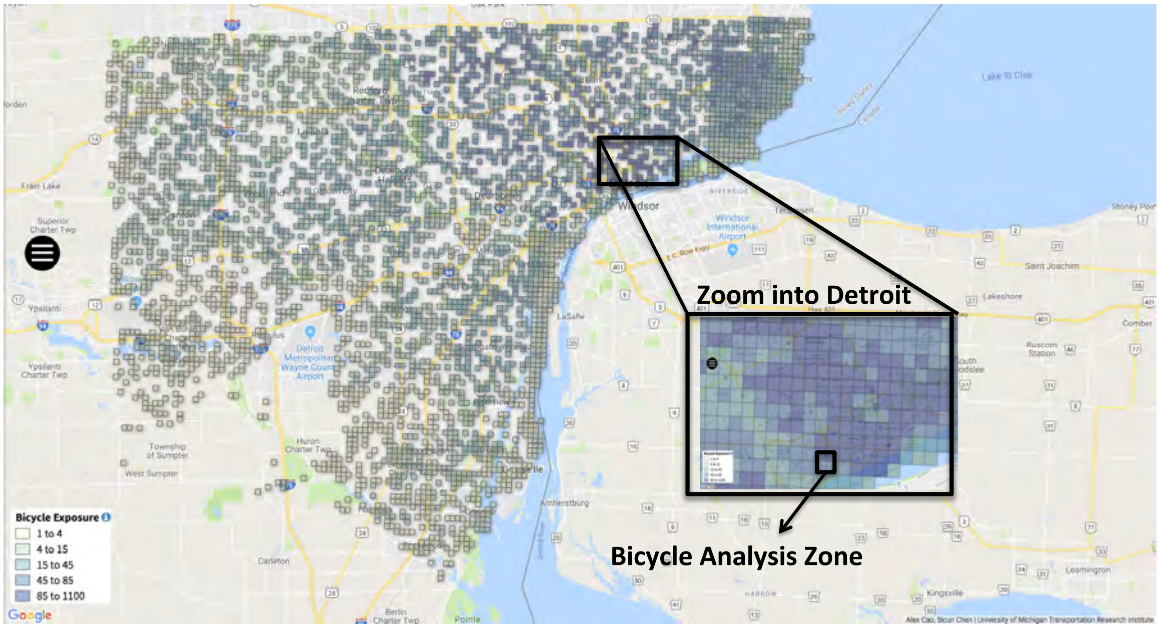

Figure 17 Map of daily bicycle exposure in Wayne County at the BAZ level.

The corresponding level of bicycle exposure at the roadway segment is below.

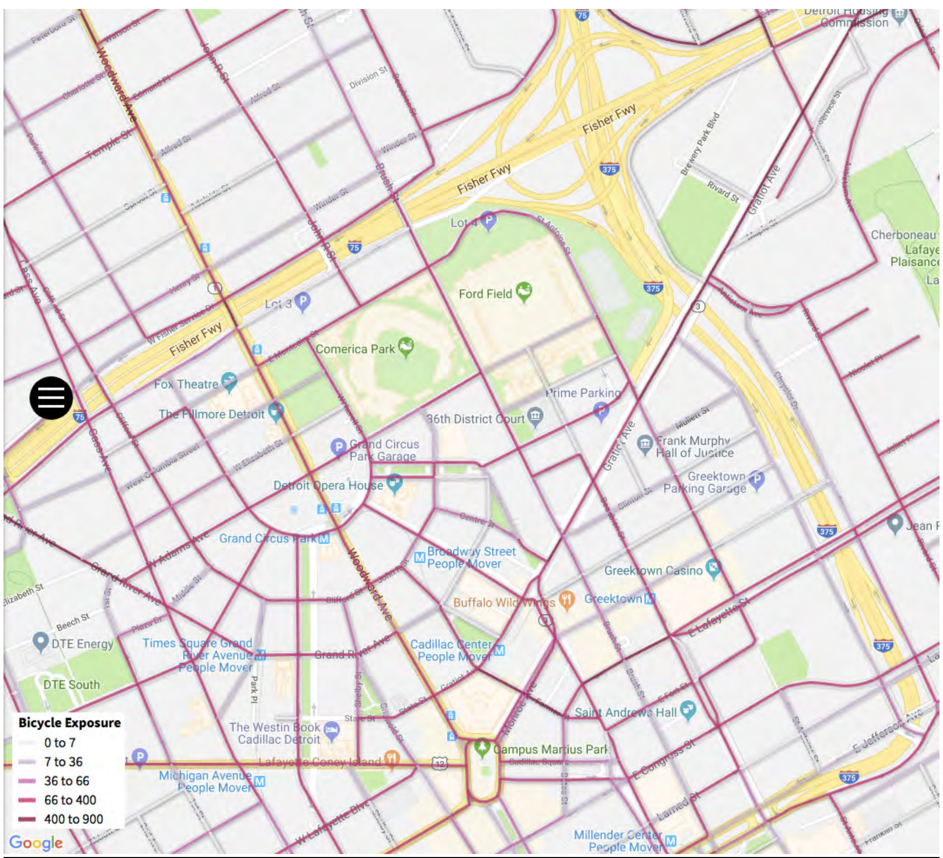

Figure 18 Map of daily bicycle exposure at the road segment level.

4.2.3 Model Validation

In order to test the accuracy of the risk score, we conducted a limited validation study. The risk score has the interpretation of the expected number of crashes in the next 11 years. We test the accuracy of our predictions by building the risk score model using only 90 percent of the PAZs. We then use the remaining 10 percent of the PAZs to test the accuracy.

However, in the last 11 years, zero crashes have occurred in most of the PAZs. You would be pretty accurate if you estimate a risk score of zero for every PAZ. However, this zero risk model does not capture the true intent of transportation planners.

To get around the issue of most PAZ having zero crashes, we can we create a list of the most dangerous PAZs. We define a PAZ to be dangerous if we expect more than 2 crashes in the next time period. Thus, we test how accurately our risk score methodology predicts the most dangerous PAZ.

To do this, we select a random subset of the PAZs to build the model using all years of data, and test on the remaining hold out PAZs. In order to test the prediction we use trained model to predict the "most dangerous" in the test data. Next, we use the actual crash data to construct a "most dangerous" list for the holdout PAZ sample.

A prediction is correct if the PAZ is a member of both "most dangerous lists." We repeat this process 20 times, selecting a random 10 percent of PAZ as the test set.

The results indicate that on average the risk score method identifies 9.8 percent of PAZs as "dangerous" when, in fact, they are actually not dangerous. This is called the false positive rate. Additionally, the relative prediction error for the number of PAZ identified as dangerous is 6 percent. Thus the risk score model reliably identifies the number of dangerous PAZs and has a modest false positive rate.

4.3 Discussion and Validity of hypotheses

We recall that the key hypothesis of the project,

> *A key hypothesis of this report is that it is possible to identify locations in need of countermeasures for pedestrian and bicycle crashes using a "risk score" that leverages information beyond observed crashes.*

We have demonstrated an approach for estimating non-motorized crash risk that goes beyond the observed number of crashes. We created fine scale risk scores and corresponding exposure measures for each county. Our limited validation study results provide evidence to support our key hypothesis. The next section describes some of the factors that affected our results.

4.3.1 Factors affecting the results

The project methodology and application have some limitations. These include our assumptions about “exposure,” as well as data quality and availability.

Definition of “exposure to risk”

Our proposed risk score and corresponding safety performance function consider a PAZ to be increasingly risky with the product of vehicle AADT and pedestrian exposure. However, our current approach does not guarantee that pedestrians and vehicles are actually interacting. For example, pedestrian walking on a separated path near a road is not distinguished from pedestrians walking on a sidewalk. This may lead to spurious predictions of high risk areas.

If we revisit the concept of pedestrian’s exposure to crashes, pedestrian exposure measures the number of potential opportunities for a crash to occur (Turner et al. 2017), although no consensus has been made about how to quantify it. In our approach, we use zonal walking trips made by pedestrian as a proxy for exposure, which is pedestrian travel demand indeed, and use demand times AADT to explore the interaction between pedestrians and vehicle traffic, which seems more like pedestrian exposure to me. Nevertheless, this might be problematic. For example, we have identified areas where pedestrian travel demand is high due to large number of employees, and crossed by highway, which makes the AADT measure affiliated to corresponding PAZ being high. As a result, the interaction term between travel demand and vehicle traffic turns out to be high as well. However, since that section of highway does not have access for pedestrians, there should be no pedestrian walking on the highway, the pedestrian exposure to crashes should be close to zero. Similar situations might exist that two PAZs have similar AADT and pedestrian travel demand, but one with relatively good pedestrian facilities (sidewalk, signalized intersection, stop signs, etc.) while the other does not. The exposure of pedestrians walking on the latter should be much higher than the former, but it has not been captured in our approach so far.

Data Availability

Our results were negatively affected by the paucity of data on sidewalk extent and bicycle facility extent. While select cities have a complete inventory of sidewalk extent and bicycle facilities, they were rare. Thus, our model ingested only a proxy measure for these variables. The risk model accuracy will surely improve if, and when, these data become widely available.

4.4 Implications

The risk score results should be used with caution until they are further validated and possibly corroborated with manual observations. Until then, we recommend that the results be used in combination with existing accepted safety analysis procedures.

5 Conclusions

5.1 Conclusions from the study

This report showed that it is possible to conduct a pedestrian and bicycle safety analysis that goes far beyond observed crashes. It also provides credence to the hypothesis that non-motorized crashes are predictable.

5.2 Recommendations for further research

In order to extract value from the project results, we recommend further research along several directions.

5.2.1 Implementation

The risk score results should be used with caution until they are further validated and possibly corroborated with manual observations. Until then, we recommend that the results be used in combination with existing accepted safety analysis procedures.

5.2.2 Validation

We recommend a follow up study to validate risk score and exposure results of this report and the current implementation of the tool. This includes research to improve the interface and usability of the webtool.

5.2.3 Improve Safety performance function

Apart from the aforementioned, areas for further studies should include improving the risk score by systematically defining “reference groups” in the Empirical Bayes computation. The reference groups will be based on roadway geometry as well as built environment and socio-demographics variables.

5.3 Recommendations for implementation

Further research is needed on the implementation of the tool into MDOT processes. Finally, the exposure results are of independent interest and are possibly relevant to crime and health outcomes.

6 Acknowledgements

The authors would like to thank Carissa McQuiston and Dean Kanitz from the Michigan Department of Transportation for their input and support for this project. Alex Cao was the GIS consultant for this project. Tian Tian and Yu-Hung Kuo

provided support for the literature review. Brian Hilbrands developed an earlier version of the pedestrian risk score. Thank you all.

7 References

1. Clifton, K., Singleton, P. A., Muhs, C. D., and Schneider, R. J. *Development of a Pedestrian Demand Estimation Tool*. Publication NITC-RR-677. National Institute for Transportation and Communities (NITC), 2015.
2. National Highway Traffic Safety Administration. Traffic Safety Facts, 2014. Pedestrians. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812375. Accessed June 30 2017.
3. National Highway Traffic Safety Administration. FARS Data. https://www-fars.nhtsa.dot.gov/Main/index.aspx. Accessed June 30 2017.
4. American Community Survey. B08101. Means of Transportation to Work by Age: Workers 16 years and over. 5-year estimates, 2009-2013.
5. American Community Survey. B08101. Means of Transportation to Work by Age: Workers 16 years and over. 5-year estimates, 2010-2014.
6. American Community Survey. B08101. Means of Transportation to Work by Age: Workers 16 years and over. 5-year estimates, 2011-2015.
7. Michigan Office of Highway Safety Planning, 2016 Annual Evaluation Report, http://www.michigan.gov/documents/msp/AER_2015_14_web_522481_7.pdf. Accessed June 30 2017.
8. Roche, J. and McFadden, J. Data-Driven Safety Analysis (DDSA). United States Department of Transportation Center for Accelerating Innovation. https://www.fhwa.dot.gov/innovation/everydaycounts/edc_4/ddsa.cfm. Accessed July 22 2017.
9. Raford, N., and D. Ragland. Space Syntax: Innovative Pedestrian Volume Modeling Tool for Pedestrian Safety. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1878, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 66-74.
10. Leden, L. Pedestrian Risk Decrease with Pedestrian Flow: A Case Study Based on Data from Signalized Intersections in Hamilton, Ontario. *Accident Analysis and Prevention*, Vol. 34, No. 4, 2002, pp. 457-464.
11. Jacobsen, P. L. Safety in Numbers: More Walkers and Bicyclists, Safer Walking and Bicycling. *Injury Prevention*, Vol. 9, 2003, pp. 205-209.
12. Keall, M. D. Pedestrian Exposure to Risk of Road Accident in New Zealand. *Accident Analysis and Prevention*, Vol. 27, 1995, pp. 729-740.
13. Hauer, E., D. W. Harwood, F. M. Council, and M. S. Griffith. Estimating Safety by the Empirical Bayes Method: A Tutorial. In *Transportation Research Record: Journal of the Research Board*, No. 1784, TRB, National Research Council, Washington, D.C., 2002. pp. 126-131.
14. Brimley, B. K., M. Saito, and G. G. Schultz. Calibration of *Highway Safety Manual* Safety Performance Function: Development of New Models for Two-Lane Two-Way Highways. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2279*, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 82-89.
15. El-Basyouny, K., and T. Sayed. Safety Performance Functions Using Traffic Conflicts. *Safety Science*, Vol. 51, No. 1, 2013, pp. 160-164.

16. Qin, X., Ivan, J. N., and Ravishanker, N. Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis and Prevention*, Vol. 36, No. 2, 2004, pp. 183-191.
17. Kerridge, J., Hine, J., Wigan, M. Agent-Based Modelling Of Pedestrian Movements: The Questions That Need To Be Asked and Answered. *Environment and Planning B: Planning and Design*, Vol. 28, 2001, pp. 327 - 341.
18. Schneider, R. J., Henry, T., Mitman, M. F., Stonehill, L., and Koehler, J. Development and Application of the San Francisco Pedestrian Intersection Volume Model. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2299, Transportation Research Board of the National Academies, Washington, D. C., 2012, pp. 65-78.
19. Pulugurtha, S.S. and Repaka, S.R. Assessment of Models to Measure Pedestrian Activity at Signalized Intersections. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2073, Transportation Research Board of the National Academies, Washington, D. C., 2008, pp. 39-48.
20. Clifton, K. J., Singleton, P. A., Muhs, C. D., Schneider, R. J., and Lagerwey, P. Improving the Representation of the Pedestrian Environment in Travel Demand Models, Phase I, 2013.
21. Ewing, R., and R. Cervero. Travel and the Built Environment: A Synthesis. In *Transportation Research Record: Journal of the Transportation Board, No. 1780*, TRB, National Research Council, Washington, D.C., 2001, pp. 87-114.
22. Saelens, B. E., and Handy, S. L. Built environment correlates of walking: a review. *Medicine and science in sports and exercise*, Vol. 40, Suppl. 7, 2008, pp. S550-S566.
23. Clifton, K. J., Singleton, P. A., Muhs, C. D., and Schneider, R. J. Representing pedestrian activity in travel demand models: Framework and application. *Journal of transport geography*, Vol. *52*, 2016, pp.111-122.
24. Müller, K., and Axhausen, K. W.. Population synthesis for microsimulation: State of the art. ETH Zürich, Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau (IVT), 2010.
25. Beckman, R. J., K. A. Baggerly, and M. D. McKay Creating Synthetic Baseline Populations. *Transportation Research Part A*, Vol. 30, No. 6, 1996, pp. 415-429.
26. Guo, J. Y., and C. R. Bhat. Population Synthesis for Microsimulating Travel Behavior. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2014*, Transportation Research Board of the National Academies, 2007, pp. 92-101.
27. American Community Survey Office. U.S. Census Bureau. American Community Survey 2011-2015 ACS 5-year PUMS Files ReadMe. https://www2.census.gov/programs-surveys/acs/tech_docs/pums/ACS2011_2015_PUMS_README.pdf. Accessed June 25 2017.
28. Infogroup. Reference USA. http://www.referenceusa.com/Home/Home. Accessed Apr. 4 2017.
29. Admmistraiion, F. H. Highway Performance Monitoring System. 1998.
30. Michigan Traffic Crash Facts. Vehicle-pedestrian crashes. https://www.michigantrafficcrashfacts.org/. Accessed Jan. 29 2017.
31. McNally, M. G. The Four-Step Model. In *Transport Modelling Handbook*: 2nd Edition (D.A. Hensher, K.J. Button, ed.), Emerald Group Publishing Limited, 2007. pp. 35-53.

32. Clifton, K. J., Singleton, P. A., Muhs, C. D., and Schneider, R. J. Development of destination choice models for pedestrian travel. *Transportation Research Part A: Policy and Practice,* Vol. 94, 2016, pp. 255-265.
33. Voas, D., and P. Williamson. An Evaluation of the Combinatorial Optimization Approach to the Creation of Synthetic Microdata. *International Journal of Population Geography*, Vol. 6, No. 5, 2000, pp. 349-366.
34. Savolainen, P. T., Gates, T. J., Lord, D., Geedipally, S. R., Rista, E., Barrette, T, Thompson P., and Thompson I. *Michigan Urban Trunkline Segments Safety Performance Functions (SPFs) Development and Support*. Publication RC-1639. Michigan Department of Transportation, 2016.
35. Turner S., Sener I., Martin M., Das S., Shipp E., Hampshire R., Fitzpatrick K., Molnar L., Wijesundera R., Colety M. and Robinson S. *Synthesis of Methods for Estimating Pedestrian and Bicyclist Exposure to Risk at Areawide Levels and on Specific Transportation Facilities*. Publication FHWA-SA-17-041. FHWA, U.S. Department of Transportation, 2017.

8 Appendix

8.1 Glossary

Throughout this report, we use the term “non-motorized” to include both pedestrian and bicyclist.

8.2 List of Acronyms, Abbreviations and Symbols

BIE – Bicycle Index of Environment

EB – Empirical Bayes

HSM – Highway Safety Manual

IPF – Iterative Proportional Fitting

PAZ – Pedestrian Analysis Zone

PIE – Pedestrian Index of Environment

MTC - MI Travel Counts III

SPF – Safety Performance Function

8.3 Other Appendices

8.3.1 Experimental data

Trip Production Model Results

HBShop Trips

```
                           Logit Regression Results
==============================================================================
Dep. Variable:             walk_or_not   No. Observations:                  949
Model:                           Logit   Df Residuals:                      935
Method:                            MLE   Df Model:                           13
Date:                 Fri, 09 Dec 2016   Pseudo R-squ.:                  0.2799
Time:                         14:42:53   Log-Likelihood:                -157.20
converged:                        True   LL-Null:                       -218.30
                                         LLR p-value:                 7.418e-20
==================================================================================
                      coef    std err          z      P>|z|      [95.0% Conf. Int.]
----------------------------------------------------------------------------------
Walk_score_Shop     0.0015      0.000      3.269      0.001         0.001     0.002
HH_Size_2          -0.7666      0.627     -1.222      0.222        -1.996     0.463
HH_Size_3           0.0326      0.867      0.038      0.970        -1.667     1.732
HH_Size_>3          1.0048      0.862      1.166      0.244        -0.684     2.694
HHIncome_Median    -0.8747      0.397     -2.201      0.028        -1.654    -0.096
HHIncome_High      -0.9461      0.520     -1.821      0.069        -1.964     0.072
HHVeh_1            -2.1217      0.648     -3.272      0.001        -3.393    -0.851
HHVeh_2            -3.6776      0.744     -4.941      0.000        -5.136    -2.219
HHVeh_>2           -6.0658      1.251     -4.850      0.000        -8.517    -3.614
HHChild_1           0.0899      0.741      0.121      0.903        -1.363     1.543
HHChild_>=2        -1.5144      0.700     -2.164      0.030        -2.886    -0.143
HHWorker_1          1.0404      0.548      1.897      0.058        -0.035     2.115
HHWorker_>1         3.0327      0.692      4.382      0.000         1.676     4.389
intercept          -1.3892      0.639     -2.173      0.030        -2.642    -0.136
==================================================================================
```

HBO Trips

Logit Regression Results

Dep. Variable:	walk_or_not	No. Observations:	4088
Model:	Logit	Df Residuals:	4074
Method:	MLE	Df Model:	13
Date:	Fri, 09 Dec 2016	Pseudo R-squ.:	0.1437
Time:	14:20:06	Log-Likelihood:	-1315.6
converged:	True	LL-Null:	-1536.3
		LLR p-value:	3.615e-86

	coef	std err	z	P>|z|	[95.0% Conf. Int.]	
Walk_score_Other	4.567e-05	3.66e-06	12.469	0.000	3.85e-05	5.28e-05
HH_Size_2	-0.2728	0.173	-1.579	0.114	-0.612	0.066
HH_Size_3	0.5823	0.226	2.576	0.010	0.139	1.025
HH_Size_>3	0.5259	0.318	1.654	0.098	-0.097	1.149
HHIncome_Median	-0.3342	0.141	-2.367	0.018	-0.611	-0.057
HHIncome_High	-0.7495	0.173	-4.341	0.000	-1.088	-0.411
HHVeh_1	-1.3177	0.249	-5.301	0.000	-1.805	-0.831
HHVeh_2	-1.8570	0.276	-6.736	0.000	-2.397	-1.317
HHVeh_>2	-2.4626	0.309	-7.966	0.000	-3.068	-1.857
HHChild_1	-0.5854	0.187	-3.125	0.002	-0.953	-0.218
HHChild_>=2	-1.0118	0.264	-3.830	0.000	-1.530	-0.494
HHWorker_1	0.8846	0.171	5.160	0.000	0.549	1.221
HHWorker_>1	1.1593	0.188	6.180	0.000	0.792	1.527
intercept	-0.9318	0.239	-3.903	0.000	-1.400	-0.464

HBS Trips

Logit Regression Results

Dep. Variable:	walk_or_not	No. Observations:	490
Model:	Logit	Df Residuals:	476
Method:	MLE	Df Model:	13
Date:	Fri, 09 Dec 2016	Pseudo R-squ.:	0.2036
Time:	15:34:12	Log-Likelihood:	-183.71
converged:	True	LL-Null:	-230.68
		LLR p-value:	2.458e-14

	coef	std err	z	P>|z|	[95.0% Conf. Int.]	
Walk_score_School	0.0103	0.003	3.069	0.002	0.004	0.017
HH_Size_2	-1.1551	0.538	-2.146	0.032	-2.210	-0.100
HH_Size_3	-1.0020	0.698	-1.435	0.151	-2.370	0.366
HH_Size_>3	-0.3246	0.712	-0.456	0.648	-1.720	1.071
HHIncome_Median	-0.6234	0.331	-1.884	0.060	-1.272	0.025
HHIncome_High	-0.5079	0.429	-1.185	0.236	-1.348	0.333
HHVeh_1	-0.2162	0.603	-0.358	0.720	-1.398	0.966
HHVeh_2	-0.3637	0.627	-0.580	0.562	-1.592	0.864
HHVeh_>2	-0.7363	0.738	-0.998	0.318	-2.182	0.709
HHChild_1	-1.3234	0.437	-3.028	0.002	-2.180	-0.467
HHChild_>=2	-2.5607	0.488	-5.251	0.000	-3.516	-1.605
HHWorker_1	0.1561	0.537	0.291	0.771	-0.896	1.208
HHWorker_>1	0.2630	0.544	0.483	0.629	-0.803	1.329
intercept	0.4277	0.754	0.567	0.571	-1.051	1.906

The results show that the different types of trips have different statistical significance patterns for these variables. However, the measure of accessibility is always statistically significant, this verifies our assumption that the ease to reach destinations would influence pedestrians' mode choice. Also, the number of vehicles within the household is significant for both trips with other purpose and shopping trips. But we cannot observe that for school trips, which might due to the fact that children can take the school buses to school. Many interesting observations can be make from these results, for example, as the number of household vehicles increases, the coefficient, which is negative value, decreases, indicating that traveler has less probability to travel by walking. These reasonable outputs can be further processed to remove non-statistical significant variables and used to estimate the walk mode choice.

8.3.2 Analytical technique details

ArcGIS Spatial Analysis

We used ArcGIS and python as our principal tools in the spatial analysis part. The following sections are about geo-reference of Michigan, data preparation for modeling, and weights calculation for selected factors.

Geo-Reference[10]

To specify limits in distortions, Michigan adopted State Plane Coordinate System of 1983, which broke the territory into three separate horizontally oriented projections. The entire Upper Peninsula makes up the northern zone, the northern half of the Lower Peninsula is the central zone, and the southern half of the Lower Peninsula is the southern zone. However, for a project that scopes for the whole state, using three different projections is inefficient. Thus, we decided to use the alternative system named Michigan GeoRef. Even though GeoRef allows more variance on distance and area, it was designed to project the State using a single zone rather than three zones. Specifically, we chose NAD 1983 Michigan GeoRef in US feet. The following graphics show the differences between the two systems.

10 Michigan Department of Natural Resources, available at: http://www.michigan.gov/documents/DNR_Map_Proj_and_MI_Georef_Info_20889_7.pdf

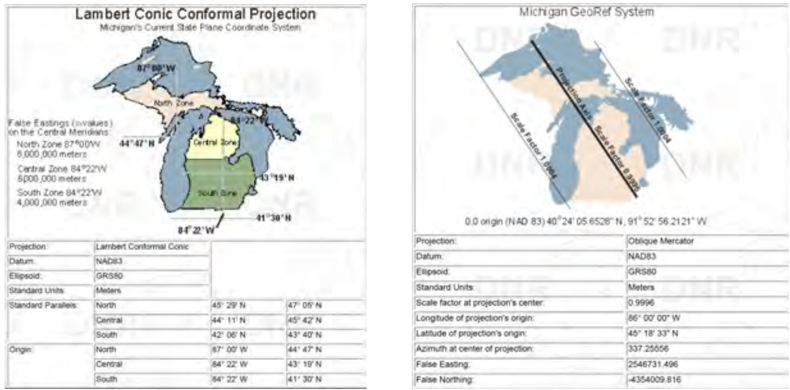

Figure 15 Geo-reference System

Pedestrian Index of Environment (PIE)

In addition, a key component, which is a comprehensive measure of the built environment, called Pedestrian Index of Environment (PIE) is introduced into our model. Ideally, the PIE will replace the accessibility in our model and functions as an important indicator of pedestrians' mode choices and destination choices. With a combination, the walk trip production model and destination choice model, we are able to estimate the walk trips productions and attractions in a much finer geographical range in near future.

The first report, OTREC-ED-510,[11] addresses the walk trip production while the second report, NITC-RR-677,[12] focuses on walk trip distribution. These two methodologies in combination were used to estimate how many walk trips are produced within each analysis unit and where these walk trips end. In comparison, these methodologies have demonstrated many advantages over our previous GIS-based Accessibility Model, especially its inclusiveness of sociodemographic characteristics of the travelers as well as its much finer geographical range to capture the subtlety of pedestrian's mode choice. Furthermore, these reports also validate how their models can achieve relatively high accuracy in estimating pedestrian demand in Portland Metropolitan Area, Oregon. It is noteworthy that bicycle demand estimation was not introduced in these reports, nevertheless, similar methodologies can be readily applied to estimate bicycle trips. Finally, we note that the first report was produced in 2013, and second report was not released until September 2015. These relatively new reports somewhat represent the recent trend in the research field with regard to modelling walk trips. Based on these advantageous features and the availability of the data, our team supposes these methodologies are worth a try in the

---

[11] Clifton, Kelly J., Patrick Allen Singleton, Christopher Devlin Muhs, Robert J. Schneider, and Peter Lagerwey. "Improving the Representation of the Pedestrian Environment in Travel Demand Models, Phase I." (2013).
[12] Clifton, Kelly, Patrick Allen Singleton, Christopher D. Muhs, and Robert J. Schneider. "Development of a Pedestrian Demand Estimation Tool." (2015).

context of Michigan, and ideally, output can be integrated into the formation of the risk score.

Data Requirement

In general, MoPeD 2.0 mainly involves travel survey data and built environment data. In terms of the nature of the method, the required data could be categorized into five groups. The groups are traveler characteristics, attractiveness, pedestrian supports, pedestrian barriers, and impedance. Each group contains several factors, some of the factors will be weighted and subtotaled to group value to serve as a variable in later logit models, and the others factors in each group will be studied individually. The team followed the data gathering and deduction process that Dr. Clifton had done for Portland, but made some minor changes based on Michigan settings. All the data are manipulated at PAZ level.

Group 1: Traveler Characteristics
The first group was called traveler characteristics, which contained socio-demographic information from the statewide travel survey, MTC III. In the original report, data from 2011 Oregon Household Activity Survey was partitioned for the modeling and validation process to see if the social-demographic status of travelers have significant impact on their choices of walking and destinations. In addition, they randomly selected 90% of the travel survey trips to estimate the model, and retained the other 10% for model validation. The team followed the report's method but applied MTC III in this case. MTC III was a travel survey conducted in 2015 by MDOT, sampled 16,276 households across the state reporting their weekday trips.

There were five factors in this group: household income, age, number of workers, number of children, and number of vehicles. All the information could be easily retrieved from the MTC III raw data, the questionnaire was designed internally for advanced analysis.

We used the merged table from "household.dbf" and "person.dbf". "person.dbf" contained all the information for a complete tour that a surveyed person made for the designated date; while "household.dbf" was about the demographic information of surveyed households, which could be joined to "person.dbf" based on sample IDs. The following table indicates the factors and its corresponding field name in MTC III dataset. We cooperated the factors into mode split and destination choice modeling later individually.

Table 5 Select MTC Survey Field Names

Factor	Field Name
Household Income	"HHINC"
Age	"AGE_AAGE"
Number of Workers	"HHWORKER"
Number of Children	"HHCHILD"
Number of Vehicles	"HHVEH"

Group 3: Pedestrian Supports

Pedestrian supports is the most complicated group of factors, which will possibly but not definitely increase the willingness of people to walk. Dr. Clifton's study was built on an existing index in Portland metro area, named Metro Context Tool. It is an index of built environment that encompasses seven dimensions: population and job density, block size, transit access, sidewalk extent, urban living infrastructure, bicycle access, and access to park.[13] Each dimension is quantified on a scale of one to five for individual PAZ cell. A subtotaled score of equally weighted seven dimensions represents the character of the built environment of each PAZ through the measurement of objective conditions. The index helps to illustrate geography specific to a finest spatial unit, or it can be aggregated to whatever level, for instance, census block group level, and city level.

However, the Metro Context Tool does not address the possibility that some components are more influential than the others. In other words, weighting each component equally will overestimate the influence of factors that have weak relationships with walking and underestimate the influence of factors that have stronger relationships with walking. Therefore, after tested and calibrated the Metro Context Tool, Dr. Clifton and the team introduced a new index, called the Pedestrian Index of Environment (PIE). They built a series of binomial logit regression model to derive weights for each dimension in Metro Context Tool. It turned that access to park had the weakest relationship with pedestrian trip mode choice, and park itself will potentially mislead the results. Therefore, access to park was dropped from consideration of the PIE, it serves as a supplementary factor of pedestrian supports.

Consequently, the pedestrian supports group is consisted of two sub-groups: PIE and access to park. PIE is an intermediate value derived from six dimension. In the report, the score of each dimension ranges from 1 to 5, but we considered it is possible in rural Michigan that people have neither access to transit, nor bicycle path. Thus, in our version, the score ranges from 0 to 5, in which a score of 0 means no access to certain infrastructure. In the following paragraphs, we will discuss how we assigned scores to each PAZ. The general information of each dimension is shown in the table.

[13] Clifton, Kelly, Patrick Allen Singleton, Christopher D. Muhs, and Robert J. Schneider. "Development of a Pedestrian Demand Estimation Tool." (2015).

Table 6 Pedestrian Supports Data Description

Dimension	Buffer	Classification (0 to 5; low to high)	Data Source
Population and Number of Jobs	¼ mile	Natural Breaks	ACS 2015, ReferenceUSA
Block Size (Intersection Density)	¼ mile	Natural Breaks	U.S. Census Tiger
Transit Access	¼ mile	Natural Breaks	AAATA, U of M Transit Service, DDOT, SMART, Google Maps
Sidewalk Extent (NFC)	¼ mile	Natural Breaks	U.S. Census Tiger
Urban Living Infrastructure	¼ mile	Natural Breaks	ReferenceUSA
Bicycle Access (AADT)	1 mile	Natural Breaks	Highway Performance Monitoring System (HPMS)
Park Access	¼ mile, ½ mile, ¾ mile, 1 mile, 1¼ mile	Linear Breaks	Esri

Transit Access

Accessing to transit has always been a key driving force for people to walk. Due to data unavailability, transit accessibility was simply measured by the number of transit stops within quarter-mile buffer of a PAZ. If more data were available, analysis involving transit frequencies should have been done to increase the models' accuracy.

Since there was no statewide transit authority, data were collected from several sources. Bus and light rail stops in Detroit metropolitan region were obtained from Data Driven Detroit,[14] credits to DDOT and SMART. Bus stops in Ann Arbor-Ypsilanti area were available at City of Ann Arbor Data Catalog,[15] credits to AAATA and U of M Transit Service. Bus stops in Kent County were downloaded and reorganized from Grand Rapids Open Data,[16] credits to The Rapid. The other regions, including City of Kalamazoo, Traverse City, Tri-County region, data were extracted from Google Maps.
We used spatial join tool in ArcMap to get counts of transit stops in each buffer zone. PAZ with no stop has no access to public transit, while PAZ with more stops has higher transit accessibility, so we assigned score according to following table.

[14] http://portal.datadrivendetroit.org/datasets?q=bus+stop&sort_by=relevance
[15] http://www.a2gov.org/services/data/Pages/default.aspx
[16] http://data.grcity.us/dataset/gtfs

Table 7 Value Breaks in Transit Access Measure

Score	Value (Number of Transit Stops)
0	0
1	0 < x ≤ 3
2	3 < x ≤ 12
3	12 < x ≤ 25
4	25 < x ≤ 55
5	55 < x

Population and Number of Jobs

We measured this factor by population and number of jobs per spatial unit. Population at census block group level was retrieved from American Community Survey (ACS) 5-Year Estimate 2011-2015 at Social Explorer.[17] Job data was obtained from ReferenceUSA,[18] which provided location based information of verified businesses, including NAICS code and actual location employees. It is noteworthy that we corrected one obvious error in ReferenceUSA employee data, the Dow Chemical Company at 1801 Larkin Center Dr., the location number of employee should be 150 instead of 88,000.[19]

Before applied spatial join tool, we created random points within each census block group based on population data, each point represents one person. It would be easier to distribute total number of population from census block groups into PAZs in this case. Afterwards, we joined both number of jobs per location and population points into quarter-mile buffered PAZ. PAZs with more population and jobs were assigned higher score according to following table.

Table 8 Value Breaks of Score for Measuring Population and Number of Jobs

Score	Value (Number of People in Total)
0	0
1	0 < x ≤ 649
2	649 < x ≤ 2,317
3	2,317 < x ≤ 6,127
4	6,127 < x ≤ 17,817
5	17,817 < x

[17] http://www.socialexplorer.com/tables/ACS2015_5yr
[18] http://www.referenceusa.com/UsBusiness/Search/Custom/
[19] https://www.dandb.com/businessdirectory/thedowchemicalcompany-midland-mi-2446318.html

Block Size

To simplify the process, we used number of road junctions per PAZ to measure block size. Usually, more junctions in a certain area mean tighter networks, which is more appealing to pedestrians.

The data was derived from road feature from U.S. Census Tiger product. First, we used vertices in the point tool to get endpoints of each road segment. Second, we used collect events tool to identify number of endpoints at intersections. The final step was to filter out road junctions out of geometric nodes. The rule was keeping nodes, which contain equal or greater than three endpoints, which means at least three road segments intersect and form a road junction. PAZs with more junctions were assigned higher score according to following table.

Table 9 Value Breaks of Score for measuring Block Size

Score	Value (Number of Junctions)
0	0
1	$0 < x \le 4$
2	$4 < x \le 18$
3	$18 < x \le 38$
4	$38 < x \le 64$
5	$64 < x$

Urban Living Infrastructure (ULI)

Certain destination types were measured within a quarter-mile radius of each PAZ. Business location data from ReferenceUSA were queried for specific NAICS codes to determine the accessibility of PAZs to day-to-day living needs, such as K-12 schools, grocery stores, cafes, restaurants, clothing and other retail stores, dry cleaners, and entertainment venues. To put it simply, ULIs are businesses that provide service and financial activities, the NAICS of involved industries are listed as follow. And the score should be assigned according to the number of ULIs with quarter-mile buffer of each PAZ as shown in the following table.

Table 10 NAICS Codes for Urban Living Infrastructure

NAICS	Category
44-45	Retail Trade
522	Credit Intermediation & Related Activities
54	Professional, Scientific & Technical Services
6111	Elementary & Secondary Schools
71	Arts, Entertainment & Recreation
722	Food Services & Drinking Places
812	Personal & Laundry Services

Table 11 Value breaks for Measuring Urban Living Infrastructure (ULI)

Score	Value (Number of ULI)
0	0
1	$0 < x \le 13$
2	$13 < x \le 50$
3	$50 < x \le 134$
4	$134 < x \le 452$
5	$452 < x$

Sidewalk Extent

Ideally, this measure should have been the interaction of total length and condition of sidewalk per PAZ. However, due to statewide sidewalk data deficits, we had to use road function classification as substitute. Again, we used the road system data, which was obtained from U.S. Census Tiger.[20] The field “NFC” in attribute represented national road function classification. It was categorized into eight levels. For interstates and other freeways, we assumed there should not be pedestrians, so we assigned 0 on these categories. For the other road categories, we constructed a comparison table of presence of sidewalk and NFC code in Kent County, where we had the most comprehensive sidewalk data. The comparison table shows that higher level of road has higher probability of having sidewalk, so they will get higher scores. The scores should be assigned according to the following table.

[20] https://www.census.gov/cgi-bin/geo/shapefiles/index.php

Table 12 Cross Comparison between the presence of Sidewalks and NFC codes.

NFC	Description	No Sidewalk/Segments	No Sidewalk/Percentage	Has Sidewalk/Segments	Has Sidewalk/Percentage	All Roads Segments	Score
1	Interstate	275	89.9%	31	10.1%	306	0
2	Other Freeway	114	91.9%	10	8.1%	124	0
3	Other Principle Arterial	50	5.8%	816	94.2%	866	5
4	Minor Arterial	65	5.6%	1091	94.4%	1156	5
5	Major Collector	92	14.2%	558	85.8%	650	4
6	Minor Collector	-	-	-	-	-	3
7	NFC Local	1029	20.2%	4064	79.8%	5093	2
0	Non-certified	692	87.3%	101	12.7%	793	1
Total		2317	25.8%	6671	74.2%	8988	

Table 13 Value breaks of Score for Measuring Sidewalk Extent

Score	Value (NFC)
0	1, 2
1	0
2	7
3	6
4	5
5	3, 4

Park Access

Park access as factor was dropped from PIE, but worked individually in pedestrian supports in Clifton's report. In any case, park is an important destination for people choosing to walk to, especially there are so many recreational walking trips in MTC III. The park data was retrieved from ArcGIS Online,[21] Esri has a default dataset of parks in the U.S., ranging from national parks and forests to local parks. We applied a series of buffers with ¼ mile interval around every park's boundary. Areas being closer to parks got higher score, vice versa. Areas away from parks more than 1¼ mile were considered as no walking access to park. Scores were assigned according to the following table.

[21] https://www.arcgis.com/home/index.html

Table 14 Value breaks for score for measuring bicycle access

Score	Value (Distance from Park Boundary/ Mile)
0	1¼ < x
1	1 < x < 1¼
2	¾ < x ≤ 1
3	½ < x ≤ ¾
4	¼ < x ≤ ½
5	0 ≤ x ≤ ¼

Group 4: Pedestrian Barriers

Barriers to pedestrian travel include the mean slope in the destination zone, the presence of freeways and the proportion of industrial-type employment as a proxy for industrial land uses. For slope, we used DEM data from Michigan Open Data Portal to generate the average slope in degree per PAZ. The freeways can be found in the attribute of road system downloaded from US Census Tiger by NFC. The presence of freeways is a dummy variable, if there is any freeway intersect with a PAZ, the score should be 1; if there is not any freeway present in a PAZ, the score should be 0. For industrial-type jobs, again, we utilized employment data ReferenceUSA. We filtered out retail, service, finance and government jobs, and the remaining categories are industry-type jobs. The NAICSs of industry-type job are 11, 21, 23, 31, 32, 33.

Group 5: Impedance

As a measure of impedance, we calculated the shortest path distance (in miles) between the centroid of production zones and attraction zones along a network that included the complete street network and major off-street paths. We have already built up the road network, but the scale of the zone remains open to questions. The production and attraction zone could be based on either PAZ or SuperPAZ, but the calculation is extremely time consuming, but the finest level zones are able to capture the most accurate pedestrian flow. We could also calculate the impedance based upon census block group, but the details will be gone.

Calculating Weights

Build upon these data, we took a further step to calculate the weights for factors in PIE. Because PIE is a union of several predictors, equal weighted factors should be nice as a general index, however, setting up weights can better predict the preference of pedestrians on built environment. We ran a series binomial logistic regressions to get coefficients for each factor. The coefficients quantified the relationship between factor values and the observed utility of choosing to walk rather than the other modes.

At first try, we followed Dr. Clifton's report but failed to standardize the weights because the coefficients of sidewalk extent and bicycle access turned to be negative. These two factors needed to be further reexamined.

After dropping sidewalk extent and bicycle access, we started to consider developing a new index. Since we are developing statewide level models, for the convenience of model maintenance, data availability should be our prior concern. For all other data-available factors under pedestrian supports category, we could group them together as a new index, we called it PIE, which contained five factors: people per acre, block size, transit access, ULI and park access. We ran the regressions again, and the results turned to be promising. From the coefficients we could tell, population density was the strongest variable, and park access was the weakest. All the factors were statistically significant. Using the coefficients, we were able to calculate the weights. To make PIE more intuitive, we set a value range from 0 to 100. When we maintained the ratios among the coefficients, we scaled them to weights respectively. The weights and maximum weighted value for each factor in PIE is listed as follow. In the future, if the data deficits issue is solved, different factors and weights will be applied.

Table 15 Coefficients for the Pedestrian Index of Environment (PIE)

Factor	Coefficient	P-Value	Pseudo R-Square
People per Acre	0.50	0.000	0.029
Block Size	0.46	0.000	0.038
Transit Access	0.36	0.000	0.027
ULI	0.32	0.000	0.015
Park Access	0.18	0.000	0.015

Table 16 Weights for Each Factor in PIE

Factor	Score Range	Weight	Maximum Weighted Value
People per Acre	0-5	5.495	27.475
Block Size	0-5	5.055	25.275
Transit Access	0-5	3.956	19.78
ULI	0-5	3.516	17.58
Park Access	0-5	1.978	9.89
PIE			100

Bicyclist Index of Environment (BIE)

This section gives a step-by-step description on how we estimate bicycle supports within each SuperPAZ. With bicycle supports, we can further estimate bicycle trip generation and bicycle destination.

Bicycle supports refer to an index of built environment that will support cycling activities and possibly encourage people to cycle. Basically, an area with higher bicycle support index, the built environment within that area will be friendlier for people to cycle. To objectively determine bicycle index of environment, we selected five factors to estimate bicycle supports. Each factor is quantified on a scale of zero to five for each SuperPAZ cell and is presented as a score. A subtotaled score of equally weighted five factors represents the character of the built environment of each SuperPAZ through the measurement of objective conditions.

- **Population and job density:** This factor was measured by population and number of jobs per spatial unit. We joined both number of jobs per location and population points into quarter-mile buffered SuperPAZ. SuperPAZs with more population and jobs were assigned higher score.
- **Block size:** We used number of road junctions per SuperPAZ to measure block size. Usually, more junctions in a certain area means more connected networks which is more appealing to cyclists. We calculated the number of junctions within each quarter-mile buffered SuperPAZ.
- **Bicycle Access:** A one-mile radius around every PAZ was used to calculate the bicycle access in that area. In this case, a one-mile radius represented the increased accessibility range of bicycles over pedestrian travel. Unfortunately, bicycle paths were unavailable statewide, we had to use AADT as substitute. Typically, high-traffic roads without bicycle facilities are considered as lower biking comfort level, and vice versa. Thus, we assumed that “bikability” here in Michigan also follow this cognition, in which higher AADT associates lower “bikability”.

We obtained AADT data directly from MDOT, which covers almost 80% of statewide road segments (630,469 out of 794,277). Scores were assigned according to the following table.

Table 17 Value breaks for score for bicycle access

Score	Value (AADT)
0	Null
1	7,584 < x
2	3,454 < x ≤ 7,584
3	1,645 < x ≤ 3,454
4	607 < x ≤ 1,645
5	0 ≤ x ≤ 607

- **Transit Access:** Transit accessibility was simply measured by the number of transit stops within quarter-mile buffer of a PAZ. We counted numbers of transit stops in each buffer zone. A SuperPAZ with no bus stop means there is no access to public transit, while a SuperPAZ with more stops has higher transit accessibility and get the higher score of transit access.

- **Urban Living infrastructure:** Certain destination types were measured within a quarter-mile radius of each PAZ. Infrastructures like K-12 schools, grocery stores, cafes, restaurants, clothing and other retail stores, dry cleaners, and entertainment venues were all recorded in this factor, which provide service and financial activities that support daily living.
- **Bicycle Facilities extent:** Bicycle facilities extent was collected from OpenStreepMap which is an open source that users can extract bicycle paths in the selected area. After we get the bicycle path length within each SuperPAZ, we used Intersect tool in ArcMap to get the length of bicycle paths in each half-mile buffer zone. As a result, we assigned scores based on Natural Breaks method, a SuperPAZ with less than 10 feet will be assigned a zero score for bicycle facilities extent, while SuperPAZ with longer bicycle paths will have higher score for bicycle facilities extent, as shown in the following table:

Table 18 Bicycle Facilities extent

Score	Value (Feet)
0	0 < x ≤ 10
1	10 < x ≤ 45
2	45 < x ≤ 108
3	108 < x ≤ 210
4	210 < x ≤ 259
5	259< x ≤ 1337

Weight Calculation

From the previous step, each factor is quantified on a scale of zero to five for each SuperPAZ cell and is presented as a score. A subtotaled score of equally weighted five factors represents the character of the built environment of each SuperPAZ through the measurement of objective conditions. However, weighting each factor equally will overestimate the influence of factors that have weaker relationships with cycling and underestimate the influence of factors that have stronger relationships with cycling. In addition, setting up different weights can better predict the preference of cyclists on each bicycle index of built environment. According to Dr. Clifton's study, they built a series of binomial logit regression model to derive weights for each factor. In this research, we utilize MI counts as the training data to estimate the correlation between people's mode choice for cycling and factors in the index of built environment. The coefficients of each factor will quantify the relationship between factor values and the observed utility of choosing to cycle rather than the other modes, as shown in the following table.

Table 19 BIE regression results

Factor	Coefficient	P-Value	Pseudo R-Square
People per Acre	0.35	0.000	0.0097
Block Size	0.30	0.000	0.0012
Transit Access	0.26	0.000	0.0105
ULI	0.15	0.000	0.0022
Bike Facilities Extent	0.15	0.000	0.0017

Conforming to our predictions, all factors have the positive relationship with people's willingness of cycling, which means people will be more likely to cycle in an area where the built environment is more supportive of cycling. Using these coefficients, we could calculate their influences on people's mode choice for cycling. To make bicycle supports more intuitive, we set a value range from 0 to 100. When we maintained the ratios among the coefficients, we scaled them to weights respectively. The weights and maximum weighted value for each factor in bicycle supports is listed as follow. Ultimately, we can get the BIE based on the foumula below.

Table 20 BIE Weights

Factor	Score Range	Weight	Maximum Weighted Value (Score = 5)
People per Acre	0-5	5.785	28.925
Block Size	0-5	4.959	24.795
Transit Access	0-5	4.298	21.49
ULI	0-5	2.479	12.395
Bike Facilities	0-5	2.479	12.395
Bike supports			100

BIE = 2.479*ULIScore + 4.298*TAScore + 5.785*PPAScore + 2.479*BFScore + 4.959*BSScore
(Total=100)

Table 21 Bicycle model data sources

Category	Element	Data Source
Bike Attractiveness	Size Term	ACS 5-Year Estimates 2011-2015, ReferenceUSA (April,2017)
	Retail Jobs	ReferenceUSA (April,2017)
	Finance Jobs	ReferenceUSA (April,2017)
	Government Jobs	ReferenceUSA (April,2017)
	Service Jobs	ReferenceUSA (April,2017)
	Other Jobs	ReferenceUSA (April,2017)
	Households	ACS 5-Year Estimates 2011-2015
	Bike Trail Mileage	GIS Open Data(Michigan DNR Designated Bicycle Trails)
Bike Supports	BIE	ACS 5-Year Estimates 2011-2015, U.S. Census Tiger Porduct,AAATA, U of M Transit Service, DDOT, SMART, Google Maps, OSM
	BIE (People per Acre)	ACS 5-Year Estimates 2011-2015
	BIE (Block Size)	U.S. Census Tiger Product
	BIE (Transit Access)	AAATA, U of M Transit Service, DDOT, SMART, Google Maps
	BIE (ULI)	ReferenceUSA (April,2017)
	BIE (Bike Facilities Extent)	OpenStreetMap
Barriers	Slope	Michigan Open Data Portal
	Freeway	U.S. Census Tiger Product
	Industrial Jobs	ReferenceUSA (April,2017)
Base Map	State Boundary	U.S. Census Tiger Product
	Counties in Michigan	U.S. Census Tiger Product
	SuperPAZ	Self-Developed
	Road System	U.S. Census Tiger Product
Crash	Pedestrian Crash Only	UMTRI
	Bike Crash Only	UMTRI
Drinking Related Places	Bars/ Pubs	ReferenceUSA (April,2017)
	Liquor Stores	ReferenceUSA (April,2017)
	Restaurants	ReferenceUSA (April,2017)
Road Geometry	AADT	U.S. Census Tiger Product, MDOT
	Number of Lanes	HPMS
	Speed Limit	HPMS
	AADT (less)	HPMS

8.3.3 User Manuals

We produced both ArcGIS files and an online web based pedestrian and bicycle risk assessment tool[22]. The manual for both of these products are available online

https://github.com/caocscar/pedbikeriskexposure/blob/master/draft.md#developing-michigan-pedestrian-and-bicycle-safety-models

8.3.4 Evidence for Factors Associated with Frequency of Non-Motorized Crashes

As part of this project we conducted a thorough scientific literature review to identify the risk factors associated with bicycle and pedestrian crashes. We considered the inclusion of each factor below into the model. We included a factor into the model if the data is widely available, effect size is large and under the control of MDOT policy actions. The table between lists the factors, magnitude of the effect on crashes, and references.

[22] see http://www.cmisst.org/pedbike-risk-exposure/

Factors Associated with Frequency of Pedestrian and Bicycle Crashes in US

Factor	Effect/Effect Size	References
Socio-demographic Factors		
Sex	Increased likelihood of pedestrian crashes among males.	Zhou et al. (2013); Campbell et al. (2004); Lee and Abdel-Aty (2005); Al-Shammari et al. (2009); Pulugurtha et al.(2004)
Age	Increased likelihood of pedestrian crashes among young people. Middle-aged males are more frequently involved in pedestrian-vehicle collision. Older pedestrians were found to experience a slightly higher number of fatal crashes per million walk trips per year.	Age 16-29: Zhou et al. (2013); Age under 15: Jang et al. (2013); Alluri et al.(2013); Baltes (1998) Mirabella and Zhang(2014) Alluri et al.(2013)
African-American or Hispanic neighborhoods; greater proportion of median-age and uneducated populations (or low income)	Significant positive correlation between pedestrian crash frequency.	Ukkusuri et al. (2011); Kravetz & Nolan (2012)
Alcohol involvement	Positively associated with pedestrian crashes.	Nolan & Quddus (2004); Jang et al. (2013); Pulugurtha et al.(2004); Kittelson(2014)
Land Use		
Greater number of schools and commercial	Increased pedestrian crashes. Center business district(CBD) area experiences more crashes per site than the residential area by about threefold.	Ukkusuri et al. (2011); Kim & Ortega (1999); Lascala et al (2000); Azam et al. (2012) Sheaffer (2008)
Residential	Greater proportion of residential land decreased likelihood of pedestrian crashes. Residential area types	Kravetz & Nolan (2012); Kim & Yamashita (2002) Zegeer, Opiela, and Cynecki

	increased risk of pedestrian crashes.	(1985)
	The majority of pedestrian crashes occurred in urban areas and especially in metropolitan areas, fatal crashes were disproportionately high in rural areas.	Alluri et al.(2013)
Roadway Characteristics		
Road width	Increased road width (wider streets) associated with higher risk of pedestrian crashes. Narrower roads and roads with higher speeds had increased risk for bicyclists. Two-way streets (compared to one-way); wider streets; high volumes of turning vehicles all increase risk. Four-legged intersections experience a greater number of crashes per site than three-legged intersections. Sidewalks and wide shoulders significantly improve pedestrian safety. Lack of bicycle facility increases the risk of bicycle safety.	Ukkusuri et al. (2012); Garder (2004); Zegeer, Opiela, and Cynecki (1985); Zegeer, Stewart, Huang, and Lagerwey(2001); Davis(1990) Hunter, Stutts, Pein, and Cox (1996) Zegeer, Opiela, and Cynecki (1985) Sheaffer (2008) Alluri et al. (2013); McMahon et al.(2002) Kittelson(2014)
Traffic volume	Higher pedestrian and vehicle volumes associated with increased risk of pedestrian crashes (pedestrian volumes one of most influential factors). Major road Average Daily Traffic (ADT) have an inverse effect on vehicle-pedestrian collision. Left-turn volume (increased volume and increased	Fernandes et al. (2012); Zegeer, Stewart, Huang, Lagerway, Feaganes, and Campbell (2005); Lyon and Persaud (2002); Zegeer, Opiela, and Cynecki (1985); Brude and Larsson(1993); Sheaffer(2008); McMahon et al.(2002); Harwood et al(2008) Leden (2002); Lyon and Persaud (2002) Davis (1987); Epperson (1994);

	proportion of total traffic, respectively) is associated with pedestrian safety.	Sorton and Walsh (1994) (peak-hour traffic volume in the curb lane); Landis (1994); Landis, Vattikuti, and Brannick (1997); Harkey, Reinfurt, Knuiman, Stewart, and Sorton (1998) (curb-lane volume); Landis, Vattikuti, Ottenburg, Petritsch, and Crider(2003) and Noel, Leclerc, and Lee-Gosselin, 2003).
Number of lanes	Two lane roads accounted for 60% of crashes. Greater number of lanes associated with more pedestrian and bicycle crashes. Intersections that have collector or arterial roadways with 4-lanes on at least one approach are a risk factor for pedestrian safety.	Hunter, Stutts, Pein, and Cox, (1996) Zegeer, Stewart, Huang, and Lagerwey (2001); Kittelson(2014) Kittelson (2014)
Raised pedestrian crosswalk, median or island	Reduced crashes, especially on multilane roads; Curb extension and pedestrian node can reduce risk of safety for pedestrian.	Zegeer, Stewart, Huang, and Lagerwey (2001); Elvik and Vaa (2004); Campbell, Zegeer, huang, and Cynecki (2004); Alluri et al. (2013); Kittelson(2014)
Crosswalk marking	On 2-lane roads, marked crosswalk alone at uncontrolled location had no association with pedestrian crash risk. On multilane roads with volume >12,000/day, increased pedestrian crash risk.	Zegeer, Stewart, Huang, Lagerway, Feaganes, and Campbell (2005)
Speed limits	Three-quarters of crashes occurred on roads with speed limits of 35 mph or less. Higher speed limits increase pedestrian safety risk. Speed limit is included as a variable in bicycle models.	Hunter, Stutts, Pein, and Cox, 1996); Kittelson(2014) McMahon et al.(2002) Davis (1987); Epperson (1994); Sorton and Walsh (1994) (vehicle speeds in the curb lane); Landis (1994); Landis, Vattikuti, and Brannick (1997); Harkey, Reinfurt, Knuiman, Stewart, and Sorton (1998) (vehicle speeds in the curb lane); and Noel, Leclerc, and

		Lee-Gosselin (2003).
Bus operations	Presence of bus operations associated with increased pedestrian risk. Bus stop is a risk factor for pedestrian safety.	Zegeer, Opiela, and Cynecki (1985) Kittelson (2014)
Traffic control	Exclusive pedestrian signal timing associated with lower pedestrian risk compared to concurrent timing at signalized intersections without pedestrian signals. An analysis of leading pedestrian interval(LPI) at downtown traffic signals shows nearly 50% reduction in pedestrian-vehicle crashes(LPI is a 3 second advance walk indication that is given to pedestrians prior to the circular green indication given to vehicles). Signalized intersection with permitted or protected left-turn phases is a risk factor for pedestrian safety.	Zegeer, Opiela, and Cynecki (1985); Kittelson (2014) Sheaffer (2008) Kittelson (2014)
Lighting conditions	The proportion of fatal crashes that occurred during nighttime were significantly greater compared to the proportion of fatal crashes in the day time.	Alluri et al. (2013); Kittelson (2014)
Safety Performance Functions		
Michigan Urban Trunkline Intersections Safety Performance Functions (SPFs):	This model is developed for pedestrian and bicycle crashes based on **vehicular** annual average daily traffic(AADT) for pedestrian and bicycle crashes. Crashes increase with respect to major road and minor road traffic volumes. Multi-linear Model Variables Coefficient 3ST AADT(maj) 0.765 AADT(min) 0.385 3SG AADT(maj) 0.402 AADT(min) 0.187 4ST AADT(maj) 0.547 AADT(min) 0.269 4SG AADT(maj) 0.364 AADT(min) 0.173	Savolainen, Gates, Lord, Geedipally, Rista, Barerrette, Russo, Hamzeie (2015) 3ST: 3-leg intersection with stop control on the minor approach. 3SG: 3-leg signalized intersections.

Highway Safety Manual Safety Performance Functions (SPFs)	Vehicle-pedestrian crashes at signalized intersections(Multi-linear Regression Model) Variables Coefficient 3SG 4SG AADT(total) 0.05 0.40 $AADT_{min}/AADT_{maj}$ 0.24 0.26 PedVol 0.41 0.45 n_{lanesx} 0.09 0.04 CMFs(crash modification factors) Bus Stops (See chapter 12 pg-46) Schools (See chapter 12 pg-46) Alcohol Sales Establishments (See chapter 12 pg-47)	AADT: vehicular annual average daily traffic PedVol: pedestrian volume n_{lanesx}: maximum number of traffic lanes crossed by a pedestrian
SPF	Pedestrian SPF (Poisson Model) Variables coefficient ADT 0.0000251 Bicycle Volume 0.000091 Number of left turn lanes 0.2296894 On-street parking 0.5712769 Presence of speed signs - 0.4470537 Presence of bus stop 0.9400843 Bicycle SPF: (Negative Binomial Model) Variables coefficient ADT 0.0000419 Pedestrian volume 0.0008022 Number of left turn lanes	Hamidreza (2014)

	0.1566364 Presence of bicycle lanes 0.5408297 Presence of bus stop 0.9032806	
Pedestrian and Bicycle Intersection Safety Indices(ISI)	Bicycle ISI Model Variables Coefficient 1 Main street ADT 0.019 2 Main street speed limit 0.815 3 Presence of turning- 0.650 vehicle traffic across the Path of through cyclists 4 Vehicle right-turn lanes 0.470 and bicycle lane present 5 Cross street ADT and 0.023 no bicycle lane 6 Traffic signal and no 0.428 bicycle lane 7 Parking on approach 0.200 and no bicycle lane Ped ISI Model Variables coefficient SIGNAL -1.867 STOP - 1.807 THRULNS 0.335 SPEED 0.018 MAINADT*SIGNAL 0.006 COMM 0.238	Carter, Hunter, Zegeer et al (2006) Variable 2,3,4,5,6,7 a value of 1 indicates that specific condition is true SIGNAL: traffic signal-controlled crossing 0=no, 1=yes STOP: stop sign controlled crossing 0=no, 1=yes THRULNS: number of through lanes on street being crossed SPEED: 85 percentile speed of street being crossed MAINADT*SIGNAL: traffic volume on street being crossed COMM: predominant land use on surrounding area is commercial development 0=no 1=yes
Behavior		
Direction in relationship to traffic	Pedestrians facing traffic had 77% reduction in fatal and injury crashes compared to pedestrians walking in direction	Luoma and Peltola (2013)

	of traffic.	
Crossing behavior	Pedestrians who crossed the road compared to those who walked along the road at higher risk. Through vehicular movements at intersections had a greater effect on crash rates than left- and right-turn movements.	Sarkar et al. (2011) Fernandes et al. (2012)
Crossing compliance	Males are more likely to cross without right of way than females. Pedestrians are more unlikely to cross against signal when heavy vehicle traffic exists. Groups of more than two individuals waiting on curbs are more likely to obey traffic laws. Push buttons positively influence compliance. Intersection with rest in walk and pedestrian recall had higher compliance rates for pedestrian and bicyclist.	Mirabella and Zhang (2014)