



*FINAL REPORT*

## **Improving Bicycle Crash Prediction for Urban Road Segments**

---

NITC-RR-756 ■ May 2018

*NITC is a U.S. Department of Transportation  
national university transportation center.*





# **IMPROVING BICYCLE CRASH PREDICTION FOR URBAN ROAD SEGMENTS**

**Final Report**

**NITC-RR-756**

by

Krista Nordback  
Sirisha Kothuri  
Geoff Gibson  
Portland State University

Wesley Marshall  
Nick Ferenchak  
University of Colorado Denver

for

National Institute for Transportation and Communities (NITC)  
P.O. Box 751  
Portland, OR 97207



**May 2018**



## Technical Report Documentation Page

1. Report No. NITC-RR-756	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle  IMPROVING BICYCLE CRASH PREDICTION FOR URBAN ROAD SEGMENTS		5. Report Date May 2018	
		6. Performing Organization Code	
7. Author(s)  Krista Nordback, Sirisha Kothuri, Geoff Gibson, Wesley Marshall, Nick Ferenchak		8. Performing Organization Report No.	
9. Performing Organization Name and Address		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No.	
12. Sponsoring Agency Name and Address  National Institute for Transportation and Communities (NITC) P.O. Box 751 Portland, OR 97207		13. Type of Report and Period Covered	
		14. Sponsoring Agency Code	
15. Supplementary Notes			
<p>16. Abstract</p> <p>The 2010 Highway Safety Manual (HSM) provides methods for predicting the number of motor vehicle crashes on various roadway facilities (AASHTO, 2010). However, it includes only a simplistic method for predicting the number of bicycle-related crashes. The research team investigated the bicycle-specific crash data in eight potential study areas around the U.S.: Arlington, VA (city/county); Bellingham, WA (city); Boulder, CO (city); Denver, CO (city/county); Minneapolis and St. Paul, MN (cities); Philadelphia, PA (city/county); Portland, OR (city); and San Diego, CA (county). The available online data from each were compared. The study city for future analysis (Boulder) was selected based on the availability of not just crash data but also the availability of continuous and short-duration bicycle and pedestrian traffic count data.</p> <p>In this analysis, a negative binomial model with log link was used to predict annual, non-fatal, motorist-bicyclist crashes on road segments per mile. This report adopts methods from the HSM used for motor vehicle safety performance functions (SPFs) in order to develop bicycle-specific SPFs for roadway segments in Boulder (9). The analysis shows that motor vehicle volume is a leading factor associated with more crashes between motor vehicles and bicyclists. Bicyclist exposure, population density, and percent retail land use are also predictive. While both vehicle volume and bicycle volume data are used in the model in order to account for the "safety in numbers" effect, the model did not demonstrate this effect that is seen so commonly in other research, including the bicycle SPF developed previously for intersections in Boulder (11). This effort at developing a bicycle-specific SPF for segments in the U.S. that utilize bicycle volumes is an important first step towards further understanding bicyclist safety and may inform future versions of the HSM. To that end, the report includes a table of motorist-cyclist crashes predicted by the model for various values. The authors hope this table may serve as a potential format template and starting point for future efforts to generalize the results of models for possible use in HSM updates.</p>			
17. Key Words  Bicycle, Safety Performance Function, Pedestrian, Negative binomial, population density, AADT		18. Distribution Statement No restrictions. Copies available from NITC: <a href="http://www.nitc.us">www.nitc.us</a>	
19. Security Classification (of this report)  Unclassified	20. Security Classification (of this page)  Unclassified	21. No. of Pages 73	22. Price

## **ACKNOWLEDGEMENTS**

This project was funded by the National Institute for Transportation and Communities (NITC), a federally funded University Transportation Center housed at Portland State University's Transportation Research and Education Center (TREC). Matching funds were provided through work with the Oregon Department of Transportation. We thank the City of Boulder staff and the Denver Regional Council of Governments (DRCOG) for making the data available.

## **DISCLAIMER**

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the material and information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation University Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the U.S. Government. This report does not constitute a standard, specification, or regulation.

## **RECOMMENDED CITATION**

Nordback, Krista, Kothuri, Sirisha, Gibson, Geoff, Marshall, Wesley, and Ferenchak, Nick. *Improving Bicycle Crash Prediction for Urban Road Segments*. NITC-RR-756. Portland, OR: Transportation Research and Education Center (TREC), 2018.

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>1</b>
<b>1.0 INTRODUCTION.....</b>	<b>2</b>
<b>2.0 LITERATURE REVIEW .....</b>	<b>5</b>
2.1 EXPOSURE MEASURES .....	5
2.1.1 Population .....	5
2.1.2 Volume.....	6
2.1.3 Time .....	6
2.1.4 Distance.....	6
2.1.5 Trip.....	7
2.2 BICYCLE SAFETY AT THE FACILITY LEVEL .....	7
2.3 BIKE SAFETY STUDIES WITH BIKE VOLUME AND SAFETY PERFORMANCE FUNCTIONS .....	10
2.4 SUMMARY .....	12
<b>3.0 CRASH DATA EXPLORATION .....</b>	<b>13</b>
3.1 BACKGROUND .....	13
3.1.1 Motorist-cyclist Crash Data Standards and Tools .....	14
3.2 STUDY COMMUNITIES .....	16
3.2.1 Observed Issues with Crash Data .....	17
3.2.2 Inconsistent Data Formats.....	17
3.2.3 Differences in Data Fields Between Communities.....	18
3.2.4 Differences in Data Coding Within Communities.....	21
3.2.5 Underreporting.....	22
3.2.6 Motor-vehicle-specific Reporting Forms.....	22
3.3 SUMMARY .....	25
<b>4.0 DATA AND METHODS .....</b>	<b>27</b>
4.1 SITE LOCATION.....	27
4.1.1 Crash Data.....	27
4.1.2 AADB .....	29
4.1.3 AADT .....	30
4.1.4 Pedestrian Volume .....	31
4.1.5 Sociodemographics.....	34
4.1.5.1 Population Density.....	34
4.1.5.2 Employment Density .....	34
4.1.5.3 Educational Attainment .....	35
4.1.6 Land Use .....	35
4.1.7 Facilities .....	35
4.2 METHODS .....	36
4.3 SUMMARY .....	37
<b>5.0 RESULTS .....</b>	<b>39</b>
5.1 DISCUSSION.....	41
5.2 INTERPRETATION.....	41
5.2.1 AADT .....	42
5.2.2 Percent Retail .....	43
5.2.3 Population Density.....	44

5.2.4	Model Fit.....	45
5.2.5	Generalizability.....	47
5.3	LIMITATIONS.....	49
5.4	SUMMARY.....	49
<b>6.0</b>	<b>CONCLUSIONS AND RECOMMENDATIONS.....</b>	<b>50</b>
6.1	RECOMMENDATIONS.....	51
6.2	FUTURE WORK.....	51

## APPENDICES

APPENDIX A: DATA PREPARATION METHODS

APPENDIX: DATA DICTIONARY

## LIST OF FIGURES

Figure 1.1:	Cyclist Fatalities and Mode Share in the U.S.....	2
Figure 3.1:	Percent of Crashes with Intersection Type, Lighting, and Weather.....	20
Figure 3.2:	Police Traffic Collision Report Form from Washington State.....	24
Figure 4.1:	Bicycle Volume and Non-fatal, Motorist-bicyclist Crash Map for Boulder.....	32
Figure 4.2:	Automobile Volumes Map for Boulder.....	33
Figure 5.1:	CURE Plot for Selected Model.....	40
Figure 5.2:	Histogram of AADT.....	42
Figure 5.3:	Histogram of Adjusted AADB.....	43
Figure 5.4:	Histogram of Land Use Percent Retail.....	44
Figure 5.5:	Histogram of Population Density per Square Mile.....	45
Figure 5.6:	Predicted and Actual Crashes vs. Adjusted AADB.....	46
Figure 5.7:	Predicted and Actual Crashes vs. AADT.....	47

## LIST OF TABLES

Table 3.1:	Motorist-cyclist Crash Data Fields from Literature.....	14
Table 3.2:	Example PBCAT Motorist-Cyclist Crash Data Types (Harkey, et. al, 2006).....	15
Table 3.3:	Crash Data for Each Community.....	17
Table 3.4:	Crash Data Variables Collected in Study Areas.....	19
Table 3.5:	PBCAT Crash Types in Boulder.....	23
Table 4.1:	Summary Statistics for the Non-fatal, Motorist-bicyclist Crashes Studied (only crashes on the 174 segments with traffic volume data are included).....	29
Table 5.1:	Final Model Estimated Parameters.....	40
Table 5.2:	Range of Predicted Motorist-bicyclist, Nonintersection Crashes (per mile per year). 48	



## EXECUTIVE SUMMARY

The 2010 Highway Safety Manual (HSM) provides methods for predicting the number of motor vehicle crashes on various roadway facilities (AASHTO, 2010). Safety performance functions (SPFs) are able to estimate the relationship between collisions and exposure by accounting for the nonlinear relationship between exposure and risk. While SPFs have been comprehensively developed for motor vehicles, the HSM includes only a simplistic method for predicting the number of bicycle-related crashes. Despite research demonstrating that bicycle volume is an important factor in estimating the number of bicycle crashes, the existing bicycle-related crash method does not include the volume of bicyclists using the roadway (Nordback et al., 2014). Thus, there is a critical need to further develop bicycle-specific SPFs. While a few studies have created SPFs for bicycle crashes at intersections, no studies have analyzed bicycle crashes along road segments.

To create models to study bicycle crashes along road segments, robust safety data are needed. After a review of the bicycle-safety literature, the first step was to investigate possible data sets for study. The research team investigated the bicycle-specific crash data in eight potential study areas around the U.S.: Arlington, VA (city/county); Bellingham, WA (city); Boulder, CO (city); Denver, CO (city/county); Minneapolis and St. Paul, MN (cities); Philadelphia, PA (city/county); Portland, OR (city); and San Diego, CA (county). The available online data from each location were compared. Boulder was selected for further study based on the availability of both crash data and count data (continuous and short-duration bicycle and pedestrian traffic count data).

To create a bicycle-specific SPF for urban roadway segments, the team used eight years (2006-2013) of crash data from Boulder. In this analysis, a negative binomial model with log link was used to predict annual, non-fatal, motorist-bicyclist crashes on road segments per mile. The analysis shows that motor vehicle volume is a leading factor associated with more crashes between motor vehicles and bicyclists. Bicyclist exposure, population density, and percent retail land use are also predictive. This study also investigated the potential of various simplified methods to include bicycle volumes in future versions of the HSM. The project examined how and if the SPFs developed could be simplified for use in future versions of the HSM. A table was created that simplified motor vehicle volumes, bicycle volumes, retail land use percentages, and population density into two or three categories (high, medium, and low) for each variable. This simplified format may reduce the need for difficult-to-find data, such as bicycle volume, and may be of help to practitioners who seek to predict the number of motorist-bicyclist crashes on a road segment. The research team provided the recommended table as a possible format for consideration in the HSM update as a way to generalize the findings of this study for HSM users.

While this research cannot provide a final HSM-ready version since we have only studied one city, it did present the results of this study in a readily usable table, investigate locations where data are available, and can inform ongoing NCHRP project 17-84 Pedestrian and Bicycle Safety Performance Functions for the HSM (Rogers, 2019). Such SPFs can help prioritize projects and inform the transportation decision-making process and future editions of the HSM.

# 1.0 INTRODUCTION

As a sustainable transportation mode, cycling offers many benefits such as lower congestion and emission levels as well as improvements in personal health via enhanced physical activity. In 2016, 840 cyclists were killed on our nation’s roadways, accounting for over 2 percent of traffic fatalities (NHTSA, 2018). According to the American Community Survey one-year estimates, cyclists accounted for only 0.6 percent of journey-to-work mode share in 2016 (U.S. Census Bureau, 2018). Thus, cyclists are overrepresented in terms of fatalities. From 2005 to 2008, cyclist mode share in journey-to-work increased while fatalities decreased, but from 2009 to 2016 the percent of those traveling to work by bicycle stayed roughly constant while cycling fatalities climbed. Overall cyclist fatalities and commute mode share can be seen in Figure 1.1. Addressing cyclist safety is especially urgent given the USDOT push to establish safety performance measures in its recently updated strategic plan (FHWA, 2017).

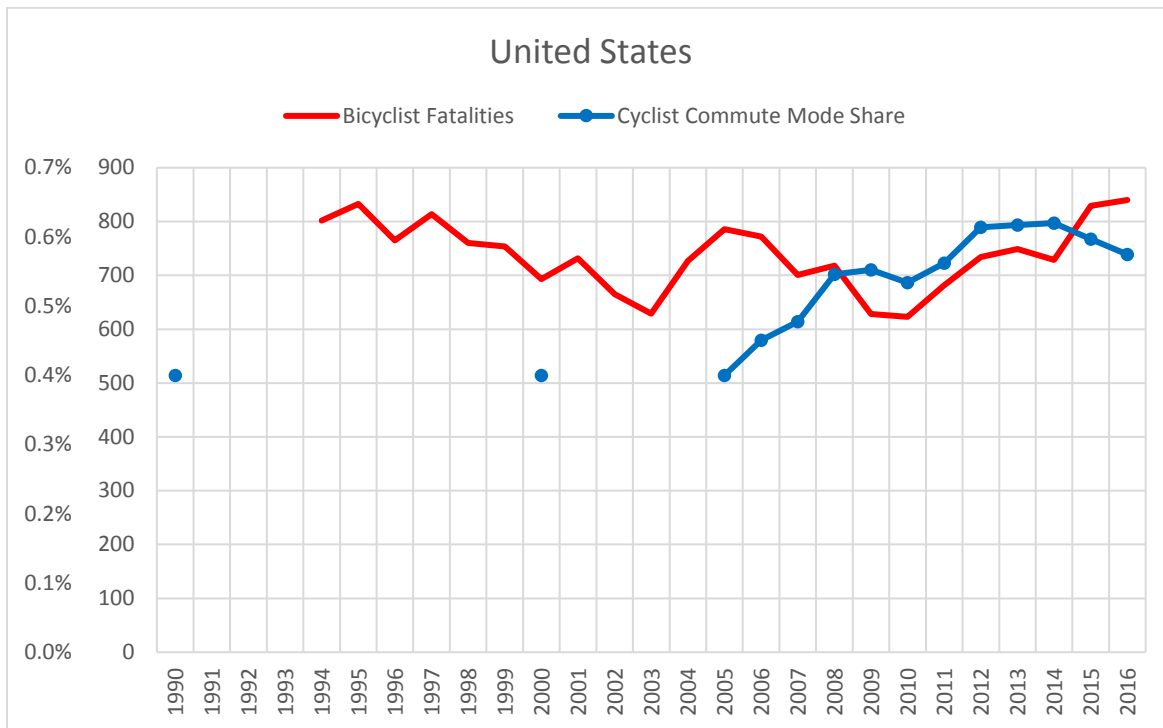


Figure 1.1: Cyclist Fatalities and Mode Share in the U.S. (NHTSA, 2018; U.S. Census Bureau, 2018)

Cyclist safety studies continue to mount, but a lack of both bicycle crash data and bicycle traffic volumes impedes what would be considered methodologically robust bicyclist safety research. This is especially apparent concerning studies that show the impact of specific infrastructure changes on cycling safety. Such research is often used to estimate crash modification factors (CMF) based on simple multiplicative factors that estimate the potential reduction in crashes due to a specific infrastructure change. Of the 136 bicycle-related CMFs listed in the online CMF Clearinghouse, there are currently no CMFs of sufficient quality to warrant the four- or five-star rating needed to earn safety improvement funds (McKenzie et al., 2014). This illustrates a basic

lack of knowledge by the field on what infrastructure changes improve cyclist safety for a given bicyclist volume. Despite the lack of data, there are many recent and ongoing efforts to understand cyclist safety, including NCHRP 17-84 and NCHRP 15-63 (USDOT, 2017). However, without adequate data, these efforts will be hampered.

While a lack of crash and exposure data continues to be a hindrance to U.S. cyclist safety research, studies from other countries can more commonly be conducted at the facility level. Canadians, Australians, Swedes, Danes, and other northern Europeans are studying cyclist safety and finding that, while cyclist infrastructure may or may not increase cyclist safety, one rule holds: the more cyclists there are, the fewer collisions per cyclist (Elvik, 2009). Peter Jacobsen brought this to the attention of the U.S. audience in his 2003 paper “Safety in Numbers,” which took a city-wide perspective (Jacobsen, 2009). In 2012, Nordback et al. (2014) created the first bicycle SPFs for signalized intersections in a U.S. city using crash data and exposure data for one example city and found that risk for cyclists decreases when more cyclists use an intersection.

This research investigated data availability in eight U.S. cities: Arlington, VA; Bellingham, WA; Boulder and Denver, CO; Minneapolis/St. Paul, MN; Philadelphia, PA; Portland, OR; and San Diego, CA. Note that Seattle, WA was specifically omitted because there were at least two bicycle safety research projects being conducted during the same time as this study, and the team wanted to avoid duplication of effort (Thomas et al., 2017; Chen 2015). The team first investigated the bicycle crash data available in each city and then also examined the bicycle volume data available, both short-duration counts and permanent counts. Based on data availability, one study city was selected that had all necessary data: Boulder. Since bicycle crashes at intersections had been previously studied in this city, crashes on road segments were investigated for this report.

The work began with estimation of annual average bicyclists on road segments using a combination of continuous count data and short-term intersection counts from which seasonal adjustment factors were computed. Next, sociodemographic factors such as population density, employment density, educational attainment, pedestrian volumes, land use, and type of facilities were combined with the crash data, and negative binomial models were estimated. In previous work based in Boulder, the impact of cyclist volume was apparent (Nordback et al., 2014). The same was not as clear for the road segments in this study, though the strong positive relationship of motor vehicle volume with motorist-cyclist crashes is clear in both studies.

This work will allow transportation professionals to better understand the relationship between bicycle and motor vehicle traffic volumes and bicyclist collisions. This is fundamental to understanding the impact of various facilities and other factors on cyclist safety. The authors intend that the results can inform future versions of the Highway Safety Manual as well as ongoing research project NCHRP 17-84. To facilitate that, we present a table of expected ranges of number of motorist-cyclist crashes for different road-segment types based on the model developed. While using a model developed purely from Boulder data to predict crashes in other places and other circumstances is not recommended, the table introduces one approach to providing a simple bicycle-motorist crash estimation tool. If more cities and location types can be studied, this approach might be an appropriate modification to the HSM for bicycle-motorist collision prediction.

The remainder of this report is laid out in the following manner: A review of the literature is next, followed by case studies of eight communities where bicycle crash and bicycle volume data were available, as well as a description of the data availability for each city. This is followed by a description of the analysis of motorist-bicyclist crashes on road segments in the study city, Boulder. Lastly, we present results, conclusions, and recommendations.

## **2.0 LITERATURE REVIEW**

In the past few years, there have been many cyclist safety studies, but few include safety performance functions (SPFs) for bicyclists. SPFs predict the average number of crashes per year at a location as a function of exposure and roadway or intersection characteristics (AASHTO, 2010). The Highway Safety Manual (HSM) provides SPFs for motor vehicle crashes; however, SPFs for bicycle crashes are not well developed (Nordback et al., 2014). The HSM recommends computing the predicted number of crashes for bicycles by multiplying predicted motor vehicle collisions by a factor that is based on motor vehicle speed and road type. The main drawback of the HSM method is that measures of cyclist exposure are not considered in the estimation. The biggest challenge with developing bicycle-specific SPFs is the lack of exposure data. Bicycle volumes are rarely available at many locations, and this precludes the creation of SPFs that consider cyclist exposure.

The objectives of this study are to investigate improved methods for cyclist exposure for collisions involving bicycles and motor vehicles; understand the impact of factors such as traffic volumes, facility type and land use on cyclist safety; and develop statistical models to predict the number of cyclist collisions. This study will not include collisions between cyclists and pedestrians or single-bicycle crashes. Pertinent literature on cyclist exposure measures, cyclist safety at the facility level, and cyclist safety analysis that includes cyclist traffic volume are reviewed in the following sections.

### **2.1 EXPOSURE MEASURES**

In road safety literature, exposure is defined as proximity to potentially harmful situations involving motor vehicles (Molino et al., 2012). To adequately understand cyclist safety, it is necessary to determine exposure. Risk is defined as the probability that a crash will occur based upon that exposure (Molino et al., 2012). Typical exposure measures that have been used in past literature are population measures, volumes, and travel measures such as time, distance, and per trip (Vanparijs et al., 2015). Each of these approaches are discussed below.

#### **2.1.1 Population**

The primary assumption with population measures is that more crashes occur in areas with higher population density than in areas with lower population density (Molino et al., 2012). Traditional statistics produced by NHTSA provide cyclist fatality rates by population (per 10,000 residents), but these metrics are not sensitive to the amount of time or distance that the cyclist is exposed to vehicular traffic (Molino et al., 2012). Another study conducted by Rodgers estimated cyclist population using data from a survey conducted on bicycle use (Rodgers et al. 1994). Based on the estimated population, a fatality rate was computed per one million cyclists. However, as stated earlier, this metric is also insensitive to local conditions and density of riders. McAndrews et al. (2013) used data from the 2001 National Household Travel Survey (NHTS) to estimate fatality and injury rates for all users per 100,000 population.

### **2.1.2 Volume**

Another measure of exposure that is often used is cyclist volume. Annual average daily bicyclists (AADB) can be used as an exposure metric, which is estimated from bicycle counts. Nordback et al. (2013) used continuous count stations in Colorado to develop factors and applied these factors to locations with short-duration counts to estimate AADB. Other studies investigated AADB estimation accuracy when varying factors are used (Miranda-Moreno and Nosal, 2011; El Elsawey, 2014; Hankey, 2014). The biggest challenge with the factoring approach for estimating AADB is developing appropriate factor groups that are representative of the cycling patterns. Others, such as Schneider, estimated exposure using models for pedestrians that can predict crossing volumes using variables such as land use and other site characteristics (Schneider et al., 2009). Similar concepts could be extended to estimate bicycle volume. NCHRP-770 (2014) also reviewed a number of methods for estimating cycling and walking volumes for planning and project development.

### **2.1.3 Time**

The amount of time that a cyclist or a pedestrian spends in certain activities can also be considered as a measure of exposure. A few studies used perceived trip duration information to estimate exposure, especially for walking trips (Keall, 1995; Chu, 2003; Bly et al., 1999). Trip duration in each of these studies was derived from travel survey data. However, the limitation of this approach is that the amount of time spent walking or bicycling is not sensitive to the facility. For example, cyclists may ride on trails or sidewalks without any motor vehicles; however, using trip duration as an exposure metric will result in exposure being overestimated (Molino et al., 2012). McAndrews et al. (2013) computed fatality and injury risk for various users using both travel and population-based exposure measures. They found that for cyclists, the fatal and injury risk estimated using a time-based exposure measure was lower compared to the risk estimated using distance and trip-based exposure measures.

### **2.1.4 Distance**

Another exposure metric that is used in the literature is distance traveled. Kaplan (1975) conducted a survey of 3,270 adult cyclists over one year and used bicycle miles traveled as a measure of exposure. However, it did not differentiate between distance traveled on shared facilities versus distance traveled on exclusive bicycle facilities. Therefore, exposure might be overestimated.

Molino et al. (2009) proposed a hundred million bicycle miles of shared facilities traveled as a measure of exposure. Observers collected cyclist volumes and crossing distances along different facilities, and annual bicycle volumes and travel distances for each type of facility shared with motor vehicles were derived through aggregation and summation of the collected data (Molino et al., 2012). The shared facilities are defined as those facilities where bicyclists and pedestrians share the space with motor vehicles. McAndrews et al. (2013) found that cyclists face 10 times the risk of fatality or hospitalized injury compared to motor vehicle occupants when exposure is measured based on per person-mile of travel ().

### 2.1.5 Trip

McAndrews et al. (2013) computed a trip-based measure of exposure across different populations of travelers in Wisconsin. Their results showed that fatality and injury risks to cyclists are twice as high compared to risks for motor vehicle occupants when measured per person by trip. According to McAndrews, using a trip-based measure of exposure normalizes the differences in speed between the modes.

## 2.2 BICYCLE SAFETY AT THE FACILITY LEVEL

With increased focus on cycling, infrastructure and facilities designed for cyclists have gained increased attention in the United States. The Model Inventory of Roadway Elements (MIRE) provides the following list of types of bicycle facilities: “None, Wide curb lane with no bicycle markings, Wide curb lane with bicycle markings (e.g., sharrows), Marked bicycle lane, Separate parallel bicycle path, Signed bicycle route only (no designated bicycle facility), and Other” (MIRE FHWA Version 1, 2010).

In Europe, however, separate facilities for cyclists have been in existence since the 1970s. The Dutch CROW design manual for bicycle traffic, NACTO's urban bikeway design guide, and the Federal Highway Administration outline the different types of infrastructure and facilities for cyclists (Dutch CROW, 2006; NACTO, 2014). In particular, they outline seven types of bicycle infrastructure:

- Bike Lanes - Defined as a part of a road that's designated for bicycle use (NACTO, 2014). These can be:
  - Conventional bike lanes with pavement markings
  - Buffered bike lanes with a designated space to provide separation
  - Contra-flow bike lanes which run in the opposite direction of vehicle traffic
  - Left-sided bike lanes that run on the left side of a one-way road or two-way divided streets.
  
- Cycle Tracks - These are defined as a physically separated lane that is used exclusively by bicycles (NACTO, 2014). These can be:
  - A one-way protected cycle track that exists at the street level but provides physical protection from automobiles
  - Raised cycle tracks that are separated from automobiles by a vertical barrier
  - Two-way cycle tracks that allow for bicycle traffic to flow in both directions and continue to be separated physically from automobile traffic.
  
- Intersection Treatments - These are defined as facilities that exist within or near intersections that provide additional safety for bicyclists. This category can include:
  - Bike boxes, which are a designated area for bicyclists upstream of a stop bar at an intersection that are designed to increase visibility for cyclists and allow them to get ahead of automobile traffic

- Intersection crossing markings, which are designed to guide bicyclists across an intersection
- Two-stage turn queue boxes that allow for a safe method for bicyclists to make a left turn at multilane signal intersections
- Median refuge islands that offer a physically protected space in the street to help bicyclists and pedestrians cross
- Combined bike lane and turn lanes, which place a bike lane within a section of a turn lane that allows for the bicyclist to continue forward through a traffic light
- Cycle track intersection approaches, which allow for guidance of a bicyclist at an intersection approach to reduce turn conflicts (NACTO, 2014).
- Bicycle Signals - These are defined as independent signals and beacons that allow for a bicyclist to safely cross roadways and larger streets (NACTO, 2014). These signals include:
  - Active warning beacons at unsignalized intersections, which are user-actuated flashing lights that signal to automobile traffic that a pedestrian and/or cyclist will cross the street, oftentimes accompanied by existing signage
  - Bicycle signal heads, which are independent signals used to give bicyclists guidance on when it is safest for them to proceed through an intersection
  - Hybrid beacons that consist of a dual red light over a single yellow light on a major street for pedestrian and bicycle crossing, and signal detection/actuation, which is a sensor that can be activated by a bicycle to alert the signal of a bicycle needing to cross.
- Bikeway Signing and Marking - Defined as any visual treatment that is used to indicate the use of the road by bicyclists. These can include: colored bike facilities that add a visual difference for the bicycle lane to identify conflict areas or to reinforce the right of bicyclists to be in the area; shared lane markings, which inform automobile drivers that the lane can be occupied at any given time by bicyclists; and bike route wayfinding signage and marking systems, which allow for bicyclists to find where the nearest bike route is and get them to their destination in a safe manner (NACTO, 2014).
- Bicycle Boulevards - These are streets designed to give bicycle traffic priority through the use of lower speed limits, automobile traffic diversions, and volume management (NACTO, 2014). This can include:
  - Route planning, which prioritizes low-volume streets over higher-volume streets
  - Bicycle boulevard signs and markings, which help to indicate to a bicyclist that they are on a bike boulevard and to indicate to automobile drivers that it is a shared street
  - Speed management to reduce the speed of automobiles to prevent cars from passing bicycles and reduce the effect of a collision should one occur
  - Volume management, which can include physical barriers to automobiles, forcing them to larger streets but allowing bicycles to ride through



- Minor street crossings to inform automobile drivers that wish to cross a bike boulevard that bicyclists and other cross traffic will not stop
  - Major street crossings, which allow for increased visibility and a reduction of time delay for cyclists wishing to cross
  - Offset intersections that allow for a non-continuous street to direct bicycle flows towards safer streets and then back to their original direction
  - Green infrastructure, which can be integrated with speed and volume management treatments.
- Sidewalks - These are paths that are intended for use primarily with pedestrians, located between the edge of the roadway and the edge of the right-of-way or the nearest property line. Although sidewalks are intended for use by pedestrians only, bicyclists have been known to use them despite sidewalk riding being illegal in many places (Federal Highway Administration, 2014).
  - Shared Use Paths - Paths that run parallel to roads but are physically separated from automobile traffic in some manner (e.g., wall, raised curb) that are also intended for use by both bicycles, pedestrians, and other recreational forms of transportation. There is no separation between bicycle traffic and other forms of traffic (Federal Highway Administration, 2014).
    - On-street shared use paths - Shared use paths that run parallel to roads and operates as a wide sidewalk.
    - Off-street shared use paths - Paths that offer transportation to bicycles and other non-motorized modes by way of a path that runs entirely separate from a road or highway. This would include trails that run in their own right-of-way such as rail-trails, as well as paths paralleling limited access highways where they do not intersect with roadways.

As stated earlier in the literature review, studies have shown that cyclist injuries are dependent on the number of cyclists present. If that is true, then the comfort level of cyclists must also be taken into account since more comfortable and physically separated facilities may attract more cyclists. According to a study by McNeil et al. (2014), 71 percent of residents living near newly built bike lanes stated that they would consider riding a bicycle if they were separated from motor vehicles by a barrier. Additional research conducted by Goodno et al. (2013) and Zangenehpour et al. (2014) shows that buffered bike lanes and cycle tracks provide an overall increase in the safety and comfort level of cyclists, leading to fewer collisions overall as well as a significant increase in the number of cyclists .

According to NACTO (2014), bike lanes offer a predictable environment for bicyclists to traverse shared roads with automobiles. That said, while conventional bike lanes offer only a single painted line to delineate between the automobile path and bicyclist path, a buffered bike lane offers increased comfort by creating more space between the bicyclist and automobile driver. Increasingly, buffered bike lanes are being preferred over conventional bike lanes due to this increased space (McNeil et al., 2014). A study by Akar and Clifton (2009), however, pointed out that the discontinuous nature of bike lanes in general created a problem of safety and reliability of the network at large, regardless of whether the lanes were buffered or conventional.

Cycle tracks, while offering increased comfort for cyclists, also have tangible safety benefits such as reducing the severity of collision injuries and reducing the likelihood of collisions at intersections, according to multiple studies (Thomas et al., 2012; Zangenehpour et al., 2014). There are some exceptions, however. One study found that two-way cycle tracks were more likely to have more collision incidents per mile traveled compared to cycling on the street (Thomas et al., 2012). Additionally, two studies found that cycle tracks can increase the amount of cyclist collisions at or near intersections. One study cited cycle tracks as being three times more dangerous but also concluded that this is due to an overall increase in the number of cyclists (Thomas et al., 2012; Zangenehpour et al., 2014). Overall, cycle tracks on the right side of the street were found to increase safety and decrease collision incidents; however, cycle tracks on the left side of the street did not have any significant decrease compared to a street without a cycle track (Zangenehpour et al., 2014). According to Harris et al. (2013), cycle tracks, in combination with lower speeds for automobiles, were found to significantly decrease collisions at intersections in both Vancouver and Toronto, Canada, compared to intersections without those same treatments. Finally, one study by Lusk et al. (2011) purposefully studied cycle tracks in Montreal in an attempt to dispute the long-held idea that cycle tracks are less safe according to the American Association of State Highway and Transportation Officials. After studying six different cycle tracks in Montreal, they found that all the cycle tracks were safer for cyclists than had the cycle tracks not existed at all.

The Dutch CROW Design Manual for Bicycle Traffic cites that accidents with bicyclists at intersections account for over half of all collision and, as such, require special requirements in order to deal with them safely (Dutch CROW, 2006). One common intersectional treatment designed to increase visibility and help with turning in Europe, and increasingly the U.S., is the bike box. According to a study by Dill et al. (2010), the number of collision incidents between cars and bicyclists, as well as pedestrians, decreased and automobile yielding increased where bike boxes were installed in Portland, OR.

### **2.3 BIKE SAFETY STUDIES WITH BIKE VOLUME AND SAFETY PERFORMANCE FUNCTIONS**

According to the Federal Highway Administration, a safety performance function (SPF) is a formula that can be used to predict the amount of collision incidents per year at a specific location due to exposure and characteristics of the road or intersection. While SPFs have mostly been used for automobile traffic, development of a SPF for predicting bicycle collisions is limited. In recent years, a few different research studies have started to look at SPFs specifically for bicycles.

Many studies have shown that the risks of injury to cyclists are nonlinear and dependent on the number of cyclists, essentially indicating a decreasing risk with increasing bicycle traffic (Larsson, 1993; Leden et al., 1998; Lyon and Persaud, 2002; Jacobsen, 2003; Robinson, 2005; Jonsson, 2005; Geyer et al., 2006; Harwood et al., 2008; Elvik, 2009; Miranda-Moreno et al., 2011; Schepers et al., 2011; Turner et al., 2011; Carlson et al., 2018). A summary of studies prior to 2009 was reported by Elvik (2009). Perhaps the most well-known study in this category was by Jacobsen in 2003. Using data from California, Denmark, and eight other European countries,

Jacobsen showed that the probability of a motorist-pedestrian or motorist-cyclist crash declines with the approximately -0.6 power of the number of persons walking or cycling. In other words, as the number of people walking and cycling increases, the less likely they are to be injured by motorists. This suggests a “safety in numbers” phenomenon (Jacobsen et al. 2003).

Most recently, Nordback et al. (2014) applied a bicycle-specific SPF to the city of Boulder, CO. It was the first of any such study conducted in the United States. Using their estimated SPFs, Nordback et al. found that bicycle volumes had a direct impact on the number of bicycle collisions in intersections, primarily in that more bicycles meant fewer collisions per cyclist. That said, it was made clear in the study that this specific SPF was not meant to be a generalization of all U.S. cities, but rather specific to Boulder. A study by Turner et al. (2011) applied bicycle-specific SPFs in both the Australian city of Adelaide and the New Zealand city of Christchurch. There, Turner et al. conducted a before-and-after study on the use of bicycle facility treatments with the use of SPFs. Their research found that the effect of bicycle lanes depended on the standards with which they were built (Turner et al., 2014). They also found that colored bicycle lanes decreased collisions by as much as 39percent. However, Turner et al. did not take note of whether bicycle volumes were higher after bicycle safety features were installed. A third study by McArthur et al. (2011) used SPFs to calculate the crash prediction rating of schoolchildren who were riding their bikes to and from schools. In it, the authors looked at Safe Routes to School (SRTS) programs and combined that with SPFs where SRTS had funded additional infrastructure and educational improvements. Specific results found that bicycle collisions increased with the number of students, and that schools on smaller streets had more collisions than schools on larger arterials. A recent study by Thomas et al. (2017) developed SPFs for three high-frequency and high-severity bicycle intersection crash types – total bicycle crashes at intersections, crashes in which the bicyclist and driver were traveling in the opposite directions, and crashes in which the bicyclist and driver were traveling perpendicular to each other. Factors that were significant in all three models included arterial classification, larger number of lanes, and two-way center turn lanes. The findings also revealed that pedestrian volume had a curved relationship with crashes, indicating that as the pedestrian volume increased above some threshold, a decreasing relationship with bicycle crashes was observed (Thomas et al., 2017).

Outside of SPF-specific studies, one study used a Bayesian model to research bicycle injury occurrence and bicycle safety at signalized intersections in order to find factors associated with volumes of bicyclists and injuries related to them (Strauss et al., 2012). In their methodology, Strauss et al. found that the flow of bicyclists and motor vehicles as well as geometric design of streets, traffic controls, and other built environment characteristics all contribute to bicycle injuries through intersections and corridors. An increase of 10percent volume of right- and left-turning automobile traffic was found to increase bicycle injury rates by 2.4percent and 1.9percent, respectively. Additionally, Strauss et al. found that low cyclist volume corridors through Montreal were the worst in terms of injury rates. In conclusion, the Bayesian model found that bicycle volumes had the greatest impact on cyclist injury and safety through both corridors and intersections (Strauss et al., 2012). Another study by Shepers et al. (2013) created a framework for road safety and applied it to bicycle safety. They looked at a number of safety studies and used their created framework to test the validity of those studies. One particular section found that studies on traffic volumes and risk were nonlinear using a framework for road

safety and that the type of cyclists (e.g., women, elderly, etc.) plays a larger role in bicycle safety than some studies have suggested. Shepers et al. suggest that future studies need to explore the causal mechanisms of cyclists, which could help explain the link between bicycle usage and risk.

It is worth highlighting that Elvik completed a literature review of the nonlinearity of risk in bicycle transportation that covered volume of bicyclists. Elvik found that as the number of pedestrians and cyclists goes up, the risk involved goes down. Additional studies looked at by Elvik found that if automobile drivers switched to walking or bicycling, the overall number of collisions would go down (Elvik, 2009).

Finally, a dissertation by Lars Ekman (1996) analyzed the effect of treatments in flows on safety and accidents. Ekman came to the conclusion that moving more bicyclists from smaller streets to larger arterials can actually be beneficial overall and that the use of SPFs would enhance any safety study in the future. Another dissertation by Thomas Jonsson (2005) looked at creating a model for predicting accidents with special emphasis on vulnerable road users (VRU), most often bicyclists and pedestrians. In it, Jonsson was able to create a model for successfully predicting 71-81percent of the systematic variation of accidents involving VRUs.

## **2.4 SUMMARY**

Both population- and travel-based exposure measures have been used in the literature to study the risk associated with cycling. According to McAndrews et al. (2013), risk estimated based on population-based exposure measures reflects overall societal risk while estimates based on travel exposure measures reflect travel risk. Trip-based measures normalize speed across modes, whereas the distance-based measures reflect the slower speeds of active transportation modes (McAndrews et al., 2013). Additionally, facilities have been studied extensively as they relate to bicycle safety. Bike lanes offer a predictable space for bicyclists and automobiles to travel together but are otherwise considered not as preferable as buffered bike lanes or as safe as cycle tracks (McNeil et al., 2014; Thomas et al., 2012; Monsere et al., 2014; Zangenehpour et al., 2014). Alternative bicycle facility treatments, such as bike boxes and median refuge islands, were also found to be important safety features, though literature on these subjects needs to be expanded (Dill et al., 2010). Four separate studies used SPFs to analyze bicycle safety and came away with varied results. One study by Nordback et al. (2014) found that bicycle volumes in Boulder had a direct impact on bicycle collisions. Another study used SPFs to gauge the effectiveness of bicycle treatments and found that colored bicycle lanes decreased collisions by 39percent (Turner et al., 2011). A third study used SPFs to look at the effectiveness of the SRTS program and treatments involved with those areas and found that bicycle collisions increased with the amount of students and that schools located on local streets had higher frequencies of crashes than larger streets (McArthur et al., 2011). A recent study developed SPFs for common bicycle-vehicle crash types in Seattle (Thomas et al., 2017). A final study applied a Bayesian model to cyclist injuries at intersections and corridors and found similar results, namely that bicycle volumes had the greatest impact on collisions and injuries (Strauss et al., 2012).

As the literature review revealed, studies involving the use of the SPFs are limited, and at the time of writing, the authors are not aware of any study that has developed SPFs for bicycle crashes and bicycle volume for a U.S. city along road segments. More research is needed to comprehensively understand cyclist safety.

## **3.0 CRASH DATA EXPLORATION**

Prior to choosing locations to study, the team reviewed existing data sources by community in order to find study areas with consistent data, or at least to identify one area with sufficient data for a bicycle-safety study. Three main types of data are needed for the development of bicycle-specific crash relationships: crash data, exposure data, and data about the environment where crashes can occur (e.g., facility types, land uses, weather, etc.). This section focuses on the first of these: crash data. Within the non-fatal crash data set involving cyclists, bicycle crashes involving motorists are the specific focus of this study, as these crashes have the potential to be more severe and more commonly reported than bicycle-only crashes, (Schepers, 2011). Our research team used non-fatal cyclist crash data from eight different communities: Arlington, VA (city/county); Bellingham, WA (city); Boulder, CO(city); Denver, CO (city/county); Minneapolis and St. Paul, MN (cities); Philadelphia, PA (city/county); Portland, OR (city); and San Diego, CA (county). Our research team examined these data for issues pertaining to quality and consistency. To date, there is little to no research about how bicycle crash data are collected, organized, and reported.

In this section, we aim to show the different ways in which communities record their non-fatal crash data for cyclists, along with a discussion of the strengths and weaknesses of the data and the collection process, with the overall goal of bringing attention to the issues and provide recommendations for improvement. First, we offer a brief review that identifies the pertinent variables that have been used in previous crash studies and which would be relevant for inclusion in crash data sets. Next, the background section reviews the current standardized methods to collect fatality data. Then, using the crash data from selected communities, we highlight the limitations of current crash data collection processes. Finally, we conclude by offering recommendations on how the crash data collection could be standardized and improved. This section contributes to the literature on bicycle data safety needs and shows that a unified bicycle crash collection method is needed to advance research-driven solutions for combatting bicycle crashes around the country.

### **3.1 BACKGROUND**

Bicycle crashes have been studied by many researchers around the world and increasingly in the United States. Many of these studies helped individual areas or regions identify and rectify issues within their bicycle infrastructure. A number of variables within the crash data sets were found to be important predictors of motorist-cyclist crashes in literature. Table 3.1 shows examples of previous research articles that found these same variables as significant.

**Table 3.1: Motorist-cyclist Crash Data Fields from Literature**

Field Type	Field Variable	Studies which found this field to be associated with motorist-cyclist crashes
Environmental	Lighting Weather Intersection Type Speed Limit Traffic Control Device	Zangenehpour et al., 2015 Yan et al., 2011 Klop et al., 1999 Allen-Munley et al., 2004 Strauss et al., 2015 Reynolds et al., 2009 Turner et al., 2011
Crash Specific	Collision Type Severity	Wang et al., 2015 Klop et al., 1999 Allen-Munley et al., 2004
Time	Year Month Day Hour	Wang et al., 2015

The variables shown in Table 3.1 are important predictors of motorist-cyclist crashes (though this is not an exhaustive list), and thus, crash data sets that include these variables are more likely to lead to more accurate prediction of bicycle crashes and identification of bicycle safety-specific countermeasures. For this reason, these variables will be specifically identified in the data sets included in this study: year, month, day, hour, intersection type, collision type, lighting, weather, speed limit, traffic control device, and collision severity.

### 3.1.1 Motorist-cyclist Crash Data Standards and Tools

Fatal crashes across the U.S. involving all modes are collected in a standardized manner and reported at the national level using the Fatality Analysis Reporting System (FARS). FARS is a national database run by the National Highway Traffic Safety Administration (NHTSA), but it only collects fatal incidents, which account for a small percentage of overall bicycle crashes (NHTSA, 2015). For example, fatal crashes are only 2percent of the crashes studied in this paper. Additionally, FARS is still heavily geared towards automobile crashes, thus making the recording of bicycle crashes problematic. However, no similar U.S. database yet exists for non-fatal crashes either for bicycles or automobiles, though many of the problems highlighted in this paper are bicycle-specific.

While crash report forms at the state level contain information on where, when, and whom the crash impacts, they tend to exclude any information regarding the sequence of events preceding the crash. This hinders the development of countermeasures due to insufficient details about the crashes (Harkey et al., 2006). Hence, the Pedestrian and Bicycle Crash Analysis Tool (PBCAT) was developed by the NHTSA as a tool for “typing” bicycle and pedestrian crashes. The objective of the crash typology is to define each crash type by a specific sequence of events, precipitating actions, predisposing factors and characteristic populations, and/or locations that can be targeted for interventions (Harkey et al., 2006). Table 3.2 shows the motorist-cyclist crash data types specified by PBCAT. These crash data types are important variables that help

researchers and practitioners understand the contributing causes of these crashes. For example, the city of Boulder reports that the most common motorist-bicyclist serious injury collision types (64percent) involved a motorist driving out (34percent) or turning (30percent) into the path of a cyclist, and Boulder is now using this information to choose appropriate countermeasures (City of Boulder Transportation Division, 2016).

**Table 3.2: Example PBCAT Motorist-Cyclist Crash Data Types (Harkey, et. al, 2006)**

General Crash Types	Specific Crash Types
Turning Error	Motorist Turning Error (Left Turn; Right Turn; Other) Bicyclist Turning Error (Left Turn; Right Turn; Other)
Loss of Control	Bicyclist Lost Control (Mechanical Problems; Oversteering, Improper Braking, Speed; Alcohol / Drug Impairment; Surface Conditions; Other / Unknown) Motorist Lost Control (Mechanical Problems; Oversteering, Improper Braking, Speed; Alcohol / Drug Impairment; Surface Conditions; Other / Unknown)
Failure to Yield – Sign-Controlled Intersection	Motorist Drive Out - Sign-Controlled Intersection Bicyclist Ride Out - Sign-Controlled Intersection Motorist Drive Through - Sign-Controlled Intersection Bicyclist Ride Through - Sign-Controlled Intersection Multiple Threat - Sign-Controlled Intersection Sign-Controlled Intersection - Other / Unknown
Failure to Yield – Signalized Intersection	Motorist Drive Out (Right Turn on Red; Signalized Intersection) Bicyclist Ride Out - Signalized Intersection Motorist Drive Through - Signalized Intersection Bicyclist Ride Through - Signalized Intersection Bicyclist Failed to Clear (Trapped; Multiple Threat; Unknown) Signalized Intersection - Other / Unknown
Crossing Paths	Crossing Paths (Uncontrolled Intersection; Intersection - Other / Unknown)
Motorist Turn / Merge	Motorist Left Turn (Same Direction; Opposite Direction) Motorist Right Turn (Same Direction; Opposite Direction) Motorist Right Turn on Red (Same Direction; Opposite Direction) Motorist Turn / Merge - Other / Unknown; Bicyclist Left Turn - Same Direction
Bicyclist Turn / Merge	Bicyclist Left Turn - Opposite Direction Bicyclist Right Turn (Same Direction; Opposite Direction) Bicyclist Ride Out - Parallel Path
Parking / Bus-Related	Motorist Drive In / Out - Parking; Bus / Delivery Vehicle Pullover
Motorist Overtaking Bicyclist	Motorist Overtaking (Undetected Bicyclist; Misjudged Space; Bicyclist Swerved; Other / Unknown)
Bicyclist Overtaking Motorist	Passing on Right; Passing on Left; Parked Vehicle; Extended Door; Other / Unknown
Head-On	Head-On (Bicyclist; Motorist; Unknown)
Parallel Paths	Parallel Paths - Other / Unknown
Failed to Yield–Midblock	Bicyclist Ride Out (Residential Driveway; Commercial Driveway / Alley; Other Midblock; Midblock - Unknown) Motorist Drive Out (Residential Driveway; Commercial Driveway / Alley; Other Midblock; Midblock - Unknown) Multiple Threat - Midblock; Crossing Paths - Midblock - Other / Unknown
Other	Bicycle Only; Motorist Intentionally Caused; Bicyclist Intentionally Caused; Backing Vehicle; Play Vehicle-Related; Unusual Circumstances; Non-Roadway; Unknown Approach Paths; Unknown Location

While FARS and PBCAT have standardized some aspects of the crash data process, they have their limitations to being adopted by jurisdictions for non-fatal crashes. FARS is for fatal crashes and therefore may need some adaptation to include non-fatal crashes. PBCAT in its present state is designed to analyze crash data after it has already been recorded in various regionalized systems and would need to be integrated into on-the-scene crash reports to work seamlessly. Thus, no standards currently exist for recording non-fatal cyclist crashes, leading to many different approaches in collecting this data. The following sections review the non-fatal crash data obtained from the eight communities and describe limitations within the data.

## **3.2 STUDY COMMUNITIES**

In this study, eight different city and county crash data were assessed. Table 3.3 shows the population estimates for each of these locations along with fatal and non-fatal crashes. These data were acquired through multiple government agencies of varying hierarchy (local and state) within their respective regions. Arlington County, Bellingham, Minneapolis/St. Paul, Philadelphia, and Portland's data were acquired through their respective state transportation departments. Boulder and Denver's data were acquired from the Denver Regional Council of Governments (DRCOG). San Diego County's data was acquired through the California Highway Patrol's Statewide Integrated Traffic Records System (SWITRS).

These locations varied in terms of population, with Bellingham having the lowest population of 80,867 residents, and San Diego County with the highest population of 3,095,013 residents. Table 3.3 shows that Philadelphia, San Diego, Minneapolis/St-Paul, and Portland, on average, had significantly higher numbers of non-fatal crashes per year compared to Arlington County, Bellingham, and Boulder. One contributing factor could be greater numbers of residents in these communities. While crashes per year per population are reported in Table 3.3, this may not be a good metric for comparison due to differences in exposure (some communities have more driving or more cycling than others) and differences in crash reporting (some communities have higher rates of motorist-bicyclist crash reporting than others).

The fact that the number of fatal crashes per year is less than four for all of the communities studied demonstrates fatalities alone are not sufficient for safety analysis at the community, area, corridor, or facility level. This is why non-fatal crashes are important for understanding bicycle safety at the facility level.



**Table 3.3: Crash Data for Each Community**

Community	2010 Population (U.S. Census Bureau, 2010)	Years (Non-Fatal)	Non-fatal Bicycle-related Crashes			Bicyclist Fatalities per Year
			Total Non-Fatal	Average per Year	Crashes per Year per 100,000 population	
Arlington	207,627	2010-2015	239	40	19	0.2
Bellingham	80,867	2001-2015	499	33	41	0.2
Boulder	97,468	2006-2012	893	99	101	0.6
Denver	600,025	2003-2009	1480	211	35	1.3
Minneapolis/St. Paul	667,646	2005-2014	3784	378	57	1.1
Philadelphia	1,526,006	2010-2014	2675	535	35	3.3
Portland	583,789	2009-2012	1271	318	54	2.2
San Diego County	3,095,313	2001-2014	6493	464	15	3.8

The crash data obtained from these communities was processed for future safety analyses. As a part of this process, several issues were encountered with the data and these are described in detail below.

### 3.2.1 Observed Issues with Crash Data

An examination of crash data across the eight locations revealed the following issues: inconsistent data formats, differences in crash data fields between communities, and differences in data within communities from year to year. Other issues included under-reporting of crashes and the use of motor vehicle-specific crash collection forms. Each of these issues are discussed separately below.

### 3.2.2 Inconsistent Data Formats

Two differences were observed at the onset of the crash data retrieval process: a) data was made available via different avenues (e.g., email, password-protected sites, nonpassword-protected sites, and FTP sites); and b) data was available in varying electronic formats (e.g., Microsoft Excel files with and without coordinates, Microsoft Access database, and GIS shapefiles). Arlington County, Bellingham, and San Diego County’s data came in an Excel file format. Arlington County’s data came with coordinates, but Bellingham and San Diego County’s data required geocoding. Boulder, Denver, Minneapolis/St. Paul, and Portland’s data were all made available as GIS shapefiles. Boulder and Denver’s GIS shapefiles were made available off the DRCOG website. Minneapolis/St. Paul’s GIS shapefile was emailed to us directly from MNDOT. Portland’s GIS shapefile was available via a public FTP site. Philadelphia’s data were sent to us via an encrypted email, which required a Commonwealth of Pennsylvania account. The data were found in an FTP site in Microsoft Access format, of which about half required further geocoding.

### 3.2.3 Differences in Data Fields Between Communities

Table 3.4 shows the crash data fields that are collected across our study communities. While there are some consistencies between them such as the year and month of the crash, every other type of crash data field has varying degrees of consistency.

Aside from Philadelphia, every city had missing fields that would be of interest to the road safety research community. The most common type of missing field was the traffic control device, with five of the eight regions omitting it entirely. It is also worth pointing out that missing fields might not always be missing entirely from a data set. Much of our data came from publicly available sources and, as such, some fields may have simply been purposely omitted from those publicly available sources. For example, the city of Boulder uses PBCAT to crash type their bicycle and pedestrian crashes, but such data are not provided in the data set we received from their metropolitan planning organization (MPO).

Even if two or more communities recorded the same basic data fields, there were often discrepancies regarding how each field was collected. For example, Portland and Philadelphia record their collision severity differently with a different amount of fields for each. Portland records its crash severity as either a fatal crash, non-fatal injury crash, or PDO crash while Philadelphia recorded it as: not injured, killed, major injury, moderate injury, minor injury, injury/unknown severity, and unknown if injured. Crash severity was entirely dependent on the scale used by the various regional reporting agencies and was equally dependent on the use of a data dictionary to figure out how a comparison could be made. This leads further to the question of how an officer might differentiate between a minor, moderate, and major injury. The thresholds are nebulous and different officers might classify the same crash injury differently.

Intersection type had the most discrepancies among data sets, with Minneapolis/St. Paul recording 17 different intersection types. Meanwhile, San Diego County recorded their intersection type as either yes, no, or unknown. Comparing such data across locations would require paring down each other region's intersection type data to a simple "yes, no, or unknown." This would result in the loss of potentially useful information since four-way intersections have been found by others to be more associated with motorist-bicyclist crashes than three-way intersections (Hauksson, 2014). San Diego, for its part, is unusual in that most of its crashes are recorded as non-intersection related. Additionally, since San Diego's data is county data from a statewide organization, their "yes, no, or unknown" issue might already have been caused by the paring down of inconsistent city-level data. Almost every other field required a similar amount of manipulation in order to accurately compare the data fields from region to region. Lighting conditions contained differences between whether sunrise and sunset were the same category or different, or whether there was mention of streetlights being on or off for dark periods of the day. Weather data contained discrepancies regarding whether cloudy conditions were recorded or not, or if snow, rain, and sleet were recorded with each other or as separate weather incidents. Figure 3.1 shows how the study communities compare with the "lowest common denominator" data. Note that cities that did not have data presented are not reflected in the graphs.

**Table 3.4: Crash Data Variables Collected in Study Areas**

Community	Time Fields				Crash Fields		Environment Fields				
	Year	Month	Day	Hour	Crash Severity	Collision Type	Lighting	Weather	Speed Limit	Traffic Control	Intersection
Arlington	X	X	X	X		X			X		
Bellingham	X	X		X	X	X	X				X
Boulder*	X	X		X	X		X	X			X
Denver	X	X	X	X	X			X	X		X
Minneapolis/St. Paul	X	X	X		X	X	X		X	X	X
Philadelphia	X	X	X	X	X	X	X	X	X	X	X
Portland	X	X	X	X	X	X	X	X		X	X
San Diego County	X	X	X	X	X	X	X	X			X

\* Boulder collects crash type data using PBCAT but it is not available publicly online.

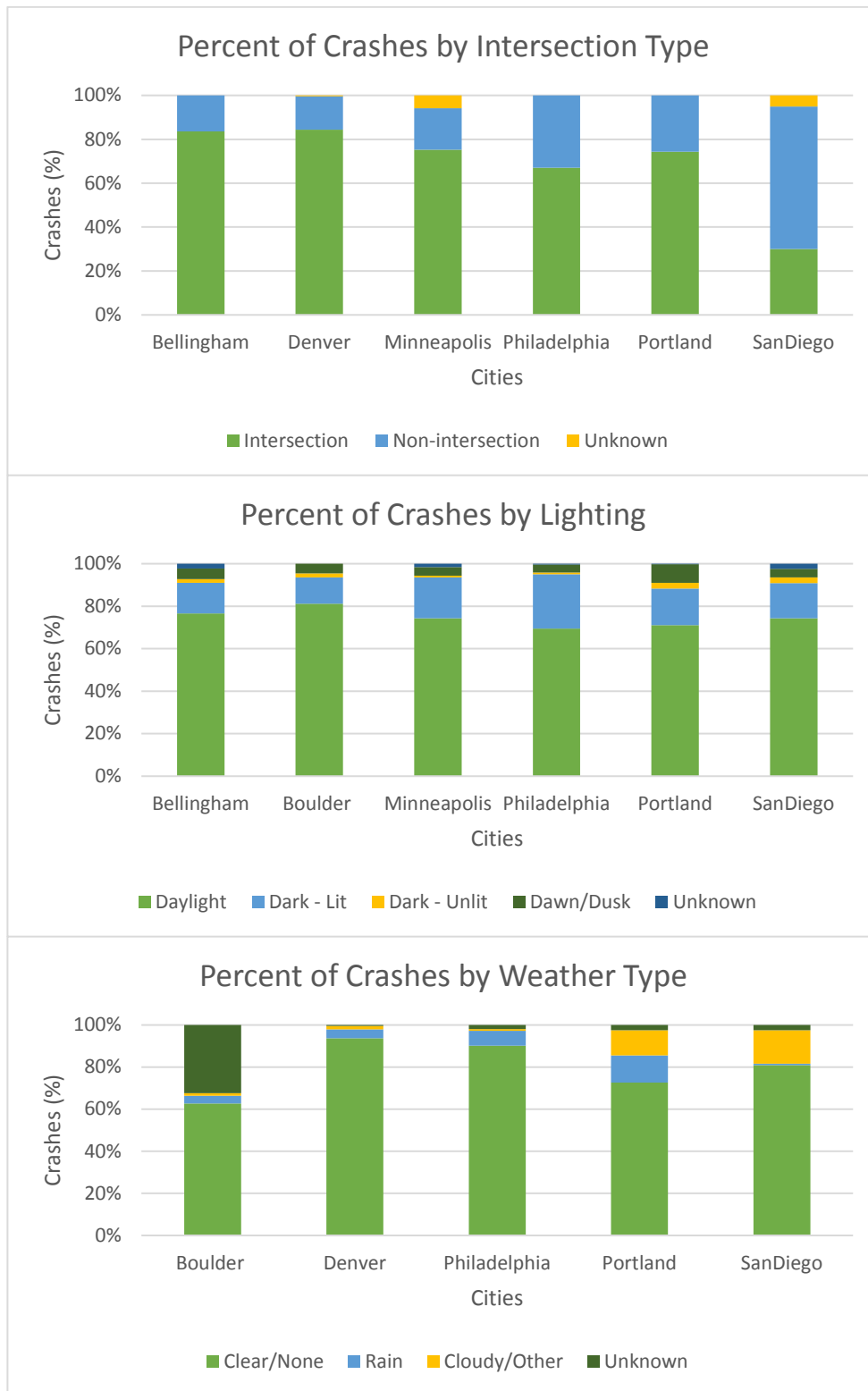


Figure 3.1: Percent of Crashes with Intersection Type, Lighting, and Weather

The only data fields that required no manipulation in order to draw comparisons were time-based. Year of crash, month of year, day of week, and hour of day were all recorded in similar, standardized formats allowing for conclusions to be drawn fairly easily between each region. Additionally, speed limit, when available, also proved to be fairly standardized across each city as, even if a city did not have a specific posted speed limit with crashes, comparisons were still able to be drawn overall.

Data dictionaries that provide information to the user about the variables in the crash data set also varied between communities. San Diego County had a data dictionary that defined various numbered scales, but not what the actual definitions were for specific named values. Minneapolis-St. Paul's data dictionary was a single page Word document with a very basic description to go with value fields. Portland's data dictionary was a 191-page document that gave specific definitions for each field as well as instructions on how the data should be read or interpreted. The range in data dictionary thoroughness from detailed to minimalistic poses a challenge to analyze and compare the data.

### **3.2.4 Differences in Data Coding Within Communities**

One other interesting problem occurred in three of the eight study communities. Discrepancies were not only found in how and what was collected, and how the data was coded, but also in an individual community's year-over-year data. This created problems in not only comparing data between different regions, but also in comparing data within the same region from one year to the next.

Three communities suffered from this specific data problem: Boulder, Denver, and Portland. In Boulder, each year from 2004-2012 was recorded in a different manner. Denver's data was recorded in a different manner for each year from 2004-2009 and then from 2010-2015 in a new, albeit consistent, manner. Portland's data was recorded in a different manner for 2007-2008 and then again in each year from 2009-2012.

Working with these three communities' data sets required significant manipulation of data from year to year in order to perform basic comparisons. This might lead to fields being left out due to a lack of coordination over time. This also led to an increase in confusion and manipulation of data when comparing these three specific communities to the other five regions.

It is worth mentioning that improvement in the recording process is something that does happen over time and that iterations are to be expected. However, data recording iterations, which do not preserve consistency with previous fields, make worthwhile safety research increasingly challenging. A good example of iterations in data recording over time with few barriers to comparison is the FARS database, which includes former codified data fields with new ones in a single document allowing researchers to easily change or use older data.

### **3.2.5 Underreporting**

One issue that could potentially impact many safety analyses is the underreporting of crashes. Many states have laws indicating that a crash may be unreported if it is a property damage only (PDO) crash with no injuries or fatalities and the damage to property is below a certain monetary threshold. Additionally, underreporting could occur due to the reluctance of involved parties to interact with law enforcement or be hospitalized. Some communities may have lower underreporting rates than others. Communities that work the hardest to overcome underreporting may appear to have higher numbers of motorist-cyclist crashes while, ironically, they may actually have fewer crashes than other communities but are simply better at reporting them.

Studies have found that pedestrian and bicycle crashes are underreported (Elvik and Mysen, 1999; Medury et al., 2017) and especially when there are no injuries, little property damage and only one party is involved (Stutts and Hunter, 1998; Sciortino et al., 2005; Loo and Tsui, 2007). Integrating hospital data with the crash reports would help provide a more complete picture of motorist-cyclist crashes. However, due to privacy laws governing medical records in the U.S., such coordination could be challenging. Sweden, for example, integrates the information from the police and hospital data into a single database – Swedish Traffic Data Acquisition (Strada). The inclusion of the hospital data addresses the issue of underreporting of crashes.

### **3.2.6 Motor-vehicle-specific Reporting Forms**

Another issue relevant to motorist-cyclist crashes is that the automobile-oriented data form precludes collection of bicycle-specific crash types. Figure 3.2 shows a typical crash form from Washington state. Every region that we examined used some variation of an automobile-oriented recording database in order to record their bicycle crash incidents. This created problems for a number of the regions as the automobile-oriented recording database simply did not match up with the specifics of a bicycle crash. In fact, a study done by Berkow et al. in 2017 reviewed all state collision reporting forms in the U.S. to determine how each form was capturing locational information about cyclist-involved collisions in report fields. The study revealed that many state collision reporting forms do not provide the level of detail for bicycle-involved collision typing, and that there are several ambiguities in collision description. For example, San Diego County recorded its collision type data as such: head-on, sideswipe, rear end, broadside, hit object, overturned, vehicle/pedestrian, other, and not stated. Of San Diego County's recorded bicycle crashes, 291 were marked "overturned" but there was no mention of whether this was for the bicycle or if there were 291 overturned vehicles hitting cyclists. This makes figuring out what actually occurred in the crash more challenging, and these recording practices seems to be occurring across the country.

Dooring is a type of motorist-cyclist crash in which the bicyclist is struck by a car door opening into the bicycle lane. This often occurs in urban areas, where bicycle lanes are placed immediately adjacent to parked cars. An example of how this crash type has turned fatal is when a cyclist who has encountered an open door is thrown or diverted into the motor vehicle travel lane and struck by a moving vehicle. While cities like Chicago have started recording data on this specific crash type, these data were not recorded in the data sets highlighted in this paper. This lack of crash typing could hamper safer and efficient design of separated bicycle

infrastructure. Though the city of Boulder does use PBCAT, which does record this type of data, this crash typing was not included in the DRCOG data set and thus is missing from the most accessible, publicly available data on crashes. Table 3.5 shows a sample of how useful this type of crash typing can be in identifying crash scenarios that can be remediated by engineering solutions.

**Table 3.5: PBCAT Crash Types in Boulder**

<i>Top 5 PBCAT crash types in Boulder 2008-2014</i>	
<b>Percent</b>	<b>Crash Group Description</b>
<b>12%</b>	Motorist Failed to Yield - Sign-Controlled Intersection
<b>8%</b>	Motorist Failed to Yield - Signalized Intersection
<b>12%</b>	Motorist Left Turn / Merge
<b>12%</b>	Motorist Right Turn / Merge
<b>7%</b>	Motorist Failed to Yield - Midblock
<i>Source data provided by City of Boulder</i>	

Data quality is also a key challenge that needs to be ensured with cyclist crash data at all levels. Law enforcement officers have many duties, need to prioritize emergency calls, and if their law enforcement department is not using the data collected, may not see the direct impact of quality data collection. Some officers may not be as familiar with bicycle-specific issues as others, which may lead to bias. Another issue is that bicyclists as vulnerable road users are more likely to be incapacitated in a collision with a car and thus less likely to be able to report their account of the crash to law enforcement officials.



STATE OF WASHINGTON  
POLICE TRAFFIC  
COLLISION REPORT



1591971

REPORT NO. 3569476

- 1
- 2
- 3
- 4
- 4a
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26

INTERSTATE     CITY STREET     FIRE RESULTED  
 STATE ROUTE     OTHER     STOLEN VEHICLE  
 COUNTY RD     PRIVATE WAY     HIT & RUN INVOLVED

CASE # \_\_\_\_\_

LOCAL AGENCY CODING \_\_\_\_\_

TOTAL # OF UNITS \_\_\_\_\_ OBJECT STRUCK \_\_\_\_\_

TRIBAL RESERVATION \_\_\_\_\_  
 DATE OF COLLISION: M M D D Y Y Y Y TIME (2400) COUNTY # MILES CITY #  
 N E IN S W OF

ON (PRIMARY TRAFFIC WAY) INTERSECTION  NON-INTERSECTION   
 BLOCK NO. \_\_\_\_\_ MILE POST \_\_\_\_\_

DISTANCE \_\_\_\_\_ OF (REFERENCE OR CROSS STREET)  
 FEET \_\_\_\_\_ MILES  N  E  S  W

**UNIT 01** MOTOR VEHICLE  PEDAL-CYCLE  DAMAGE THRESHOLD MET YES  NO  PHONE \_\_\_\_\_

LAST NAME \_\_\_\_\_ FIRST NAME \_\_\_\_\_ MIDDLE INITIAL \_\_\_\_\_

STREET NEW ADDRESS \_\_\_\_\_

CITY \_\_\_\_\_ ST \_\_\_\_\_ ZIP \_\_\_\_\_

CDL \_\_\_\_\_ ENDORSEMENTS \_\_\_\_\_ RESTRICTIONS \_\_\_\_\_

DRIVER'S LICENSE # \_\_\_\_\_ STATE \_\_\_\_\_ SEX \_\_\_\_\_ D.O.B. MMDDYYYY \_\_\_\_\_

ON DUTY  STATUS \_\_\_\_\_ AIRBAG \_\_\_\_\_ RESTR. \_\_\_\_\_ EJECT \_\_\_\_\_ HELMET USE \_\_\_\_\_ INJURY CLASS \_\_\_\_\_ NATURE OF INJURIES \_\_\_\_\_

LICENSE PLATE # \_\_\_\_\_ STATE \_\_\_\_\_ VIN# \_\_\_\_\_

TRAILER PLATE # \_\_\_\_\_ STATE \_\_\_\_\_ TRAILER PLATE # \_\_\_\_\_ STATE \_\_\_\_\_

VEH. YEAR \_\_\_\_\_ MAKE \_\_\_\_\_ MODEL \_\_\_\_\_ STYLE \_\_\_\_\_ VEHICLE TOWED YES  NO  TOWED BY \_\_\_\_\_ GOVT. VEHICLE YES  NO

REGISTERED OWNER INFO. LIABILITY INSURANCE IN EFFECT  INSURANCE CO & POLICY # \_\_\_\_\_

VEHICLE LOCALLY PARKING YES  NO  CITATION # \_\_\_\_\_ CHARGE \_\_\_\_\_



**UNIT 02** MOTOR VEHICLE  PEDAL-CYCLE  PEDESTRIAN  PROPERTY OWNER  DAMAGE THRESHOLD MET YES  NO  PHONE \_\_\_\_\_

LAST NAME \_\_\_\_\_ FIRST NAME \_\_\_\_\_ MIDDLE INITIAL \_\_\_\_\_

STREET NEW ADDRESS \_\_\_\_\_

CITY \_\_\_\_\_ ST \_\_\_\_\_ ZIP \_\_\_\_\_

CDL \_\_\_\_\_ ENDORSEMENTS \_\_\_\_\_ RESTRICTIONS \_\_\_\_\_

DRIVER'S LICENSE # \_\_\_\_\_ STATE \_\_\_\_\_ SEX \_\_\_\_\_ D.O.B. MMDDYYYY \_\_\_\_\_

ON DUTY  STATUS \_\_\_\_\_ AIRBAG \_\_\_\_\_ RESTR. \_\_\_\_\_ EJECT \_\_\_\_\_ HELMET USE \_\_\_\_\_ INJURY CLASS \_\_\_\_\_ NATURE OF INJURIES \_\_\_\_\_

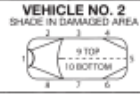
LICENSE PLATE # \_\_\_\_\_ STATE \_\_\_\_\_ VIN# \_\_\_\_\_

TRAILER PLATE # \_\_\_\_\_ STATE \_\_\_\_\_ TRAILER PLATE # \_\_\_\_\_ STATE \_\_\_\_\_

VEH. YEAR \_\_\_\_\_ MAKE \_\_\_\_\_ MODEL \_\_\_\_\_ STYLE \_\_\_\_\_ VEHICLE TOWED YES  NO  TOWED BY \_\_\_\_\_ GOVT. VEHICLE YES  NO

REGISTERED OWNER INFO. LIABILITY INSURANCE IN EFFECT  INSURANCE CO & POLICY # \_\_\_\_\_

VEHICLE LOCALLY PARKING YES  NO  CITATION # \_\_\_\_\_ CHARGE \_\_\_\_\_



OFFICER'S NAME (PRINT) \_\_\_\_\_ BADGE OR ID # \_\_\_\_\_ AGENCY \_\_\_\_\_

- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42

DISTRIBUTION: ORIGINAL - WASHINGTON STATE PATROL, P.O. BOX 42628, OLYMPIA WA 98504-2628;  
 Figure 3.2: Police Traffic Collision Report Form from Washington State



### 3.3 SUMMARY

In order to improve bicycle safety research, the U.S. needs consistent and quality data being collected throughout its major cities and metropolitan areas. As shown in this section, bicyclist crash data is currently being collected and recorded in an inconsistent manner, often with missing information and with data recording methods used primarily for automobiles. Between agencies, there are inconsistencies in the data dictionaries and their respective definitions as well as the amount of data fields being recorded by each agency, particularly along small-large region lines. Bicycle crashes may also be underreported based on the severity of the crash and whether there has been property damage above a specific threshold or if the police were ever called in the first place. Finally, there is a distinct lack of bicycle-specific crash types represented in the data, of particular note being “dooring” crashes. To remedy these issues, we propose the following recommendations:

Standard nationwide formats for motorist-cyclist crash data in the U.S. (such as PBCAT) should be used. The current method of using different state-by-state or region-by-region recording methods paints an inconsistent picture and makes nationwide research on the subject of bicycle safety from motor-vehicle crashes impractical. This is particularly true for communities that are already working with limited data of their own due to lower volumes of cyclists than automobile drivers. Standardization could be further informed by PBCAT as a potential reference system, but some basics fields that should be included are listed below:

- Intersection Type – A method for determining where specifically the bicyclist is with relation to the intersection, including: type of intersection (four-way, t-intersection, driveway, roundabout, etc.) and method of approach to the intersection by way of facility.
- Crash Type – Bicycle-specific crash types with definitions that inform whether the bicyclist or the motorist made the error and what the specific error was. For example:
  - Motorist Right Turn – Same Direction
  - Bicyclist Ride Through – Signalized Intersection
  - Bicyclist Overtaking Vehicle – Extended Door
- Crash Severity – A standardized crash severity scale that would allow for quick comparisons to be made between communities and would have a set standard for PDO crashes. This could include coordination between hospital records and crash records for a more consistent reporting of severity when an injury is involved.

Not only could PBCAT be used as a reference system, it could also be directly embedded within a police officer’s recording database. Police officers are already increasingly moving towards electronic interface recording systems (often internet enabled) that help with data consistency on a local level. Integrating these electronic systems with PBCAT, however, could prove to be an effective method of standardizing motorist-bicyclist crashes.

In addition to creating a national standard and database for non-fatal crashes, including motorist-cyclists crashes, integrating crash report databases with hospital records may yield additional data. It has been documented that motorist-bicycle crashes are underreported even when an injury is involved (Elvik and Mysen, 1999; Medury et al., 2017). The sparse number of reported

crashes would benefit from the addition of these unreported crashes present in hospital and emergency responder data.

Future research on exposure and volume data are needed to more accurately illuminate the impact that a lack of data is having on bicycle crash safety research. Exposure and volume data have their own implications with regards to bicycle safety, but data regarding exposure and volume are also recorded in an inconsistent manner across municipalities. Future research would combine this section's analysis with municipal exposure and volume data to create a clearer picture of the overall bicycle safety data needed to accurately research bicycle crashes in this country. Similarly, data on when bicycle facilities are installed and where they exist are also important to studying bicycle safety. There is currently no national standardized data set of bicycle infrastructure.

Without improvements in data consistency and quality, studies of bicycle safety will continue to be inadequate. Additionally, without accurate research studies, bicycle safety facilities will never be ranked as either a four- or five-star CMF, which will inhibit the use of funds for bicycle-related projects. With bicycle crash data already comprising a small segment of a city's crash data, communities will need accurate studies from a national pool of data if they wish to expand their existing facilities and encourage more ridership.

## **4.0 DATA AND METHODS**

An important insight in the pursuit of improved bicycling safety is a better understanding of what factors are associated with motorist-bicyclist collisions. As mentioned previously, safety performance functions (SPFs) can be developed for cyclists to account for effects such as “safety in numbers” and its relationship to risk.

Existing bicycle SPFs primarily focus on intersections, which is where the majority of bicycle crashes occur (Nordback et al., 2014). However, no bicycle SPFs in the U.S. utilizing bicycle volume data exist for segments. This section adopts methods from the Highway Safety Manual (HSM) used for motor vehicle SPFs in order to develop bicycle-specific SPFs for roadway segments in Boulder, CO (AASHTO, 2010). In order to account for the “safety in numbers” effect, both vehicle volume and bicycle volume data are used in the model. This is a preliminary effort at developing a bicycle-specific SPF for segments in the U.S. that utilizes bicycle volumes, is an important first step towards further understanding bicyclist safety, and may inform future versions of the HSM.

This chapter presents a description of the data and methods used in this study to create bicycle-specific SPFs for road segments in Boulder.

### **4.1 SITE LOCATION**

Boulder was chosen for this study because of its data availability. It also has a high commute-to-work bicycle mode share (10.8percent) compared to other U.S. cities (McKenzie, 2014). Additionally, Boulder has an extensive counting program, dating back to the early 2000s, comprised of both manual short-term counts and permanent counters, thus allowing for safety analyses. Annual average daily bicyclists (AADB) estimates were developed by expanding the short-duration counts using factors developed from the permanent counters.

Although the majority of motorist-bicyclist crashes (65percent) in Boulder occur at intersections, these have already been studied (Nordback et al., 2014). The 32percent that occur along segments have not received much attention. Therefore, the objective of this study is to model the non-intersection crashes as a function of geometric and traffic variables.

Various data were gathered for the safety analysis. The data types assembled included the crash data, bicycle, pedestrian and auto volumes, as well as supplemental data such as land use, population and employment density, educational attainment, and bicycle facilities. We describe each of these in more detail below.

#### **4.1.1 Crash Data**

Crash data for Boulder were gathered for 2006-2013 from the Denver Regional Council of Governments (DRCOG). The DRCOG dataset includes 717 non-fatal, motorist-bicyclist crashes. Of these, 232 (32percent) were non-intersection and non-intersection-related crashes. Of those,

174 occurred on the 346 road segments for which the bicycle, pedestrian and motor vehicle volumes and facility type variables are known (i.e., 0.5 motorist-bicyclist crashes per segment on average). The crashes analyzed are all non-fatal crashes including severe and PDO crashes. From 2006-2013, three bicyclist fatalities occurred within Boulder.

Many non-intersection crashes without latitude and longitude were still initially geocoded to an intersection. This anomaly was corrected by re-mapping crashes to road segments using the direction noted in the data set for the primary vehicle causing the crash (Pendleton, 2016). Of the 174 crashes studied, 2006 and 2013 were the two years with the most crashes. The highest proportion of these crashes occurred in September (17 percent), followed by August (14 percent) and April (11 percent). Since this study only analyzed crashes that were not intersection related, the location of these crashes was further analyzed. Of all crashes, 53 percent were driveway-access related, 43 percent were non-intersection related, and 4 percent were alley-related crashes. Most of the crashes also occurred on straight, level grade (79 percent), on dry pavement (73 percent), during daylight (82 percent), and during no adverse weather conditions (66 percent). Table 4.1 shows descriptive statistics associated with these crashes.

**Table 4.1: Summary Statistics for the Non-fatal, Motorist-bicyclist Crashes Studied (only crashes on the 174 segments with traffic volume data are included)**

Descriptor	Crashes
<b>Road Description</b>	
Driveway-Access Related	53%
Non-Intersection	43%
Alley Related	4%
<b>Contour</b>	
Straight-on Level	79%
Straight-on Grade	18%
Curve-on Grade	2%
Unknown	1%
<b>Condition</b>	
Dry	73%
Wet	24%
Muddy	2%
Snowy	0.5%
Unknown	0.5%
<b>Lighting</b>	
Daylight	82%
Dawn or Dusk	6%
Dark – Lighted	10%
Dark – Unlighted	1%
Unknown	1%
<b>Weather</b>	
None	66%
Snow/Sleet/Hail	3%
Wind	1%
Unknown	30%

#### 4.1.2 AADB

The short-duration bicycle counts were obtained from the city of Boulder’s turning movement counts that were conducted during the AM, mid-day, and PM peak periods. First, the morning-midday index (AMI) was computed using the following formula (Miranda-Moreno et al., 2013).

$$AMI = \frac{V_{am}}{V_{mid}}$$

Where:

$V_{am}$  = volume between 7:45 – 8:45 am and

$V_{mid}$  = volume between 11:45 am – 12:45 pm

The AMI is used to determine if traffic patterns are utilitarian or recreational. Higher AMI values indicate commute patterns, and lower values indicate recreational patterns. To expand the short-duration counts into AADB values, three continuous count sites within Boulder were used. The AMI was also computed for these sites in order to classify them into commute and recreational sites to develop the appropriate factors. Only one site (US-36) exhibited non-commute pattern, whereas the other two sites (13<sup>th</sup> Street and Folsom Street) had commute patterns. Factors for each day of the short-duration counts were computed by dividing the sum of the short-duration counts taken on that day by the AADB computed using the continuous counts. These factors were then used to expand the short-duration counts to AADB, as shown below.

$$AADB = c_{kp} * F_{df}$$

Where:

$c_{kp}$  = the known bicycle count for sum of three peak hours 7:45-8:45 a.m., 11:45 a.m.-12:45 p.m., and 4:45-5:45 p.m. (8,12,5) on a Tuesday, Wednesday, or Thursday (TWorR)

$F_{df}$  = adjustment factor for bicycle traffic for a given date,  $d$ , for a factor group,  $f$ .

$$\text{Adjusted AADB} = \text{AADB (2013-Y)} * G$$

Where:

$G$  = the growth factor per year, assuming linear growth (for Boulder,  $G$  is negative for 2012-2014:  $G = -16.3$  for commute pattern sites and  $-36.4$  for non-commute pattern sites).

$Y$  = the year in which the manual count was collected

For modeling purposes, AADB and Adjusted AADB were divided by 100 in order to be similar in magnitude to other variables modeled. For the rare situation when volumes from the three-hour counts were zero, a value of 0.5 was used for AADB and Adjusted AADB for the purposes of modeling.

Figure 4.1 shows the bicycle volumes and crashes along segments in Boulder. The figure illustrates that crash rates are higher along segments with lower bicycle volumes.

### 4.1.3 AADT

The turning movement counts for the AM, midday, and PM peak periods were adjusted to daily counts by summing the three peak-hour counts and dividing the sum by 0.225. These daily estimates were then multiplied by the daily and monthly factors and annual growth factor to estimate AADT. For the rare situations when volumes from the three-hour counts were zero, a value of 0.5 was used for AADT for the purposes of modeling.

$$AADT = c_{kp} * D_{pyf} * M_{pyf} / H + (2013-Y)*G$$

Where:

AADT = the estimated annual average daily motorized traffic

$c_{kp}$  = the known motor vehicle count for sum of three peak hours (8,12,5) on TWorR

$D_{pyf}$  = the daily motor vehicle factor for a given month in a given year, y, for a factor group, f, for TWorR

$M_{pyf}$  = the monthly motor vehicle factor for a given month in a given year, Y, for a factor group f.

H = 0.225 = hourly adjustment factor from Boulder (2010)

G = the growth factor per year, assuming linear growth

Y = the year in which the manual count was collected

Figure 4.2 shows the map of automobile volumes in Boulder. For the purposes of modeling, the AADT was divided by 1,000 so that it is of similar magnitude to other variables modeled.

#### 4.1.4 Pedestrian Volume

As with AADT and AADB, pedestrian volume was determined by turning movement counts taken on a given day during three peak hours. Unlike the AADT and AADB metrics previously described, the pedestrian volumes used were simply the sum of the three peak-hour counts. No adjustment was applied because no continuous pedestrian count data were available in the city during the years in which the turning movement counts were collected. For the rare situation when volumes from the three-hour counts were zero, a value of 0.5 was used for pedestrian volume for the purposes of modeling.

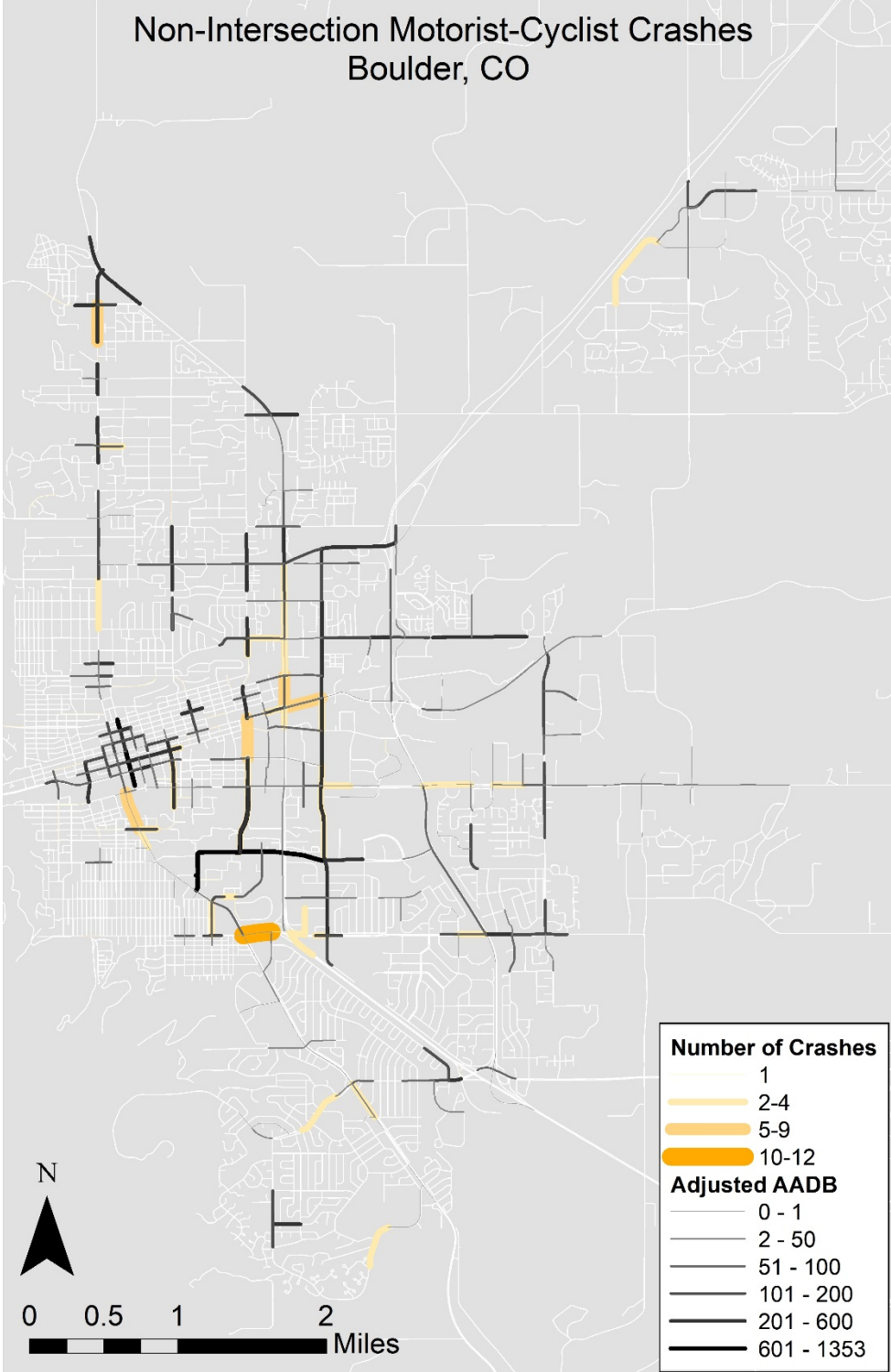


Figure 4.1: Bicycle Volume and Non-fatal, Motorist-bicyclist Crash Map for Boulder



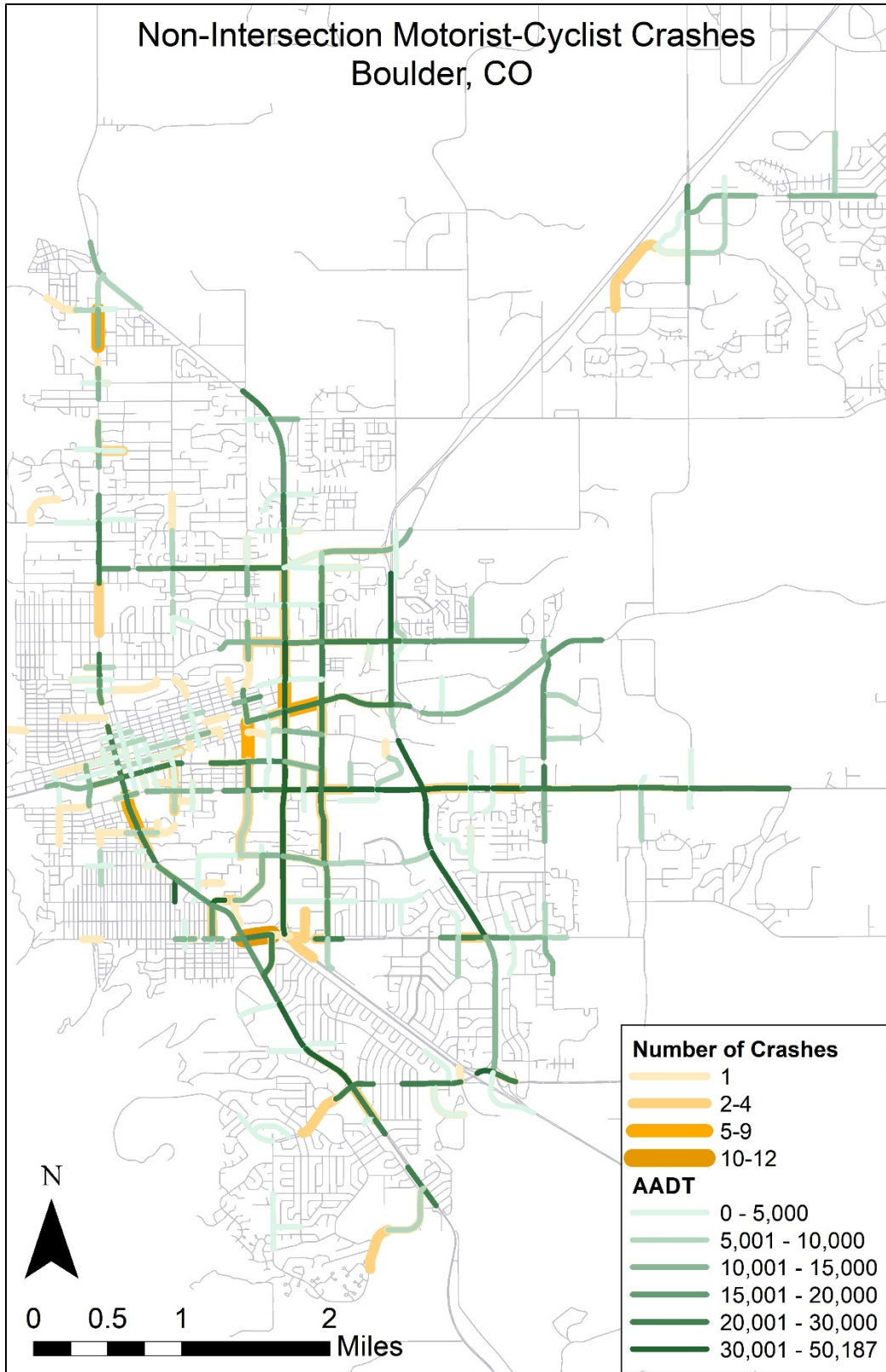


Figure 4.2: Automobile Volumes Map for Boulder

Looking at Figure 4.1, it is apparent that there are several road segments with high numbers of motorist-bicyclist crashes but lower bicycle volumes. From Figure 4.2, it is apparent that these segments also have relatively high motor vehicle volumes. To understand the bicycle-involved crashes on these segments, it is necessary to know how bicyclists are using them. Many of these segments are four lanes, with at-grade rectangular rapid flash beacon (RRFB) crossings, and multiuse paths (wide sidewalks) used for pedestrian and bicyclist travel. The crashes on these segments may involve bicyclists crossing driveways on the sidepaths or crossing the roadway using the RRFBs. A report by the city of Boulder confirms this, indicating that 58percent of all bicycle-involved crashes in the city are at a crosswalk (City of Boulder, 2016). This percent may include driveways. From the city's bicyclist crash database, which is processed using the Pedestrian Bicycle Crash Analysis Tool (PBCAT), 58percent of non-intersection related bicyclist crashes from 2008 to 2014 are driveway, sidewalk, multiuse path, or crosswalk related, while only 30percent were in the travel lane (FHWA, 2016).

#### **4.1.5 Sociodemographics**

Sociodemographic variables such as population density, employment density, and educational attainment were investigated during the modeling effort. Each of these were obtained from the American Community Survey and are further described below (U.S. Census Bureau, 2016).

##### ***4.1.5.1 Population Density***

Population density was computed by intersecting census tracts with a 500-foot buffer around the road segment. The percent of the area within the buffer in each of three categories (high, medium, and low) was computed. The categories were determined using Jenks natural breaks algorithm and are as follows:

- Low < 79
- 79 < Medium < 202
- High > 202

To aggregate these percentages into one number, the percentage of the area in each category was multiplied by the midpoint value of each category in order to get an estimated average population density for the 500-foot buffer around the road segment.

##### ***4.1.5.2 Employment Density***

Employment density was another sociodemographic variable that was considered in the analysis. This variable was also divided into three categories (high, medium, and low) again based on the Jenks natural breaks within a 500-foot buffer. The categories for low, medium, and high are as follows:

- Low < 45
- 45 < Medium < 116
- High > 116

### **4.1.5.3 Educational Attainment**

Educational attainment of the people who live in the area within a 500-foot buffer of the road segment was also considered in the analysis. This variable was divided into three categories – no high school, high school-some college, or college degree and above. These categories were further divided into high, medium, and low based on the Jenks natural breaks method.

#### **No High School**

- Low: < 51
- 51 < Medium < 170
- High: > 170

#### **High School – Some College**

- Low: < 15
- 15 < Medium < 49
- High: > 49

#### **College Degree or above**

- Low: < 20
- 20 < Medium < 55
- High: > 55

## **4.1.6 Land Use**

In this study, four types of land use were considered in the analysis: percent industrial, percent retail, percent residential, and percent university. The data were obtained from Boulder's open data portal (City of Boulder, 2016). The percentage of area surrounding the segment within a 500-foot buffer that fell into each of these four categories was computed and used in the analysis. For example, the land use surrounding a particular segment could be 89 percent industrial, 6 percent residential, 5 percent retail, and 0 percent university.

## **4.1.7 Facilities**

Types of facilities were also considered in the analysis. Ten facility types were considered: bike lanes, multipurpose sidewalks, neighborhood trails, paths, cul-de-sacs, paved trails, paved trails next to road, protected bicycle lanes, separated sidewalks, shared lane bicycle routes, and unpaved trails. The percentages for each of these categories was computed with respect to the percentage of the road segment that contains each of these facilities. For example, a segment could contain 20 percent bike lane and 80 percent separated sidewalk.

## 4.2 METHODS

The purpose of the SPF is to provide estimates of predicted motorist-bicyclists crashes on road segments. Creating an SPF requires the statistical modeling of the above variables with the dependent variable being the count of the number of motorist-bicyclist crashes. Linear models are rarely appropriate for safety studies, so many researchers use a generalized linear model (GLM) that is able to account for a non-normal distribution. A Poisson distribution is often employed with count data such as ours; however, that is only appropriate when the mean equals the variance (Lord et al., 2010). The negative binomial GLM is the generalized form of the Poisson model, which allows it to account for this overdispersion issue that is common in crash data by introducing a stochastic component to the log-linear Poisson mean function relationship (Long, 1997; Marshall et al., 2011). Thus, we created the SPFs using a negative binomial model with a generalized linear model with log link, modeled in SAS using the Genmod procedure.

The basic form of the negative binomial generalized linear regression model used is:

$$\ln\mu_i = \varepsilon + \sum X_i\beta_n$$

Where:

$\mu_i$  = the randomized version of the expected number of collisions at a given road segment  $i$

$\varepsilon$  = the random error term, used to account for overdispersion, estimated by the model

$X_i$  = independent variables

$\beta_n$  = estimated model parameters for motor vehicle ( $n=1$ ) and  $n$  other independent variables

The negative binomial probability distribution is determined by (Long, 1997):

$$P(y_i|X_i) = \frac{\Gamma(y_i + v_i)}{y_i! \Gamma(v_i)} \left(\frac{v_i}{v_i + \mu_i}\right)^{v_i} \left(\frac{\mu_i}{v_i + \mu_i}\right)^{y_i}$$

Where:

$\Gamma$  = the gamma distribution function

$v_i$  = the gamma distribution parameter = 1/dispersion parameter

$y_i$  = the number of crashes at intersection  $i$

The variance of the negative binomial distribution is (Long, 1997):

$$\text{Var}(y_i|X_i) = \mu_i + \frac{\mu_i^2}{v_i}$$

Since the crashes occur on road segments of different lengths and in an eight-year period, an offset of logarithm value of (segment length \* number of years) was used when developing the models, thus the developed SPFs could be used to predict the number of crashes occurring per year per mile on a segment. Since the crashes are occurring on road segments of different lengths over the eight-year period, there are two options to model the crash counts:

Option 1: The model predicts the number of crashes per year, per mile. This is the most common model to predict crashes on a segment. The underlying assumption is that the number of crashes is linear to the segment length.

Option 2: An alternative model is to predict the number of crashes on a segment per year. Segment length becomes an independent variable.

Models were evaluated based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), common model comparison criteria, for which lower values indicate better fit (Akaike, 1973; Schwarz, 1978). The criteria are expressed as shown in the equations below:

$$AIC = 2k - 2\ln(L)$$

$$BIC = k\ln(n) - 2\ln(L)$$

Where:

$k$  = number of free parameters in the model

$n$  = sample size

$L$  = maximized value of the likelihood function

A larger maximized value of the likelihood function ( $L$ ) produces a smaller value for both metrics. Likewise, fewer parameters in the model ( $k$ ) also produces a smaller value for both criteria. The difference in the two criteria is the magnitude of the penalty assigned for the number of parameters in the model. The AIC uses a multiplier of 2.0 and does not account for sample size. The BIC accounts for the size of the sample with the multiplier  $\ln(n)$ . The goal is to select the model with the best fit and the fewest parameters. Thus, the BIC was chosen as the preferred criterion for comparison of the models. However, the AIC was also used when there was no or little change in the BIC. The magnitude of the change in these criteria when choosing among models must also be considered. It can be neglected if the change in both AIC and BIC is less than 2. The difference can be deemed as substantial if the change is between 2 and 6. If the change is greater than 6, then the models with lower values are strongly favored.

We tried both options and found that the parameter estimates from each were comparable. The model comparison criteria AIC are almost the same. The BIC for Option 1 is 2.7 lower than that for Option 2, indicating Option 1 is better than Option 2. Thus Option 1, the crash rates model, was recommended.

### 4.3 SUMMARY

In this study, Boulder was chosen as the location for estimating SPFs for non-fatal bicycle crashes on segments. A number of data sources were needed to estimate SPFs. Crash data were obtained from DRCOG and mapped to a network. Similarly, AADB and AADT were estimated from short-duration counts using expansion factors that were developed from continuous counters. Finally, sociodemographic data such as population density, employment density,

educational attainment, and other data such as land use and types of bicycle facilities were also gathered. The SPFs were created using a negative binomial model with a generalized linear model with log link, modeled in SAS using the Genmod procedure.

## 5.0 RESULTS

Many negative binomial models were explored, including various forms of variables and combinations of variables. An early simple model explored the relationship between number of crashes and AADT and Adjusted AADB, using a GLM negative binomial model. While both variables were significant, the AIC and BIC were much higher than for other models that included additional variables.

In order to explore the possible relationship between bicycle crashes and bicyclist exposure, two forms of exposure were used: Adjusted AADB divided by 100 (Adjusted AADB/100) and the natural logarithm of the Adjusted AADB (ln(Adjusted AADB)). In the first option, the Adjusted AADB was scaled to Adjusted AADB/100 in order to be comparable with ln(Adjusted AADB). During the model exploration, these two forms of exposure (Adjusted AADB/100 and ln(Adjusted AADB)) were both included in a single model and independently in separate models. Similarly, two forms of traffic volume (AADT/1000 and ln(AADT)) and pedestrian volume in the form of pedestrian volume/10 and ln(pedestrian volume) were included in the initial variables list that was explored for the SPF development. Variables with significance above the 5percent level of significance were excluded in the final SPF model. It is worth mentioning that those variables that were correlated either strongly or moderately were not used in the same models.

The model with the lowest AIC included pedestrian volume and percentage of facilities that are bike lanes but not any of the AADB variables. Because this finding was counterintuitive, a model with slightly lower BIC (difference of 1 BIC) was chosen that included adjusted AADB. Since AADB was included, pedestrian volume and percent of facilities that are bike lanes were not significant. The team concluded that the two variables (pedestrian volume and percentage of facilities that are bike lanes) were acting as a surrogate for bicycle volume (Adjusted AADB), and it would be clearer to simply include Adjusted AADB directly.

Based on previous work by the authors and others, an exponential relationship with bicycle volume was expected to be the best fit (Elvik et al., 2009). However, the model with lowest AIC and BIC included Adjusted AADB/100, but not ln(Adjusted AADB).

The final model is shown in Table 5.1 and can be expressed as:

$$C = e^{\varepsilon + \beta_1 AADT/1000 + \beta_2 Adj\_AADB/100 + \beta_3 L + \beta_4 D} \quad (1)$$

where:

$C$  = motorist-bicyclist crashes per year per mile of roadway

$\varepsilon$  = the random error term, used to account for overdispersion, estimated by the model

$\beta_n$  = estimation parameters for the model

$AADT$  = Annual Average Daily Traffic adjusted to the year 2013

$Adj\_AADB$  = AADB adjusted to the year 2013

$L$  = Percent of surrounding area within 500 feet of segments that is retail land use

$D$  = Approximated population density per square mile

$$C = e^{-3.616 + 0.050 * AADT/1000 + 0.139 * Adj\_AADB/100 + 1.973 * L + 0.0002 * D} \quad (2)$$

**Table 5.1: Final Model Estimated Parameters**

Parameter	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
<b>Model AIC = 592, BIC = 615</b>						
<b>Intercept</b>	-3.616	0.391	-4.382	-2.849	85.51	<.0001
<b>AADT/1000</b>	0.05	0.011	0.029	0.071	21	<.0001
<b>Adj AADB/100</b>	0.139	0.063	0.016	0.262	4.93	0.0264
<b>L = % retail land use</b>	1.973	0.354	1.279	2.667	31.05	<.0001
<b>D = approximated population density/square mile</b>	0.0002	0.000	0.0001	0.0003	19.28	<.0001
<b>Dispersion</b>	1.369	0.312	0.876	2.138		

A Cumulative Residual (CURE) plot for the predicted number of crashes for each road segment using eight years of data is shown in Figure 5.1 (Hauer and Bamfo, 1997). The CURE plot indicates that the model fit is relatively good, with the central data within the upper and lower bounds.

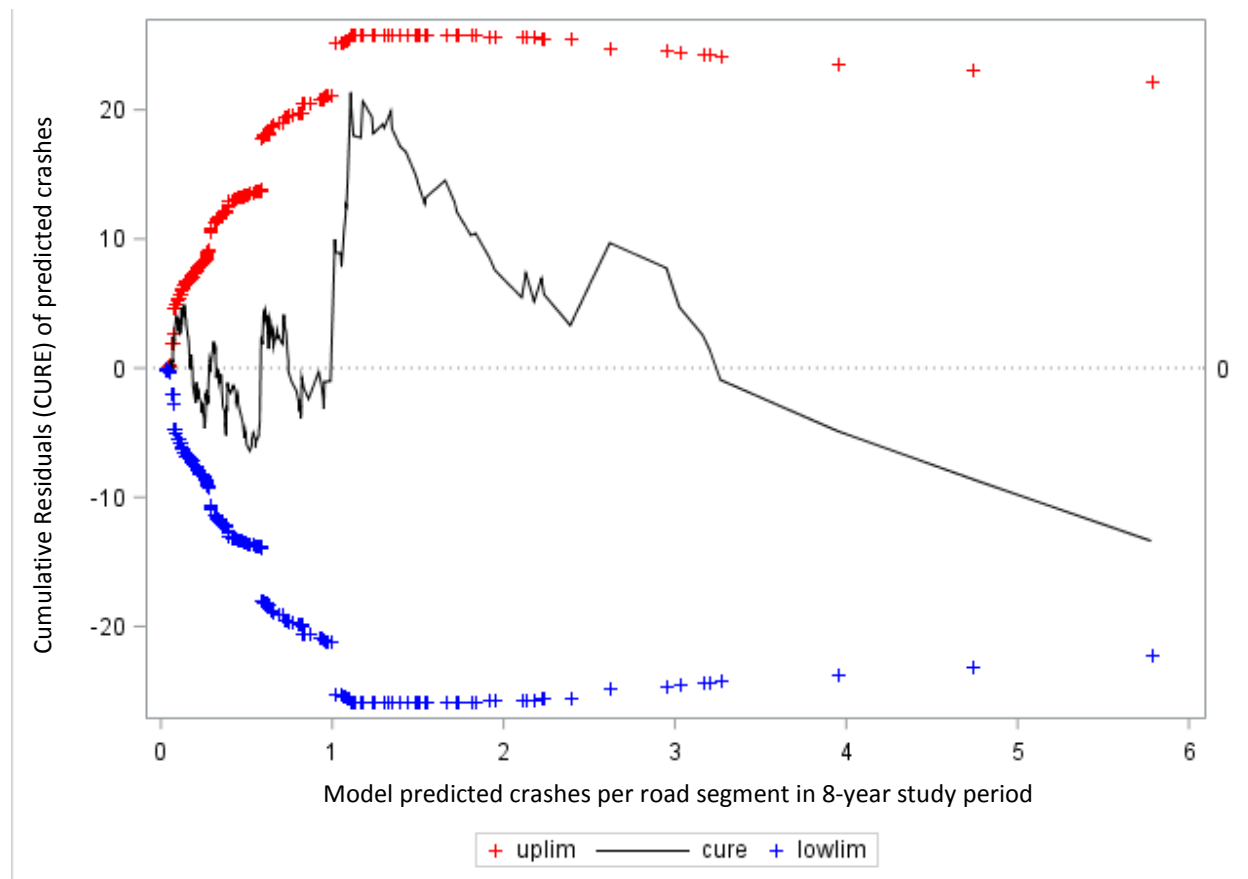


Figure 5.1: CURE Plot for Selected Model



## 5.1 DISCUSSION

The analysis shows that higher motor vehicle volume is a leading factor associated with an increase in crashes between motor vehicles and bicyclists. Bicyclist traffic volume, population density, and percent retail land use are also significant.

Higher levels of retail land use and population density were consistently found to be significantly associated with motorist-bicyclist crashes on road segments across the various models tested. While certain land uses do not directly cause such crashes, areas with more retail development may be more confusing to drivers, perhaps due to more curb cuts, which could complicate the driving task and cause drivers not to see bicyclists either in the road, at a driveway crossing of a sidepath, or at a midblock crossing. Alternatively, bicyclists may make sudden turns or ride on the sidepaths to access retail establishments. Boulder has suburban-style retail developments where parking lot access driveways cross multiuse paths. Pedestrian volumes are moderately correlated with percent retail, indicating that there may be some collinearity between these two variables.

The types of crashes on road segments include crashes at driveways (53percent) as well as bicyclists crossing roadways at midblock locations. From Boulder's bicyclist crash database, which is processed using the Pedestrian Bicycle Crash Analysis Tool (PBCAT), 58percent of non-intersection related bicyclist crashes from 2008 to 2014 are driveway, sidewalk, multiuse path, or crosswalk related, while only 30percent were in the travel lane.

As mentioned previously, several of the road segments with high numbers of crashes shown in Figure 4.1 have RRFs used by bicyclists and multiuse paths. The data used in our analysis do not indicate if the bicyclist was in the roadway, bike lane, midblock crosswalk, or on a sidewalk/multiuse path driveway. However, according to Safe Streets Boulder, only 7percent of total crashes occur in bicycle lanes (including one fatality), and another 7percent occur in the travel lane while 58percent occur in crosswalks (City of Boulder, 2016). If the crash types on these segments are primarily crosswalk related, this may explain why pedestrian volume might be a significant predictor, since these crossings are primarily designed for pedestrians. Thus, pedestrian volume may be a surrogate for bicyclists using midblock or driveway crosswalks. However, since pedestrian traffic volume data are rare, it is unlikely that pedestrian volumes would be available where bicycle volume data are not available.

Also of interest are the variables that were explored but not included in the final model. Some variables, like employment density and education level, were correlated with population density and thus not included to reduce potential multicollinearity issues. Other variables, such as infrastructure variables for the presence of sidewalks, presence of protected bike lane, etc. were not included because they were often not present, meaning that many of the values were zero.

It should be noted that since all the data have been collected from Boulder, the results may not be applicable to other communities. A full list of limitations of this work follow.

## 5.2 INTERPRETATION

After the models were developed, the next step was to interpret what these mean for the safety of road users. The final model includes the following significant variables: AADT, AADB, percent retail land use, and population density. The following sections provide a better understanding of the range of values for which this model is applicable.

## 5.2.1 AADT

Figure 5.2 shows a histogram of the AADT values for the 347 road segments studied. Figure 5.2 clearly illustrates that the majority of the AADT observations that were used in the model were between 0 and 30,000. Thus, extrapolating beyond 30,000 AADT is not recommended.

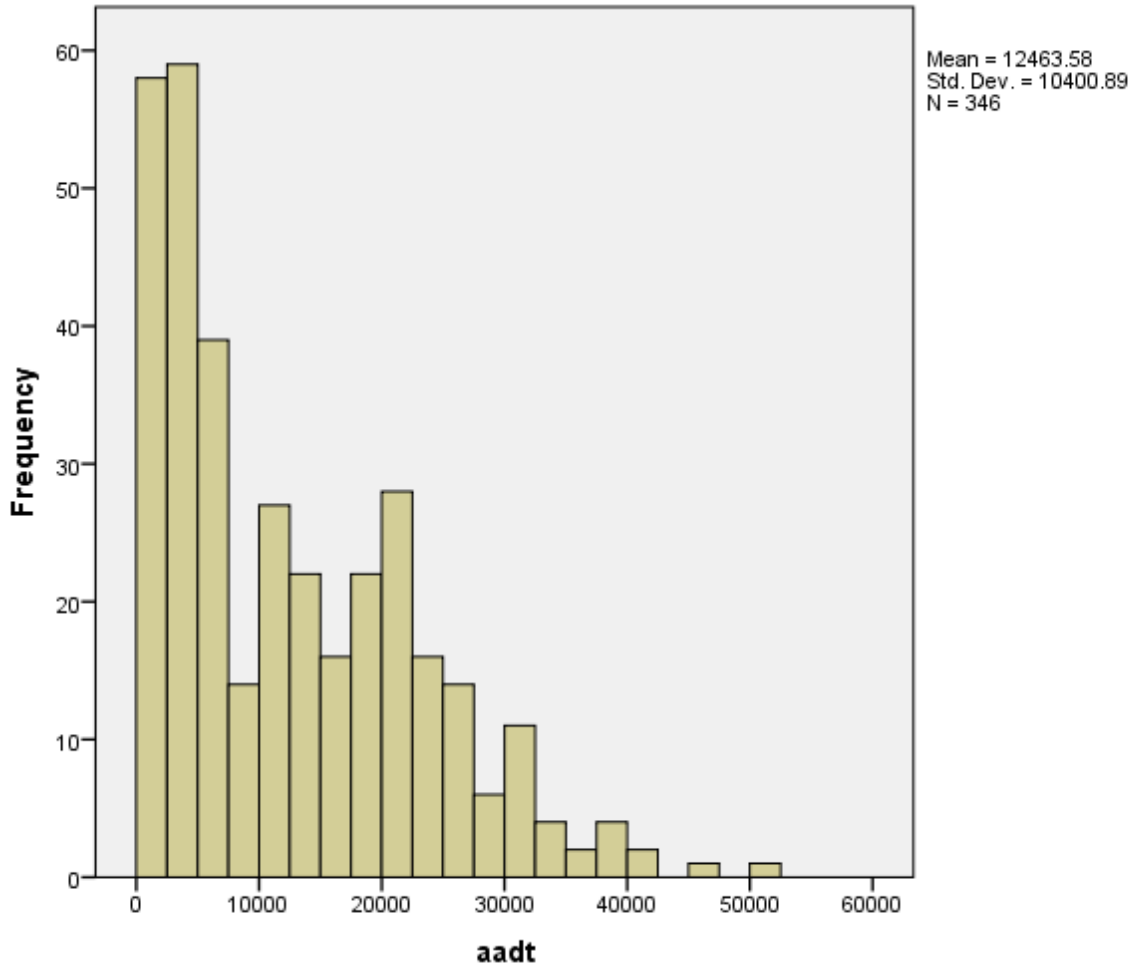


Figure 5.2: Histogram of AADT

Figure 5.3 shows a histogram of the Adjusted AADB values for the road segments studied. The model was built based primarily on Adjusted AADB values between 0 and 600. Thus, extrapolating beyond 600 AADT is not recommended.

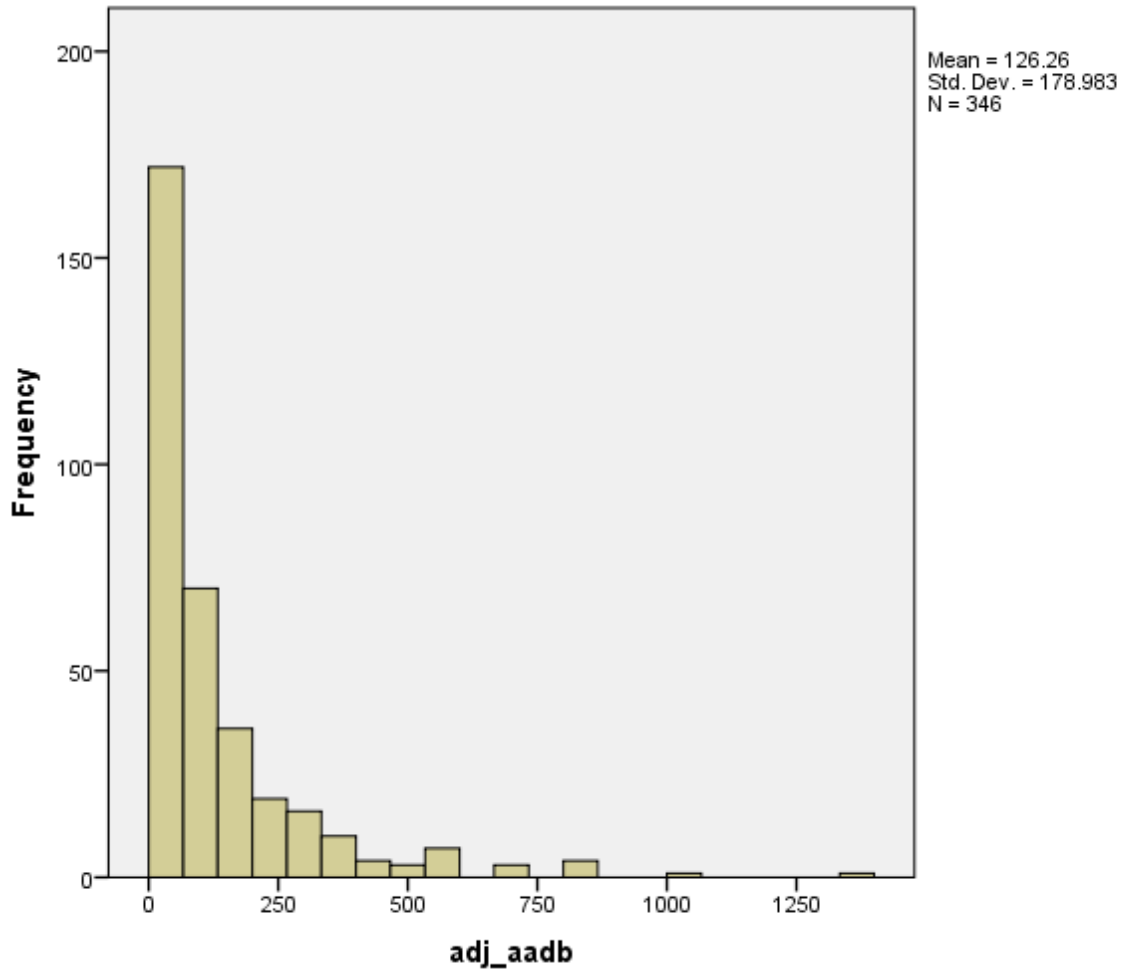


Figure 5.3: Histogram of Adjusted AADB

## 5.2.2 Percent Retail

Figure 5.4 shows a histogram of the percent residential values for the road segments studied. Though there are a high number of road segments with no retail land use, the model was built based on values spread somewhat evenly between 0 and 100 percent. Thus, the model may be applicable for any percent retail but is especially suited to lower percent retail values.

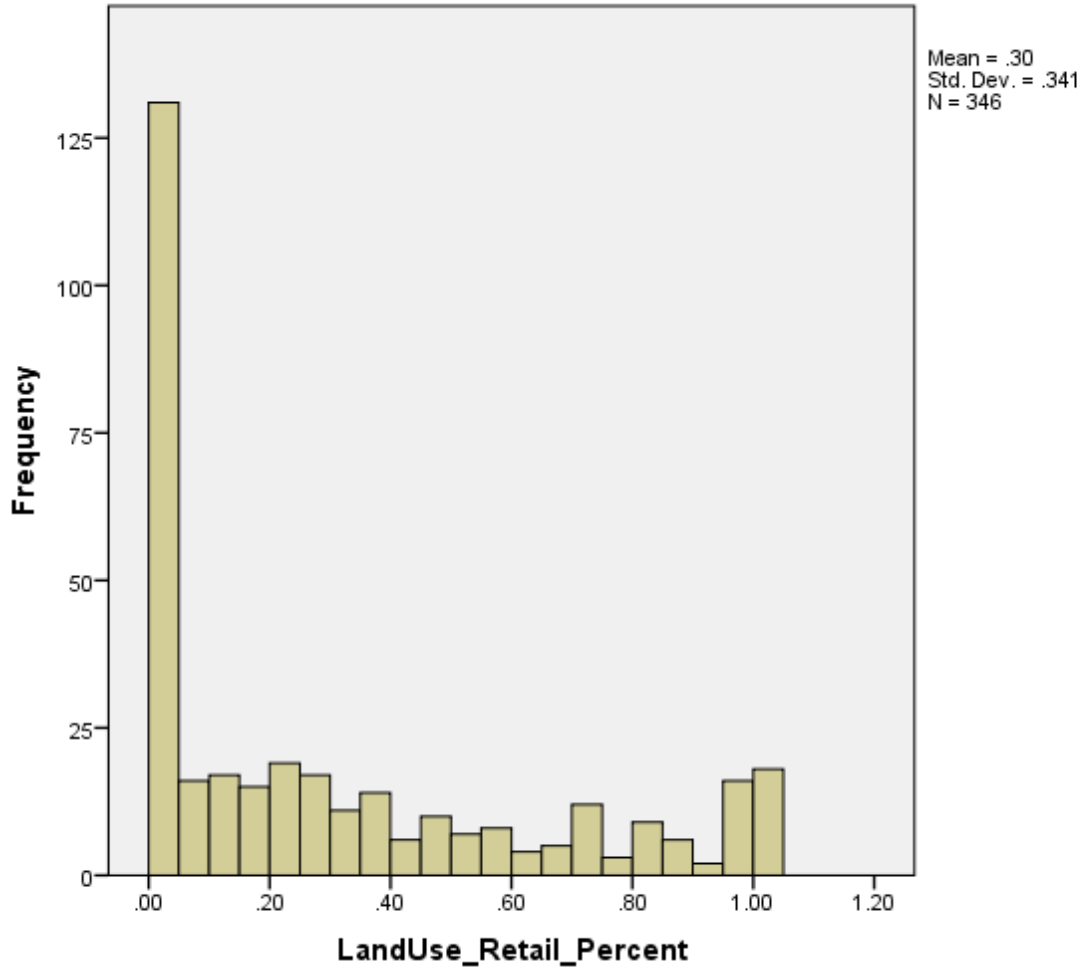


Figure 5.4: Histogram of Land Use Percent Retail

### 5.2.3 Population Density

Similarly, because there are few areas in Boulder that have population densities above 12,000 per square mile (Figure 5.5), extrapolating the results of the model to denser, more urban areas would not be appropriate. The study city is a university town of roughly 100,000 people with an urban core surrounded by suburban-style development, mountain parks, and open space. Thus, using the results of this model to apply to New York City, for example, may not be appropriate.

The U.S. Census Bureau uses a population density threshold of 1,000 per square mile as part of its definition of urban areas (U.S. Commerce Department, 1994). All of the areas studied are above 2,000 people per square mile, well above the minimum for urban areas.

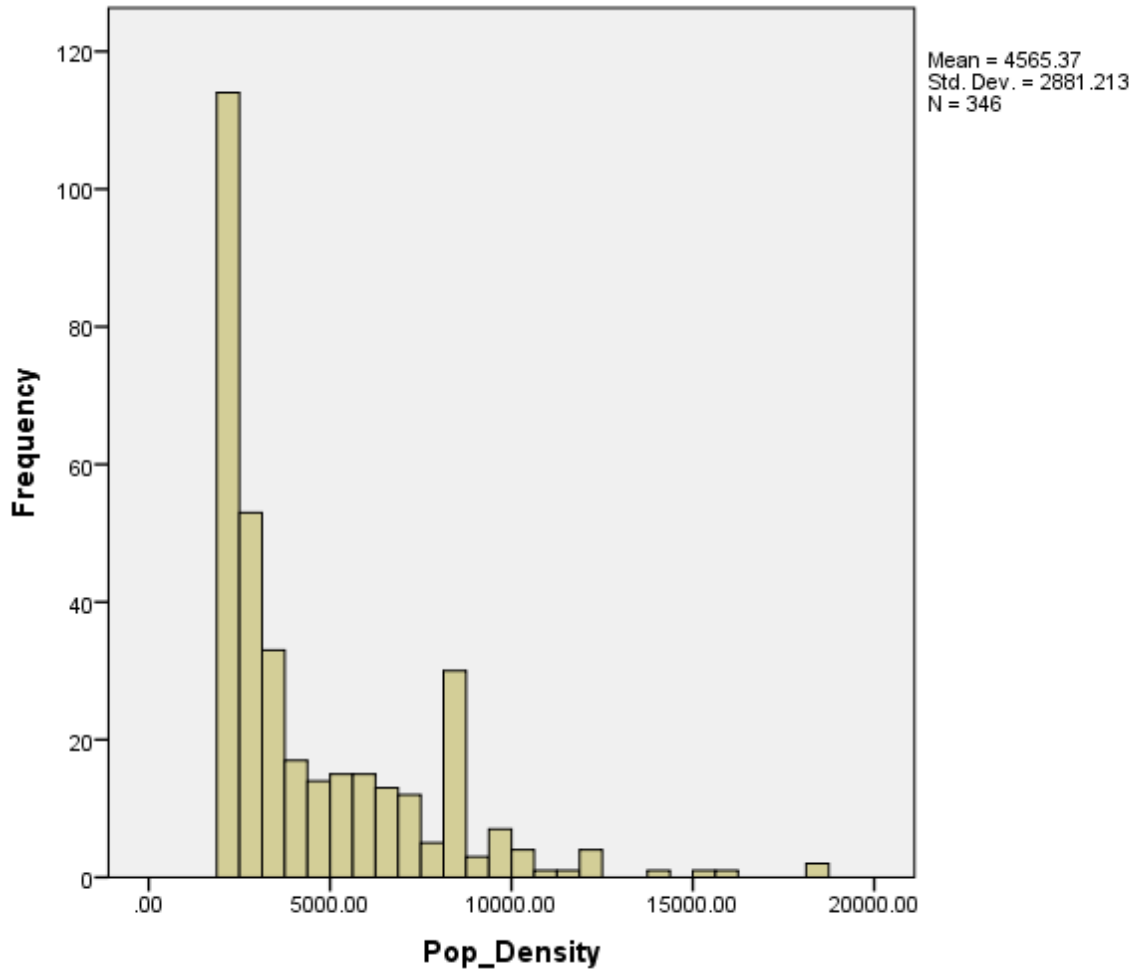


Figure 5.5: Histogram of Population Density per Square Mile

### 5.2.4 Model Fit

To better visualize how the model fits the data, the relationship between the predicted and actual crashes and Adjusted AADB for a given AADT value (10,000 AADT) were plotted in Figure 5.6. Here we see that while the predicted curve does trend up for values above 300 Adjusted AADB (reverse of the safety in numbers effect), the predicted line is nearly linear for values below 300 Adjusted AADB. This is not in accord with the expected safety in numbers effect, since it predicts an approximately linear relationship between Adjusted AADB and crashes.

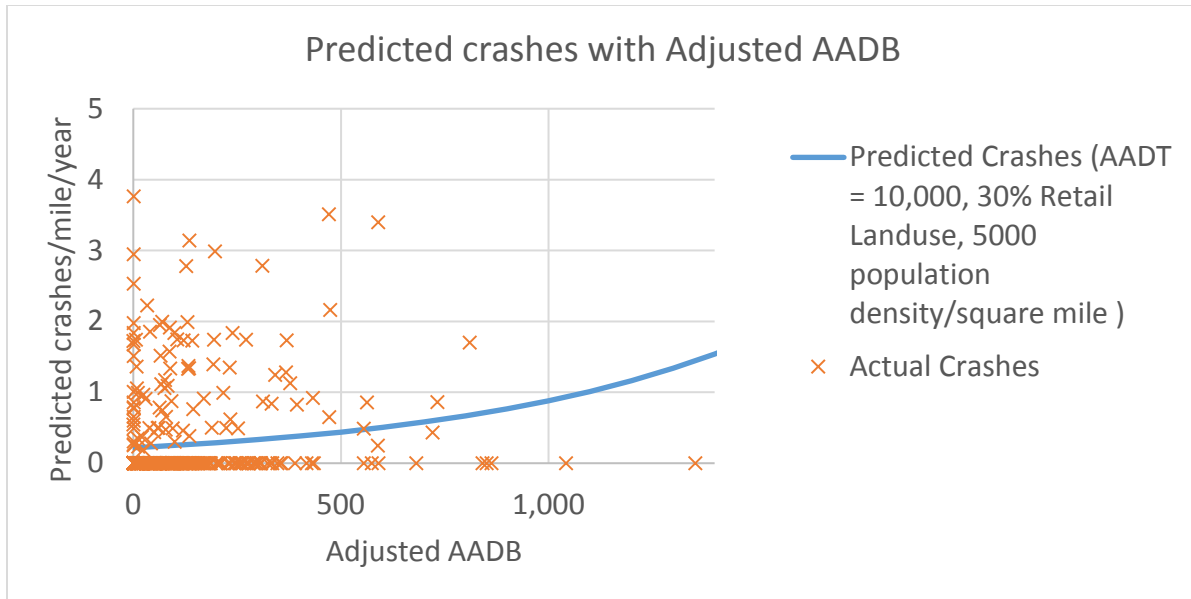


Figure 5.6: Predicted and Actual Crashes vs. Adjusted AADB

Similarly, model-predicted and actual number of crashes in relation to AADT can be visualized for different values of Adjusted AADB as shown in Figure 5.7 for 30 percent retail land use and 5,000 population density per square mile. As for AADB, the upward trending curve is the opposite of what is expected based on the “safety in numbers” hypothesis. Here again we see that there are few segments above 30,000 AADT and that the curves are close to linear for values below 30,000 AADT. The figure also shows that there is relatively little difference in predicted crashes between 5 and 100 AADB, but there is a relatively larger increase in predicted crashes for higher bicycle volumes at 600 AADB.

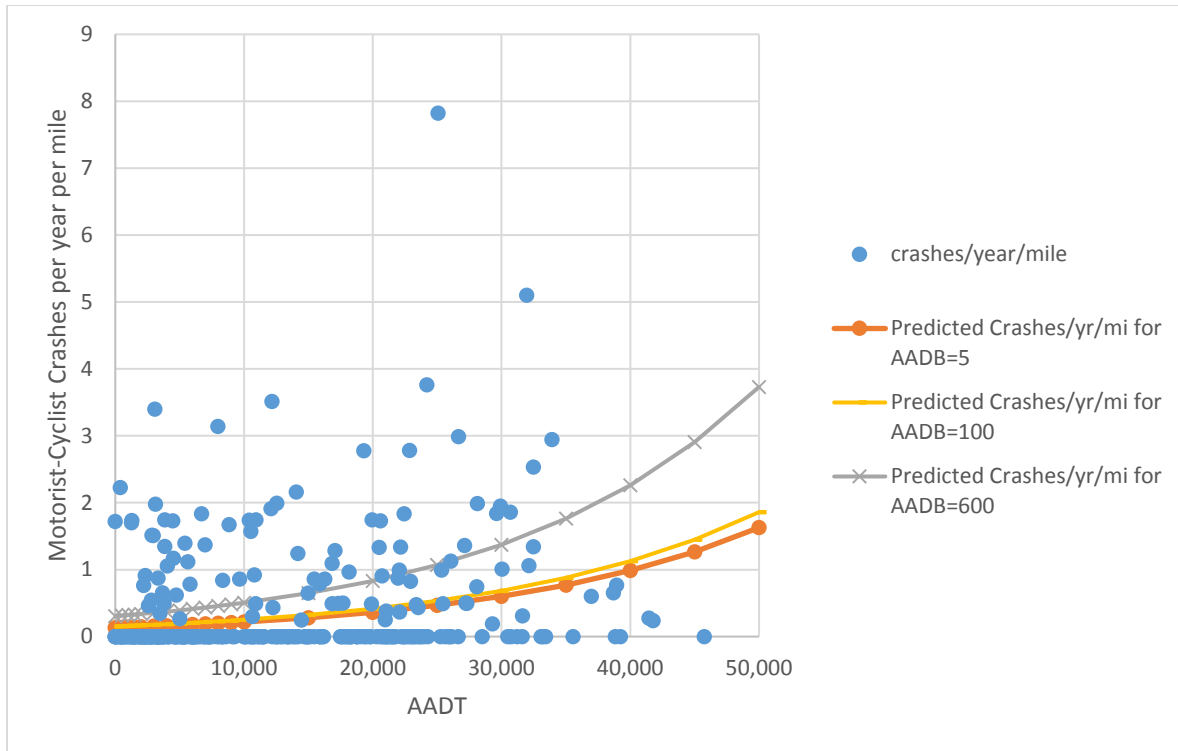


Figure 5.7: Predicted and Actual Crashes vs. AADT

### 5.2.5 Generalizability

The current HSM uses a very simplistic method to estimate bicycle crashes, which involves applying a factor to total motor-vehicle crashes without including the number of cyclists. In this study, we incorporated bicycle volumes to generate a bicycle-specific SPF. However, the model was created for a single small university city in Colorado. Caution should be exercised when applying it beyond the area in which it was created. However, the research team wanted to present the findings in a way that might be useful for practitioners. In addition, the research team also wanted to investigate what format might be adoptable by the next edition of the HSM. Table 5.2 shows the range of predicted motorist-bicyclist intersection crashes based on ranges of input variables that were significant predictors in the model.

**Table 5.2: Range of Predicted Motorist-bicyclist, Non-intersection Crashes (per mile per year)**

Population Density	AADB	Percent Retail Land Use	Motor Vehicle Traffic: AADT		
			Low (<10,000)	Medium (10k-20k)	High (20k-30k)
Low (2,000-5,000/sq. mile)	Low (0-200)	Low (0-20%)	0.04 - 0.2	0.06 - 0.4	0.1 - 0.6
		Medium (21-60%)	0.06 - 0.5	0.1 - 0.8	0.2 - 1
		High (60-100%)	0.1 - 1	0.2 - 2	0.3 - 3
	Medium (200-600)	Low (0-20%)	0.05 - 0.4	0.08 - 0.6	0.1 - 1
		Medium (21-60%)	0.08 - 0.8	0.1 - 1	0.2 - 2
		High (60-100%)	0.2 - 2	0.3 - 3	0.5 - 5
Medium (5,000-12,000 /sq. mile)	Low (0-200)	Low (0-20%)	0.07 - 0.8	0.1 - 1	0.2 - 2
		Medium (21-60%)	0.1 - 2	0.2 - 3	0.3 - 5
		High (60-100%)	0.2 - 4	0.4 - 6	0.6 - 10
	Medium (200-600)	Low (0-20%)	0.09 - 1	0.2 - 2	0.2 - 4
		Medium (21-60%)	0.1 - 3	0.2 - 5	0.4 - 8
		High (60-100%)	0.3 - 7	0.5 - 11	0.8 - 18

Average daily bicycle volumes have been broken into categories developed by past research:

- Colorado bicycle count sites were categorized by AADB into three categories: Low <200 per day; Medium 200-600; and High >600 (Nordback et al., 2013).
- NCDOT categorized Average Daily Bicyclists into three categories: Low <100; Medium 100 to 500; and High >500 (O'Brien et al., 2016, Page 10-8, Figure 12);

For the purposes of this study, based on the histogram in Figure 5.3 and Jenks natural breaks analysis, which broke the distribution into groups at 0-165, 165-500, and >500, we are categorizing our sites into 0-200 and 200-600. As mentioned previously, sites with over 600 are high volume sites, and we do not have enough such sites in the sample data to extrapolate above 600 annual average bicyclists per day.

Percent retail is categorized into three categories using Jenks natural breaks: 0-20percent retail is considered Low Retail, 20-60percent retail is considered Medium Retail, and >60percent retail is considered High Retail.

Population density is divided into categories based on the Jenks natural breaks discussed earlier, with categories being defined as Low (<4,700 per square mile), Medium (4,700 to 12,000 per square mile), and High (>12,000 per square mile). To simplify slightly, we propose breaking these into similar groups: Low density 2,000-5,000 per square mile; Medium density 5,000 to 12,000 per square mile. We do not use the highest group (>12,000 per square mile) because we have too few road segments in that category. We also excluded very low density (below 2,000 people per square mile) since these are not represented in the studied data set.

While the team does not recommend that the numbers in Table 5.2 be adopted by the HSM, since they were estimated based on data from one city only, the research team proposes Table 5.2 as a possible format for helping practitioners use the results of research such as the model presented herein. Such a format may be useful in future updates to the HSM.



## 5.3 LIMITATIONS

There are many limitations of our findings based on the data and the analysis. Data-related limitations include:

- The bicycle and pedestrian volumes used have not been validated. It is possible that the pedestrians or bicycles were not correctly counted by the video image processing software used in the turning movement count process or that bicycles may be counted as pedestrians or vice versa.
- Data were only collected for the city of Boulder, a college town of approximately 100,000 people, identified by the League of American Bicyclists as one of only four Platinum Bicycle Friendly Communities in the United States. Results may not apply to other communities with different sizes, characteristics, types of bicycle infrastructure and lower levels of bicycle friendliness.
- Many variables that might impact bicycle safety were not included in this analysis because they were not readily available. For example, motor vehicle posted and actual speed was not included.
- Bicycle infrastructure included in this study was primarily standard bike lanes, generally 4 to 5 feet in width. While Boulder does have other infrastructure types like bicycle routes, multiuse paths paralleling streets, and protected bike lanes, there were either not enough of these or not a large enough impact on safety to be included as a significant variable in the analysis.
- Though midblock crossings may be impacting safety on the road segments studied, we did not have a comprehensive list of locations of marked midblock crossings, nor of crossings with the rectangular rapid flash beacons (RRFBs) common in the city.
- Based on the range of values included in the data, predicted crashes with values beyond the following are not reliable. The model applies best to the values in the following ranges:
  - AADT between 0 and 30,000
  - AADB between 0 and 600 bicyclists per day
  - Population density 2,000 to 12,000 per square mile
  - Any value of percent retail land use

## 5.4 SUMMARY

In this section, we present the best-fit negative binomial model with log link to predict annual non-fatal, motorist-bicyclist crashes on road segments per mile. It requires four input variables: AADT, AADB, percent retail land use, and population density. Surprisingly, the exponential form of the model does not support the “safety in numbers” effect, seen so commonly in other research including the bicycle SPF developed previously for intersections in Boulder (Nordback, et al., 2014). For the lower values of AADT and AADB on which the model was built, the model appears close to linear in form.

We then discuss over what range of values the model should be applied and create a table to present the results in a format that might serve as a potential starting point for future efforts to generalize the results of models for possible use in the updated HSM.

## 6.0 CONCLUSIONS AND RECOMMENDATIONS

On-road bicycling inevitably carries risk of motorist-bicyclist collisions. However, the health benefits of bicycling have been found to outweigh safety risks (de Hartog et al., 2010). Improving bicycling safety is therefore a valuable pursuit. An important insight in this pursuit of improved bicycling safety is a better understanding of what factors are associated with motorist-bicyclist collisions. However, the common practice of deducing a rate of collisions per vehicle represents a fundamental misunderstanding of this risk (Hauer, 2010). Safety performance functions (SPFs) are able to estimate the relationship between collisions and exposure by accounting for the nonlinear relationship between exposure and risk. While SPFs that estimate motor vehicle collisions have been comprehensively developed, those that estimate bicycle collisions have not (AASHTO, 2010).

Existing bicycle SPFs primarily focus on intersections, which is where the majority of bicycle crashes occur (Nordback, 2014). However, no bicycle SPFs in the United States that utilize bicycle volume data exist for segments. This study adopts methods from the Highway Safety Manual (HSM) used for motor vehicle SPFs in order to develop bicycle-specific SPFs for roadway segments in Boulder, CO (AASHTO, 2010). This is a preliminary effort at developing a bicycle-specific SPF for segments in the U.S. that utilizes bicycle volumes, is an important first step towards further understanding bicyclist safety, and may inform future versions of the HSM.

First a comprehensive literature review was undertaken to review existing literature on exposure measures, bicycle infrastructure, and studies that incorporate bicycle volumes to determine bicycle safety impacts. Both population- and travel-based exposure measures have been used in the literature to study the risk associated with cycling. Additionally, facilities have been studied extensively as they relate to bicycle safety. The literature review also revealed that studies involving the use of the SPFs are limited and, at the time of writing, the authors are not aware of any study that has developed SPFs for bicycle crashes and bicycle volume for a U.S. city along road segments.

Prior to choosing locations to study, the team reviewed existing data sources by community in order to find study areas with consistent data or at least to identify one area with sufficient data for bicycle-safety study. Our research team used non-fatal cyclist crash data from eight different communities: Arlington, VA(city/county); Bellingham, WA(city); Boulder, CO (city); Denver, CO(city/county); Minneapolis and St. Paul, MN(cities); Philadelphia, PA(city/county); Portland, OR(city); and San Diego, CA (county). Our research team examined these data for issues pertaining to quality and consistency and ultimately selected Boulder for further analysis since it had the best available data.

In this analysis, a negative binomial model with log link was used to predict annual, non-fatal, motorist-bicyclist crashes on road segments per mile. While both vehicle volume and bicycle volume data are used in the model in order to account for the “safety in numbers” effect, the model did not demonstrate this effect seen so commonly in other research, including the bicycle SPF developed previously for intersections in Boulder (Nordback et al., 2014). The analysis shows that motor vehicle volume is a leading factor associated with increases in crashes between motor vehicles and bicyclists. Bicyclist exposure, population density, and percent retail land use are also predictive. This study also investigated the potential of various simplified methods to include bicycle volumes in future versions of the HSM. The project

examined how and if the SPFs developed could be simplified for use in future versions of the HSM.

## **6.1 RECOMMENDATIONS**

There are many implications of this work for practice:

- Jurisdictions could build on the work we have done here and construct similar SPFs to model bicycle crashes on their roadways.
- Others can compare the models they develop for motorist-bicyclist crashes on urban road segments to our findings to see if they find similar curve shapes and similar significant variables.
- The developed SPFs can then be used to predict crashes along road segments and determine where unexpectedly high numbers of crashes exist, using the Empirical Bayes approach from the HSM. This information can then be used to prioritize road segments for improvements.
- Based on the model, cities may also work to reduce motor vehicle volumes, especially on roads with high bicycle volumes, in pursuit of reducing motorist-cyclist crashes.
- Similarly, jurisdictions may focus efforts on concentrating bicycle travel on corridors with low motor vehicle volumes.
- Special care should be taken in areas with high retail activity where cyclists and drivers make frequent turns and other maneuvers to access businesses. Reducing speeds in such areas may be a low-cost solution.
- Providing safer midblock crossings for bicyclists as well as pedestrians by using crossing treatments similar to those found effective for pedestrians (median refuge islands, rectangular rapid flash beacons, pedestrian hybrid beacons, etc.) may also reduce motorist-bicyclist crashes.
- Separated bicycle lanes should also be considered to reduce motorist-bicyclist crashes.

## **6.2 FUTURE WORK**

Having bicycle SPFs for road segments is essential if we seek to rely upon quantitative analysis in our transportation decision making. It also represents a necessary step to improving our holistic understanding of bicycling safety. This work puts us on that path to a broader awareness of bicycle safety. We look at Boulder as an exploration of how a bicycle-specific SPF for segments may better inform bicycle safety.

## REFERENCES

AASHTO, Highway Safety Manual, 2010.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle,” In Petrov, B.N. and F. Csaki (Eds.), Second International Symposium on Information Theory, 267-281. Budapest, Akademiai Kiado.

Akar, G., Clifton, K. 2009. Influence of Individual Perceptions and Bicycle Infrastructure on Decision to Bike. *Transportation Research Record*, Issue 2140, pp. 165-172.

Berkow, Matthew, et. al. (2017). Improvements to Statewide Collision Reporting to Understand Sidewalk-Related Bicycle Collisions, Transportation Research Board, Washington DC, United States.

Bly, P., Dix, M., and Stephenson, C. (1999). Comparative Study of European Child Pedestrian Exposure and Accidents, MVA Limited, Surrey, England.

Boulder. *Transportation Division Traffic Count e-resource*. 2010 May 7, 2010]; Available from: [www.bouldercolorado.gov/index.php?option=com\\_content&task=view&id=12226&Itemid=1198](http://www.bouldercolorado.gov/index.php?option=com_content&task=view&id=12226&Itemid=1198).

Brüde, U., Larsson, J., 1993. Models for predicting accidents at junctions where pedestrians and cyclists are involved. How well do they fit? *Accident Analysis and Prevention* 25, 499–509.

Carlson, K., Murphy, B., Ermagun, A., Levinson, D., and A. Owen. Safety in Numbers: Pedestrian and Bicyclist Activity and Safety in Minneapolis. Final Report. Center for Transportation Studies, University of Minnesota, 2018.

Chen, P. (2015). Built environment factors in explaining the automobile-involved bicycle crash frequencies: A spatial statistic approach. *Safety Science*, 79, 336-343.

Chu, X. (2003). The Fatality Risk of Walking in America: A Time-Based Comparative Approach, 4th International Conference on Walking in the 21st Century, Portland, OR.

City of Boulder Transportation Division, “Safe Streets Boulder Report.” 2016, City of Boulder: Boulder, Colorado. p. 64.

de Hartog, J. J., H. Boogaard, H. Nijland, and G. Hoek. Do the Health Benefits of Cycling Outweigh the Risks? *Environmental Health Perspectives*, Vol. 118, No. 8, 2010, pp. 1109-1116.

El Esawey, M. Estimation of Annual Average Daily Bicycle Traffic with Adjustment Factors. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2443, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 106–114.

Ekman, L., 1996. *On the Treatment of Flow in Traffic Safety Analysis. A non-parametric approach applied on vulnerable road users*. Ph.D. Dissertation. Bulletin 136. Lund Institute of Technology, Department of Traffic Planning and Engineering, Lund.

Elvik, R. The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis and Prevention* 41 (2009) 849–855.

Elvik, R., and A. Mysen. Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1665, Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 133-140.

FHWA. (2017). *FHWA Strategic Plan*. FHWA-PL008-027. U.S. Department of Transportation, Washington, D.C. Originally published October 2008, Revised January 2017.

Ferrara, T. C. *Statewide Safety Study of Bicycles and Pedestrians on Freeways, Expressways, Toll Bridges, and Tunnels*. Mineta Transportation Institute, San Jose, CA, 2001, pp. 165.

Gates, T. J., P. T. Savolainen, S. Stapleton, T. Kirsch, and S. Miraskar. *Development of Safety Performance Functions and Other Decision Support Tools to Assess Pedestrian and Bicycle Safety*. Transportation Research Center for Livable Communities, Kalamazoo, MI, 2016.

Geyer, J., Raford, N., Pham, T., Ragland, D.R., 2006. Safety in numbers. Data from Oakland, California. *Transportation Research Record* 1982, 150–154.

Goodno, M., McNeil, N., Parks, J., Dock, S. 2013. Evaluation of Innovative Bicycle Facilities in Washington, DC Pennsylvania Avenue Median Lanes and 15th Street Cycle Track. *Transportation Research Record*, Issue 2387, pp. 139-148.

Hamann, C., and C. Peek-Asa. On-Road Bicycle Facilities and Bicycle Crashes in Iowa, 2007-2010. *Accident Analysis & Prevention*. Vol. 56, 2013, pp. 103-109.

Hankey, S., Lindsey, G., and Marshall, J. (2014). “Day-of-Year Scaling Factors and Design Considerations for Non-motorized Traffic Monitoring Programs.” *Transportation Research Record*. *Journal of the Transportation Research Board*, No. 2468, Transportation Research Board of the National Academies, Washington, D.C., pp. 64–73.

Harkey, D.L., et al., Pedestrian and Bicycle Crash Analysis Tool (PBCAT): Version 2.0 Application Manual. 2006, University of North Carolina: Chapel Hill. p. 241.

Harwood, D.W., Torbic, D.J., Gilmore, D.K., Bokenkroger, C.D., Dunn, J.M., Zegeer, C.V., Srinivasan, R., Carter, D., Raborn, C., Lyon, C., Persaud, B., 2008. Pedestrian Safety Prediction Methodology. NCHRPWeb-only Document 129: Phase III. Transportation Research Board, Washington, DC.

- Hauer, E. and Bamfo, J. (1997). Two Tools for Finding What Function links the Dependent Variable to the Explanatory Variables. Proceedings of ICTCT 1997 Conference, Lund.
- Hauer, E. On Exposure and Accident Rate. *Traffic Engineering & Control*, Vol. 36, No. 3, 1995, pp. 134-138.
- Hauksson, R.G., Bicycle Safety in Gothenburg: A case study of bicycle – motor vehicle collisions on one and two-way cycle paths at intersections, in Department of Civil and Environmental Engineering. 2014, Chalmers University of Technology: Goteborg, Sweden.
- Hunter, W., Harkey, D., Stewart, J., Birk, M. 2000. Evaluation of blue bike-lane treatment in Portland, Oregon. Pedestrian and BIcycle Transportation Research, Issue 1705, pp. 107-115.
- Hunter, W. W., J. C. Stutts, W. E. Pein, and C. L. Cox. *Pedestrian and Bicycle Crash Types of the Early 1990's*. Highway Safety Research Center, University of North Carolina, Chapel Hill, NC, 1996, pp. 192.
- Jacobsen, P.L., 2003. Safety in numbers: more walkers and bicyclists, safer walking and cycling. *Injury Prevention* 9, 205–209.
- Jonsson, T., 2005. *Predictive models for accidents on urban links*. A focus on vulnerable road users. Ph.D. Dissertation. Bulletin 226. Lund Institute of Technology, Department of Technology and Society, Traffic Engineering, Lund.
- Kaplan, J.A. (1975). Characteristics of the Regular Adult Bicycle User, 132, University of Maryland College Park, College Park, MD.
- Keall, M.D. (1995). “Pedestrian Exposure to Risk of Road Accident in New Zealand,” *Accident Analysis and Prevention*, 27, 729–740.
- Kononov, J., and B. K. Allerly. Level of Service of Safety: Conceptual Blueprint and Analytical Framework. *Transportation Research Board: Journal of the Transportation Research Board*, No. 1840, 2004, pp. 57-66.
- Kroyer, H. R. G. Pedestrian and Bicyclist Flows in Accident Modeling at Intersections: Influence of the Length of Observational Period. *Safety Science*, Vol. 82, 2016, pp. 315-324.
- Kuzmyak, J., Walters, J., Bradley, M. and K. Kockelman. Estimating Bicycling and Walking for Planning and Project Development: A Guidebook. National Cooperative Highway Research Program, Report 770, Transportation Research Board, Washington DC, 2014.
- Leden, L., Gårder, P., Pulkkinen, U. An Expert Judgement Model Applied to Estimating the Safety Effect of a Bicycle Facility. *Accident Analysis & Prevention*, Vol. 32, No. 4, 2000, pp. 589-599.

Leden, L., Gårder, P., Pulkkinen, U., 1998. Measuring the safety of raised bicycle crossings using a new research methodology. In: Paper Presented at the Transportation Research Board Annual Meeting, Washington, DC.

Long, J.S., 1997. Regression models for categorical and limited dependent variables SAGE Publications, Thousand Oaks.

Loo, B.P.Y., and K.L.Tsui. Factors Affecting the Likelihood of Reporting Road Crashes resulting in Medical Treatment to the Police. *Injury Prevention*, Vol 13, Issue 3, 2007.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice* 44 (5), 291-305.

Lyon, C., Persaud, B.N., 2002. Pedestrian collision prediction models for urban intersections. *Transportation Research Record* 1818, 102–107.

Marshall, W.E., Garrick, N.W., 2011. Does street network design affect traffic safety? *Accident Analysis & Prevention* 43 (3), 769-781.

McAndrews, C., Beyer, K., Guse, C., and P. Layde. Revisiting exposure: Fatal and non-fatal traffic injury risk across different populations of travelers in Wisconsin, 2001–2009. *Accident Analysis and Prevention*, [Volume 60](#), November 2013, Pages 103–112.

McArthur, A., Savolainen, P., and T. Gates. 2014. Spatial Analysis of Child Pedestrian and Bicycle Crashes. *Transportation Research Record*, Issues 2465, pp. 57-63.

McKenzie, B. Modes Less Traveled-Bicycling and Walking to Work in the United States: 2008-2012, in *American Community Survey Reports 2014*. U.S. Census Bureau: Washington, D.C. p. 18.

McNeil, N., Monsere, C., Dill, J. 2014. The Influence of Bike Lane Buffer Types on Perceived Comfort and Safety of Bicyclists and 1 Potential Bicyclists.

Medury, A., Grembek, O., Loukaitou-Sideris, A., and K. Shafizadeh. Investigating the Underreporting of Pedestrian and Bicycles Crashes in and Around University Campuses – A Crowdsourcing Approach. In *Accident Analysis and Prevention*, August 2017.

Milne, A. and M. Melin, *Bicycling and Walking in the United States: 2014 Benchmarking Report*, 2014, Alliance for Biking and Walking.

Minikel, E., “Cyclist Safety on Bicycle Boulevards and Parallel Arterial Routes in Berkeley, California.” Presented at the 90<sup>th</sup> Meeting of the Transportation Research Board, Washington, D.C., (2011).

Miranda-Moreno, L. F., Nosal, T., Schneider, R. J., and Proulx, F. (2013). "Classification of

bicycle traffic patterns in five North American Cities." *Transportation Research Record: Journal of the Transportation Research Board* (2339), 68-79.

Miranda-Moreno, L.F., Strauss, J., Morency, P., 2011. Exposure measures and injury frequency models for analysis of cyclist safety at signalized intersections. In: *90th Annual Meeting of the Transportation Research Board*. Transportation Research Board of the National Academies, Washington, DC.

Molino, J., Kennedy, J., Inge, P., Bertola, M., Beuse, P., Fowler, N., Emo, A., and A. Do. A Distance-Based Method to Estimate Annual Pedestrian and Bicyclist Exposure in an Urban Environment. Report FHWA-HRT-11-043, Federal Highway Administration, 2012.

Monsere, C., Dill, J., McNeil, N., Clifton, K., Foster, N., Goddard, T., Berkow, M., Gilpin, J., Voros, K., vn Hengel, D., Parks, J. 2014. Lessons from the Green Lanes: Evaluating Protected Bike Lanes in the U.S. NITC.

National Highway Traffic Safety Administration, (NHTSA). *Fatality Analysis Reporting System*. 2018.

National Highway Traffic Safety Administration (NHTSA). (2018). Quick Facts 2016. Washington, DC. (US Department of Transportation)

National Highway Traffic Safety Administration. Traffic Safety Facts. Bicyclists and Other Cyclists, 2013 Data. Accessed at <http://www-nrd.nhtsa.dot.gov/Pubs/812151.pdf>

Nordback, K. Estimating Annual Average Daily Bicyclists and Analyzing Cyclist Safety at Urban Intersections. University of Colorado Denver, 2012.

Nordback, K., Marshall, W. E., & Janson, B. N. (2013). Development of Estimation Methodology for Bicycle and Pedestrian Volumes Based on Existing Counts. Denver, CO: Colorado Department of Transportation (CDOT).

Nordback, K., Marshall, W. and B. Janson. Bicyclist safety performance functions for a U.S. city. In *Accident Analysis and Prevention* 65 (2014) 114–122.

Nordback, K, Marshall, W., Janson, B., and Stolz, E. Estimating Annual Average Daily Bicyclists Error and Accuracy. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2339, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 90–97

Nosal, T. and L.F. Miranda-Moreno. "Cycle-tracks, bicycle lanes & on-street cycling in Montreal: a preliminary comparison of the cyclist injury risk." Presented at the 91<sup>st</sup> Annual Meeting of the Transportation Research Board, January 22-26, Washington, D.C., 2012.



Nosal, T., Miranda-Moreno, L., and Z. Krstulic. Incorporating weather: a comparative analysis of Average Annual Daily 2 Bicyclist estimation methods. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2468, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 100–110.

O'Brien, S., Jackson, K., Searcy, S., Rodriguez, D. A., & Stolz, E. (2016). *Bicycle and Pedestrian Data Collection, Phase I Final Report*. Raleigh, NC: Institute for Transportation Research and Education. (North Carolina Department of Transportation)

Park, J., M. Abdel-Aty, J. Lee, and C. Lee. Developing Crash Modification Functions to Assess Safety Effects of Adding Bike Lanes for Urban Arterials with Different Roadway and Socio-Economic Characteristics. *Accident Analysis & Prevention*, Vol. 74, 2015, pp. 179-191.

*Pedestrian and Bicycle Crash Analysis Tool (PBCAT)*, FHWA.

Pendleton, J. (2016). DRCOG. In G. Gibson (Ed.) (Telephone communication ed.). Denver, Colorado.

Robinson, D.L.(2005). Safety in numbers in Australia: more walkers and bicyclists, safer walking and cycling. *Health Promotion Journal of Australia* 16, 47–51.

Rodgers, G.B. (1994). “The Characteristics and Use Patterns of Bicycle Riders in the United States,” *Journal of Safety Research*, 25, 83–96.

Rogers, William. (2019) “Pedestrian and Bicycle Safety Performance Functions for the Highway Safety Manual.” MRI Global. NCHRP 17-84. Research underway, expected completion date 3/26/2019.

Santos, A., McGuckin, N., Nakamoto, H., Gray, D., and S. Liss. Summary of Travel Trends. 2009 National Household Travel Survey. Report FHWA-PL-11-022, Federal Highway Administration.

Schepers, P., Does more cycling also reduce the risk of single-bicycle crashes? *Injury Prevention*, 2011.

Schepers, P., Hagenzieker, M., Methorst, R., and B. van Wee. 2013. A conceptual framework for road safety and mobility applied to cycling safety. *Accident Analysis and Prevention*, Vol. 62, pp. 331-340.

Schepers, J.P., Kroeze, P.A., Smeets, W., Wüst, J.C., 2011. Road factors and bicycle-motor vehicle crashes at unsignalized priority intersections. *Accident Analysis & Prevention* 43 (3), 853–861.

Schneider, Robert, Arnold, Lindsay, Ragland, David, (2009.) Pilot Model for Estimating Pedestrian Intersection Crossing Volumes. *Transportation Research Record: Journal of the Transportation Research Board*, 2140, pp 13-26.

Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics*, Vol 6(2), 461-464.

Sciortino, S., Vassar, M., Radetsky, M., and M.M. Knudson. San Francisco Pedestrian Injury Surveillance: Mapping, Under-Reporting, and Injury Severity in Police and Hospital Records. *Accident Analysis and Prevention*, Vol 35, 2005, pp 1102-1113.

Smeed, R.J. Some Statistical Aspects of Road Safety Research. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 112, No. 1, 1949, pp. 1-34.

Strauss, J., Miranda-Moreno, L., and P. Morency. 2013. A Bayesian Modeling Approach for Cyclist Injury Risk Analysis at Intersections and Corridors. *Accident Analysis and Prevention*, Vol. 59, pp. 9-17.

Strauss, J., Miranda-Moreno, L., and P. Morency. Mapping Cyclist Activity and Injury Risk in a Network Combining Smartphone GPS Data and Bicycle Counts. *Accident Analysis and Prevention*, Vol. 83, 3026, pp. 132-142.

Stutts, J., and W. Hunter. Police Reporting of Pedestrians and Bicyclists Treated in Hospital Emergency Rooms. *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, Washington D.C., 1635, pp 88-92, 1998.

Thomas, B., DeRobertis, M. 1995. The Safety of Urban Cycle Tracks: A Review of the Literature. *Accident Analysis and Prevention*.

Thomas, B., Lan, B., Sanders, R.L., Frackelton, A., Gardner, S., and M. Hintze. Predicting Safety: Development and Application of Bicycle Safety Performance Functions in Seattle, WA. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2659, 2017, pp. 212–223.

Transportation Research Board. *Development of a Bicycle Safety Prediction Methodology: Research Needs Statement*. Transportation Research Board of the National Academies, Washington D.C., 2010.

Turner, S., Wood, G., Hughes, T., Singh, R. Safety performance functions for bicycle crashes in New Zealand and Australia. *Transportation Research Record: Journal of the Transportation Research Board* No 2236, 66–73, 2011.

U.S. Census Bureau, 2018 American Community Survey 1-Year Estimates.

U.S. Census Bureau, QuickFacts. 2010, U.S. Department of Commerce.

U.S. Census Bureau. (2016). "American Factfinder." American Community Survey.

United States Department of Transportation, (USDOT). American Community Survey, 2009. U.S. Census Bureau.

U.S. Census Bureau (1994). *Geographic Areas Reference Manual, Chapter 12: The Urban and Rural Classifications*: US Department of Commerce.)

United States Department of Transportation, (USDOT). Safer People, Safer Streets: Summary of U.S. Department of Transportation Action Plan to Increase Walking and Biking and Reduce Pedestrian and Bicyclist Fatalities, 2014.

Vanparijs, J., Panis, L., Meeusen, R., and B. de Geus. Exposure Measurement in Bicycle Safety Analysis: A Review of the Literature. *Accident Analysis and Prevention*, Vol 84, 2015, pp 9-19.

## **APPENDIX A: DATA PREPARATION METHODS**

### **Shapefiles at large**

This project was largely assembled within ArcMap and exported to Microsoft Excel for further analysis and data assembly. The original files for streets, bicycle facilities, land use, zip codes, and boundaries were acquired from the City of Boulder. Crash data points were acquired from the Denver Regional Council of Governments (more on that below). All other shapefiles/layers were assembled in house using existing data.

### **ArcMap**

ESRI's ArcMap was used to put together all the maps and database files found below. The master ArcMap file is found on both the Dropbox and PSU's I:Drive.

### **Road Segments**

Road segments were originally created based on the intersectional points of Boulder's manual traffic count data. One manual count form would account for 3 or 4 directions (east, west, north, and south). The road segments were created in ArcMap via the following method:

1. Find the intersection using Boulders' PDF count sheets.
2. Clip out the relevant road segments leading out of the intersection point.
3. When all intersections were completed, they were merged into a single shapefile.

Due to Boulder's original road segments being broken at each individual street cross section, we opted to expand our own street segments in order to create longer distances from count point to count point. This was done to account for the fact that bicyclists likely traveled along the same route for longer distances. The following explains our methodology for road segment length. A road segment was merged with other sections until:

- It hit another count location.
- It hit either a stop signal or an arterial.
- It hit three small blocks.
- It ended.

In some cases, such as highways that continued on for long stretches without any original breaks, we performed no manipulation.

### **Crash points**

Crash points were given to us from the Denver Regional Council of Governments and include crashes from 2004 – 2013 (later years were not available at the time of this writing). The crash data files had to first be combined (since they came in individual year formats). From there, all non-bicycle-motorist crashes were pulled out. All intersection crashes were pulled out. And, finally, all fatal crashes were removed. With lat/long data we were able to map the crash points into ArcMap. Unfortunately, despite all crashes being non-intersection, many points were still situated within an intersection. After consulting with DRCOG the following method was used to map these crashes to road segments:

1. First vehicle of travel was noted in the records.

2. According to DRCOG, “first vehicle of travel” was the primary vehicle causing the incident.
3. Therefore, a vehicle traveling southbound would logically sit north of the intersection.
4. The crash was then mapped to the road segment north of where the crash point lay in the intersection.

If an intersection point was mapped on a road segment, that road segment was then used and superseded the above method.

### **Bicycle and automobile volume mapping**

Volumes were mapped using both the road segments and the PDFs. First, the volume data were entered into a Microsoft Excel sheet. Each directional road volume was created by adding up the various directions corresponding with the road. For example, a road segment north of the intersection was created by adding up the various volumes of the:

- Southbound thru
- Southbound u-turn
- Northbound thru
- Eastbound right turn
- Westbound left turn

Each road direction was given its own ID number and then each intersection was given its own ID number. After that the road segment shapefile in ArcMap was exported. Road segment IDs from ArcMap were matched up with the ID numbers from the spreadsheet. The new file was then added to ArcMap, and joined with the road segment layer based on segment ID. This gave us our volume mapping along road segments. Note that bicycles in crosswalks were not used. Only bicycles on the roadway were used in this volume mapping and count.

### **Ped Volumes**

Ped volumes were constructed in a similar manner as the bike and automobile maps, but were not mapped. The data was simply used as a reference point within the data as no factors were ever created to extrapolate the manual counts into annual averages. The data does exist within the master statistics spreadsheet, however, instead of as a number, the data was broken into “low, medium, or high” amounts based on the Natural Breaks (Jenks) methodology. Those breaks are as follows:

- Low: < 145
- Medium: > 512 , < 1246
- High: > 1246

### **Land use, demographic data, and bicycle facilities along road segments**

For our report, we decided to use the following variables to compare the volume and crashes to:

- Land use/zones: City of Boulder. (2016). Open Data Catalog (Publication.: <https://bouldercolorado.gov/open-data>)
- Population density: ACS
- Employment density: ACS

- Educational attainment: ACS
- Bicycle facilities: City of Boulder

In mapping land use, demographic, and bicycle facility data we used the following method:

1. Created a 500ft buffer around all road segments (50ft for bicycle facilities).
2. Loaded the land use shapefile into the document.
3. Intersected the segment buffer and the land use shapefile with the buffer on top.
4. Exported the results into an excel file.
5. Created a pivot table of the data in Microsoft Excel.
6. Used SiteID and land use/demographic/bicycle facilities area/length to compute

The results were then combined with the volume and crash data within a single spreadsheet using the original road segment ID numbers as the matching point to keep everything consistent. For population density, employment density, and educational attainment levels the quantitative data was converted to a “low, medium, or high” scale based on the Natural Breaks (Jenks) method within ArcMap. These data were obtained from the American Community Survey (*U.S. Census Bureau. (2016). "American Factfinder." American Community Survey.*) Those breaks are as follows:

#### Population Density (per square mile)

- Low: < 4,700
  - 42 census tracts in this group. (intersecting Boulder city footprint)
  - Median is 1,560 people per square mile
- Medium: > 4700, < 12,000
  - 37 census tracts this group. (intersecting Boulder city footprint)
  - Median is 7,570 people per square mile.
- High: > 12,000
  - Highest in GIS is 24,870 people per square mile.
  - 8 Census Tracts in Boulder are in the High category.

#### Employment Density (per square mile)

- Low: < 2,700
- Medium: > 2,700, < 6,900
- High: > 6,900

#### Educational Attainment Levels (per square mile)

##### No High School

- Low: < 3,000
- Medium: > 3,000, < 10,000
- High: > 10,000

##### High School – Some College

- Low: < 900

- Medium: > 900, < 2,900
- High: > 2,900

College Degree or above

- Low: < 1,200
- Medium: > 1,200, < 3,300
- High: > 3,300

Bicycle facilities are simply in length based on feet. Note, however, that a bicycle facility's length may be longer than that of the road length. This is likely due to the fact that bike facilities could be counted twice (northbound/southbound for example) or due to variations in a trail zigzagging near a road segment. Direct comparisons from bicycle facility length to road length are challenging due to this.

## APPENDIX B: DATA DICTIONARY

**GenID** – General individual ID number for each row. No significant impact on data.

**IntersectionID** – ID number assigned to each intersection from 2 hour volume counts. Use to easily sort out which crashes and volumes are associated with each intersection. Specific intersections can be found in the latest version of the “Boulder Crash and Volume Master Spreadsheet” file under ‘volume.’

**GIS\_ID** - ID number created within ArcMap that aligns with the GIS road segments. Use this ID number to join with the road segment “OBJECTID” number.

**AADT** – Annual average daily automobile traffic.

**AADB** – Annual average daily bicycle traffic.

**StreetID** – Duplicate of GIS\_ID for redundancy. Ignore or remove in your own version of this file. Do not remove in the master version of this file.

**LandUse\_Industrial\_Percent** – Percent of the area surrounding the road segment that is industrial based on a 500ft buffer.

**LandUse\_Residential\_Percent** – Percent of the area surrounding the road segment that is residential based on a 500ft buffer.

**LandUse\_Retail\_Percent** – Percent of the area surrounding the road segment that is retail based on a 500ft buffer.

**LandUse\_University\_Percent** – Percent of the area surrounding the road segment that is university based on a 500ft buffer.

**LandUse\_Industrial\_Area** – Square foot area surrounding the road segment that is industrial based on a 500ft buffer.

**LandUse\_Residential\_Area** – Square foot area surrounding the road segment that is residential based on a 500ft buffer.

**LandUse\_Retail\_Area** – Square foot area surrounding the road segment that is retail based on a 500ft buffer.

**LandUse\_University\_Area** – Square foot area surrounding the road segment that is university based on a 500ft buffer.

**Pop\_Density\_High\_Percent** – Percent of the area surrounding the road segment that has a “high” percentage of density based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Pop\_Density\_Low\_Percent** – Percent of the area surrounding the road segment that has a “low” percentage of density based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Pop\_Density\_Medium\_Percent** – Percent of the area surrounding the road segment that has a “medium” percentage of density based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Pop\_Density\_High\_Area** – Square footage of the area surrounding the road segment that has a “high” percentage of density based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Pop\_Density\_Low\_Area** – Square footage of the area surrounding the road segment that has a “low” percentage of density based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.



**Pop\_Density\_Medium\_Area** – Square footage of the area surrounding the road segment that has a “medium” percentage of density based on Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Emp\_Density\_High\_Percent** – Percent of the area surrounding the road segment that has a “high” percentage of employment density based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Emp\_Density\_Low\_Percent** – Percent of the area surrounding the road segment that has a “low” percentage of employment density based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Emp\_Density\_Medium\_Percent** – Percent of the area surrounding the road segment that has a “medium” percentage of employment density based on Breaks (Jenks) breaks created in ArcMap based on a 500ft buffer.

**Emp\_Density\_High\_Area** – Square footage of the area surrounding the road segment that has a “high” percentage of employment density based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Emp\_Density\_Low\_Area** – Square footage of the area surrounding the road segment that has a “low” percentage of employment density based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Emp\_Density\_Medium\_Area** – Square footage of the area surrounding the road segment that has a “medium” percentage of employment density based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_No\_High\_High\_Percent** – Percent of the area surrounding the road segment that has a “high” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_No\_High\_Low\_Percent** – Percent of the area surrounding the road segment that has a “low” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_No\_High\_Med\_Percent** – Percent of the area surrounding the road segment that has a “medium” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_No\_High\_High\_Area** – Square footage of the area surrounding the road segment that has a “high” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_No\_High\_Low\_Area** – Square footage of the area surrounding the road segment that has a “low” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_No\_High\_Med\_Area** – Square footage of the area surrounding the road segment that has a “medium” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_High\_High\_Percent** – Percent of the area surrounding the road segment that has a “high” percentage of high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_High\_Low\_Percent** – Percent of the area surrounding the road segment that has a “low” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_High\_Med\_Pecent** – Percent of the area surrounding the road segment that has a “medium” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_High\_High\_Area** – Square footage of the area surrounding the road segment that has a “high” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_High\_Low\_Area** – Square footage of the area surrounding the road segment that has a “high” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_High\_Med\_Areat** – Square footage of the area surrounding the road segment that has a “high” percentage of no high school diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_Coll\_High\_Percent** – Percent of the area surrounding the road segment that has a “high” percentage of bachelor’s degree or higher diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_Coll\_Low\_Percent** – Percent of the area surrounding the road segment that has a “low” percentage of bachelor’s degree or higher diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_Coll\_Med\_Percent** – Percent of the area surrounding the road segment that has a “medium” percentage of bachelor’s degree or higher diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_Coll\_High\_Area** – Square footage of the area surrounding the road segment that has a “high” percentage of bachelor’s degree or higher diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_Coll\_Low\_Area** – Square footage of the area surrounding the road segment that has a “low” percentage of bachelor’s degree or higher diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**Edu\_Coll\_Med\_Area** – Square footage of the area surrounding the road segment that has a “medium” percentage of bachelor’s degree or higher diploma attainment based on Natural Breaks (Jenks) created in ArcMap based on a 500ft buffer.

**StreetID2** – Duplicate of GIS\_ID for redundancy. Ignore or remove in your own version of this file. Do not remove in the master version of this file.

**Bike\_Street\_ID** – Duplicate of GIS\_ID for redundancy. Ignore or remove in your own version of this file. Do not remove in the master version of this file.

**Bike\_Lane\_Percent** – Percent of the road segment that has a “bike lane” as compared to other bicycle facilities. Does not measure how much of the road segment the facility has.

**Multi\_Purpose\_Sidewalk\_Percent** – Percent of the road segment that has a “multi-purpose sidewalk” as compared to other bicycle facilities. Does not measure how much of the road segment the facility has.

**Neighborhood\_Trail\_Percent** – Percent of the road segment that has a “neighborhood trail” as compared to other bicycle facilities. Does not measure how much of the road segment the facility has.

**Path\_Cul\_De\_Sac\_Links\_Percent** – Percent of the road segment that has a “cul de sac path link” as compared to other bicycle facilities. Does not measure how much of the road segment the facility has.

**Paved\_Trail\_Percent** – Percent of the road segment that has a “paved trail” as compared to other bicycle facilities. Does not measure how much of the road segment the facility has.

**Paved\_Trail\_Next\_to\_Road\_Percent** – Percent of the road segment that has a “paved trail next to the road” as compared to other bicycle facilities. Does not measure how much of the road segment the facility has.

**Protected\_Bicycle\_Lane\_Percent** – Percent of the road segment that has a “protected bicycle lane” as compared to other bicycle facilities. Does not measure how much of the road segment the facility has.

**Separated\_Sidewalk\_Percent** – Percent of the road segment that has a “separated sidewalk” as compared to other bicycle facilities. Does not measure how much of the road segment the facility has.

**Shared\_Lane\_Bicycle\_Route\_Percent** – Percent of the road segment that has a “shared lane bicycle route” as compared to other bicycle facilities. Does not measure how much of the road segment the facility has.

**Unpaved\_Trail\_Percent** – Percent of the road segment that has a “unpaved trail” as compared to other bicycle facilities. Does not measure how much of the road segment the facility has.

**Bike\_Lane\_Length** – Length in feet of the “bike lane” as compared to other bicycle facilities. Not compared to full road segment length.

**Multi\_Purpose\_Sidewalk\_Length** – Length in feet of the “multi-purpose sidewalk” as compared to other bicycle facilities. Not compared to full road segment length.

**Neighborhood\_Trail\_Length** – Length in feet of the “neighborhood trail” as compared to other bicycle facilities. Not compared to full road segment length.

**Path\_Cul\_De\_Sac\_Links\_Length** – Length in feet of the “cul-de-sac path link” as compared to other bicycle facilities. Not compared to full road segment length.

**Paved\_Trail\_Length** – Length in feet of the “paved trail” as compared to other bicycle facilities. Not compared to full road segment length.

**Paved\_Trail\_Next\_to\_Road\_Length** – Length in feet of the “paved trail next to road” as compared to other bicycle facilities. Not compared to full road segment length.

**Protected\_Bicycle\_Lane\_Length** – Length in feet of the “protected bicycle lane” as compared to other bicycle facilities. Not compared to full road segment length.

**Separated\_Sidewalk\_Length** – Length in feet of the “separated sidewalk” as compared to other bicycle facilities. Not compared to full road segment length.

**Shared\_Lane\_Bicycle\_Route\_Length** – Length in feet of the “shared lane bicycle route” as compared to other bicycle facilities. Not compared to full road segment length.

**Unpaved\_Trail\_Length** – Length in feet of the “unpaved trail” as compared to other bicycle facilities. Not compared to full road segment length.

**Crash\_ID** – Individual ID number given to each crash.

**STREETID3** – Duplicate of GIS\_ID for redundancy. Ignore or remove in your own version of this file. Do not remove in the master version of this file.

**CRASH\_COUNT** – Individual crash counts. Multiples can be linked to a single GIS\_ID number.

**CRASH\_NUMBER\_SUMMED** – Individual crash counts summed up in a single row. From DRCOG database, 8 years of data 2006 to 2013 non-intersection and not intersection related crashes, motor vehicle bicycle nonfatal crashes only.

**Crash\_Location** – Direction from the intersection where the crash occurred based on the DIR1 field.

**Crash\_Street** – Street where the crash occurred.

**LOC\_01** – Primary location of the first vehicle in crash.

**LOC\_02** – Location of the second vehicle in crash.

**LOCATION** – Location of the crash.

1. On Roadway
2. Ran Off Left Side
3. Ran Off Right Side
4. Ran Off 'T' Intersection
5. Vehicle crossed center median into opposing lanes
6. On private property

**DIR1** – Direction of travel for first vehicle

1. North
2. Northeast
3. East
4. Southeast
5. South
6. Southwest
7. West
8. Northwest

**DIR2** – Direction of travel for second vehicle

1. North
2. Northeast
3. East
4. Southeast
5. South
6. Southwest
7. West
8. Northwest

**DIR3** – Direction of travel for third vehicle

1. North
2. Northeast
3. East
4. Southeast
5. South
6. Southwest
7. West
8. Northwest

**VM1** – First vehicle movement prior to impact

1. Going straight
2. Slowing
3. Stopped in traffic
4. Making a right turn
5. Making a left turn

6. Making a u-turn
7. Passing
8. Backing
9. Entering/leaving parked position
10. Parked
11. Changing lanes
12. Avoiding object in roadway
13. Weaving
14. Spun out of control
15. Drove wrong way
16. Other

**VM2** – Second vehicle movement prior to impact

1. Going straight
2. Slowing
3. Stopped in traffic
4. Making a right turn
5. Making a left turn
6. Making a u-turn
7. Passing
8. Backing
9. Entering/leaving parked position
10. Parked
11. Changing lanes
12. Avoiding object in roadway
13. Weaving
14. Spun out of control
15. Drove wrong way
16. Other

**VM3** – Third vehicle movement prior to impact

1. Going straight
2. Slowing
3. Stopped in traffic
4. Making a right turn
5. Making a left turn
6. Making a u-turn
7. Passing
8. Backing
9. Entering/leaving parked position
10. Parked
11. Changing lanes
12. Avoiding object in roadway
13. Weaving
14. Spun out of control
15. Drove wrong way

16. Other

**KILLED** – Number killed

**INJURED** – Number injured

**VEHICLES** – Number of vehicles involved? Can't find specific info

**INJLEVEL\_0** – Severity of injury

0. No injury
1. Complaint of injury
2. Evident – non incapacitating
3. Evident – incapacitating
4. Fatal

**INJLEVEL\_1** – Severity of injury

0. No injury
1. Complaint of injury
2. Evident – non incapacitating
3. Evident – incapacitating
4. Fatal

**INJLEVEL\_2** – Severity of injury

0. No injury
1. Complaint of injury
2. Evident – non incapacitating
3. Evident – incapacitating
4. Fatal

**INJLEVEL\_3** – Severity of injury

0. No injury
1. Complaint of injury
2. Evident – non incapacitating
3. Evident – incapacitating
4. Fatal

**INJLEVEL\_4** – Severity of injury

0. No injury
1. Complaint of injury
2. Evident – non incapacitating
3. Evident – incapacitating
4. Fatal

**MONTH\_**

**DAY\_**

**YEAR\_**

**HOUR\_**

**ROAD** – Description of the road

1. At intersection
2. Driveway access related
3. Intersection related

4. Non-intersection
5. Alley related
6. Roundabout
7. Highway interchange
8. Parking lot

**CONTOUR** – Contour of the road

1. Straight On-level
2. Straight On-grade
3. Curve On-level
4. Curve On-grade
5. Hillcrest

**SURFACE** – Road surface type

1. Concrete
2. Blacktop
3. Brick or block
4. Gravel, slag or stone
5. Dirt
6. Other
7. Unknown

**CONDITION** - Condition of the road

1. Dry
2. Wet
3. Muddy
4. Snowy
5. Icy
6. Slushy
7. Foreign Material
8. Dry w/ visible icy road treatment
9. Wet w/ visible icy road treatment
10. Snowy w/ visible icy road treatment
11. Icy w/ visible icy road treatment
12. Slushy w/ visible icy road treatment

**LIGHTING** – Lighting during time of crash

1. Daylight
2. Dawn or dusk
3. Dark – lighted
4. Dark – Unlighted

**WEATHER** – Weather during time of crash

0. None
1. Rain
2. Snow/sleet/hail
3. Fog
4. Dust

5. Wind

**VT1** – Vehicle type? Can't find specifics.

**VT2**

**VT3**

**SPEED1** – Estimated speed limit for vehicles only

**SPEED2** – Estimated speed limit for vehicles only

**SPEED3** – Estimated speed limit for vehicles only

**LATITUDE**

**LONGITUDE**

**DRV\_ACT1** – Action of first driver (officer opinion)

0. No action
1. Exceeded safe/posted speed
2. Impeded traffic
3. Failed to yield ROW
4. Disregard stop sign
5. Failed to stop at signal
6. Disregarded other device
7. Improper turn
8. Turned from wrong lane or position
9. Other improper turns
10. Lane violation
11. Improper passing on left
12. Improper passing on right
13. Followed too closely
14. Improper backing
15. Signaling violation
16. Reckless driving
17. Careless driving

**DRV\_ACT2** – Action of second driver (officer opinion)

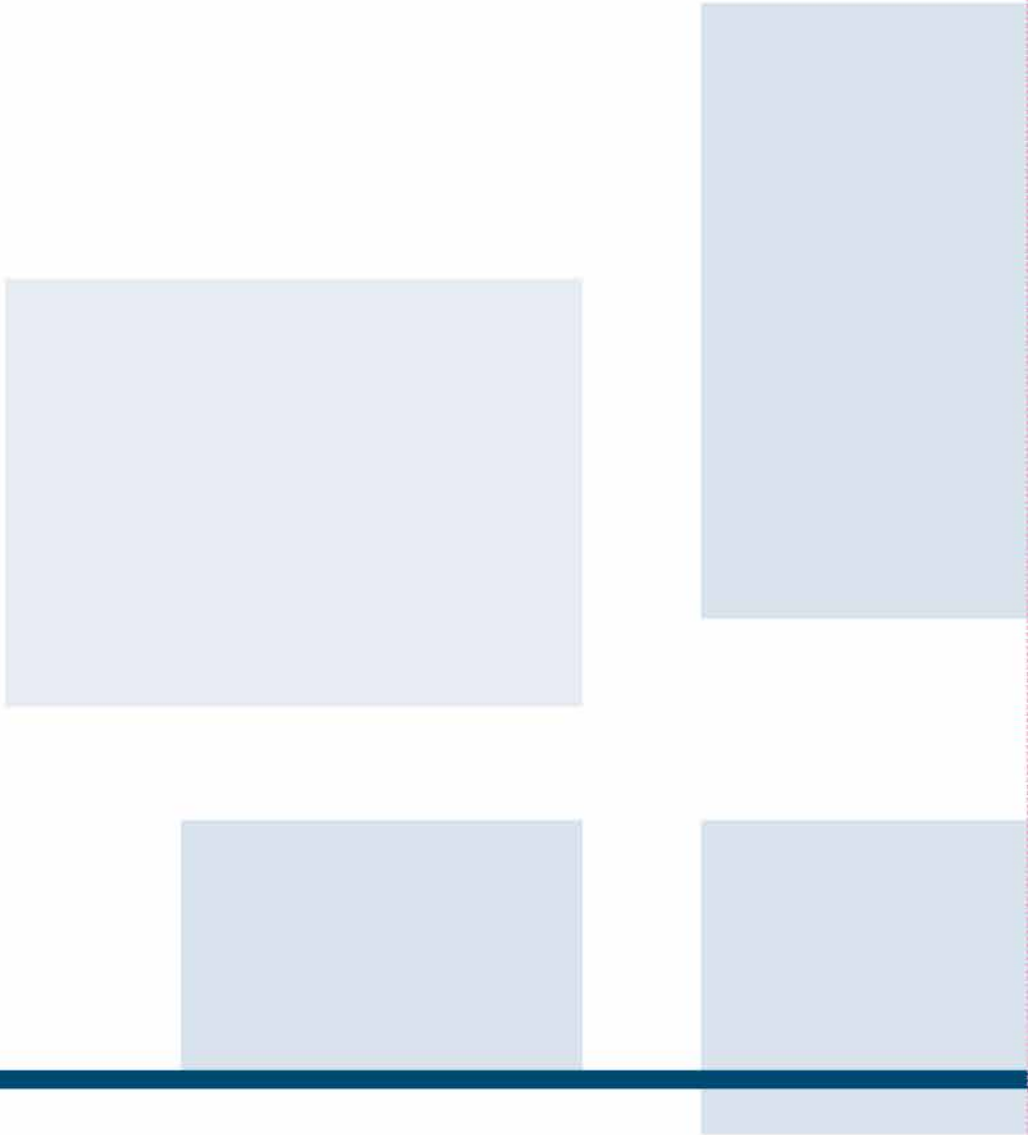
0. No action
1. Exceeded safe/posted speed
2. Impeded traffic
3. Failed to yield ROW
4. Disregard stop sign
5. Failed to stop at signal
6. Disregarded other device
7. Improper turn
8. Turned from wrong lane or position
9. Other improper turns
10. Lane violation
11. Improper passing on left
12. Improper passing on right
13. Followed too closely
14. Improper backing
15. Signaling violation



16. Reckless driving
17. Careless driving

**DRV\_ACT3** – Action of third driver (officer opinion)

0. No action
1. Exceeded safe/posted speed
2. Impeded traffic
3. Failed to yield ROW
4. Disregard stop sign
5. Failed to stop at signal
6. Disregarded other device
7. Improper turn
8. Turned from wrong lane or position
9. Other improper turns
10. Lane violation
11. Improper passing on left
12. Improper passing on right
13. Followed too closely
14. Improper backing
15. Signaling violation
16. Reckless driving
17. Careless driving



NATIONAL INSTITUTE for  
TRANSPORTATION and COMMUNITIES  
Transportation Research and Education Center  
Portland State University  
1900 S.W. Fourth Ave., Suite 175  
Portland, OR 97201