DEVELOPING A PEDESTRIAN SAFETY PERFORMANCE FUNCTION FOR OREGON

Final Report

SPR 841

RESEARCH

Oregon Department of Transportation

DEVELOPING A PEDESTRIAN SAFETY PERFORMANCE FUNCTION FOR OREGON

SPR 841 PHASE II

by

Josh Roll (Oregon Department of Transportation Research Unit)
Jason Anderson (Portland State University)
Nathan McNeil (Portland State University)

for

Oregon Department of Transportation
Research Section
555 13th Street NE, Suite 1
Salem OR 97301

and

Federal Highway Administration
1200 New Jersey Avenue SE
Washington, DC 20590

February 2026

Technical Report Documentation Page

1. Report No. FHWA-OR-RD-26-06	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Developing a Pedestrian Safety Performance Function for Oregon		5. Report Date January – 2026 6. Performing Organization Code
7. Author(s) Josh Roll – Oregon Department of Transportation Jason Anderson (Ph.d) – Portland State University Nathan McNeil – Portland State University		8. Performing Organization Report No. SPR 841
9. Performing Organization Name and Address Oregon Department of Transportation Research Section 555 13th Street NE, Suite 1 Salem, OR 97301		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 16292 20-05
12. Sponsoring Agency Name and Address Oregon Dept. of Transportation Research Section 555 13th Street NE, Suite 1 Salem, OR 97301	Federal Highway Admin. 1200 New Jersey Avenue SE Washington, DC 20590	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This report details the development of a pedestrian safety performance function (SPF) for intersections in urban areas in the state of Oregon. The report documents the process to develop SPFs for pedestrian crashes including the process to develop the data for analysis. A key innovation of this work is the use of pedestrian traffic volume estimates from traffic signals push button actuations. These data and other pedestrian traffic volume data collected through traditional methods are utilized in a machine learning data fusion model to estimate pedestrian volumes for all intersections in Oregon urban areas. SPFs are constructed for various intersection types with different traffic control. Crash prediction models using estimates of pedestrian volumes and also proxies for pedestrian volume are compared to give agencies a sense of how models perform with better pedestrian volume data.

17. Key Words	18. Distribution Statement
Safety performance function, SPF, pedestrian volume, AADPT, data fusion, crash prediction, HPMS, Inrix, intersection, direct demand,	Copies available from NTIS, and online at www.oregon.gov/ODOT/TD/TP_RES/

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	210	

Technical Report Form DOT F 1700.7 (8-72) Reproduction of completed page authorized Printed on recycled paper

SI* (Modern Metric) Conversion Factors
Approximate Conversions to SI Units

Physical Quantity	Symbol	When You Know	Multiply By	To Find	Symbol
Length	n	inches	25.4	millimeters	mm
Length	ft	feet	0.305	meters	m
Length	yd	yards	0.914	meters	m
Length	mi	miles	1.61	kilometers	km
Area	in2	square inches	645.2	square millimeters	mm2
Area	ft2	square feet	0.093	square meters	m2
Area	yd2	square yard	0.836	square meters	m2
Area	ac	acres	0.405	hectares	ha
Area	mi2	square miles	2.59	square kilometers	km2
Volume	fl oz	fluid ounces	29.57	milliliters	mL
Volume	gal	gallons	3.785	liters **	L
Volume	ft3	cubic feet	0.028	cubic meters	m3
Volume	yd3	cubic yards	0.765	cubic meters	m3
Mass	oz	ounces	28.35	grams	g
Mass	lb	pounds	0.454	kilograms	kg
Mass	T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
Temperature (exact degrees)	oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	oC
Illumination	fc	foot-candles	10.76	lux	lx
Illumination	fl	foot-Lamberts	3.426	candela/m2	cd/m2
Force and Pressure or Stress	lbf	poundforce	4.45	newtons	N
Force and Pressure or Stress	lbf/in2	poundforce per square inch	6.89	kilopascals	kPa

*SI is the symbol for the International System of Measurement
** Volumes greater than 1000 L shall be shown in m^3

SI* (Modern Metric) Conversion Factors
Approximate Conversions from SI Units

Physical Quantity	Symbol	When You Know	Multiply By	To Find	Symbol
Length	mm	millimeters	0.039	inches	in
Length	m	meters	3.28	feet	ft
Length	m	meters	1.09	yards	yd
Length	km	kilometers	0.621	miles	mi
Area	mm2	square millimeters	0.0016	square inches	in2
Area	m2	square meters	10.764	square feet	ft2
Area	m2	square meters	1.195	square yards	yd2
Area	ha	hectares	2.47	acres	ac
Area	km2	square kilometers	0.386	square miles	mi2
Volume	mL	milliliters	0.034	fluid ounces	fl oz
Volume	L	liters	0.264	gallons	gal
Volume	m3	cubic meters	35.314	cubic feet	ft3
Volume	m3	cubic meters	1.307	cubic yards	yd3
Mass	g	grams	0.035	ounces	oz
Mass	kg	kilograms	2.202	pounds	lb
Mass	Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
Temperature (exact degrees)	oC	Celsius	1.8C+32	Fahrenheit	oF
Illumination	lx	lux	0.0929	foot-candles	fc
Illumination	cd/m2	candela/m2	0.2919	foot-Lamberts	fl
Force and Pressure or Stress	N	newtons	0.225	poundforce	lbf
Force and Pressure or Stress	kPa	kilopascals	0.145	poundforce per square inch	lbf/in2

For More Information see: https://www.fhwa.dot.gov/publications/convtabl.cfm

ACKNOWLEDGEMENTS

The authors would like to thank the Technical Advisory Committee including Jessica Horning, ODOT's Bike and Pedestrian Program Manager, for their expert guidance and direction throughout this research effort. The research team would also like to thank supportive ODOT staff that keep pushing the agency forward towards data driven analysis to improve traffic safety for all users of Oregon roads.

DISCLAIMER

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

The need for systemic understanding of safety on our transportation networks is crucial to properly intervening at the highest risk locations. Pedestrian injuries continue to rise in the United States and in Oregon so improving technical approaches to identifying high risk locations for interventions is essential for prioritization of scarce resources. This report documents the process of developing a high-quality safety performance function for pedestrian crash injuries in Oregon by harnessing available data on traffic operations, roadway geometry, built environmental characteristics and socio-demographics. To support the pedestrian safety performance functions and emerging practices of the 2nd edition of the Highway Safety Manual, this research develops measures of pedestrian volumes using novel data sources and an advanced data fusion modeling technique. And lastly, to support the work present on pedestrian SPFs, this report also presents a Background chapter that summarizes information about how other state departments of transportation are currently using SPFs. It is this research team's hope that the information contained in this report can help transportation authorities in Oregon more effectively retrofit areas of the transportation system that present the most risk for pedestrian injury.

2.0 BACKGROUND

2.1 SYSTEMIC SAFETY APPROACH

The systemic approach to transportation safety seeks to understand and act to reduce the risk of crashes, injuries and fatalities occurring based on characteristics of the roadway and environment. This systemic approach is in contrast to a hot spot approach that looks to identify where crashes have occurred in the past and mitigate risk at those locations. As described by Federal Highway Administration (FHWA) on their "The Systemic Approach to Safety" web page, "a system-based approach acknowledges crashes alone are not always sufficient to determine what countermeasures to implement, particularly on low volume local and rural roadways where crash densities are lower, and in many urban areas where there are conflicts between vehicles and vulnerable road users (pedestrians, bicyclists, and motorcyclists)" (FHWA, n.d.). Thomas et al elaborate that "a systemic approach is a data-driven, networkwide (or system-level) approach to identifying and treating high risk roadway features correlated with specific or severe crash types. Systemic approaches seek to not only address locations with prior crash occurrence but also those locations with similar roadway or environmental crash risk characteristics" (Thomas, Sandt, et al., 2018, p. 4).

Acknowledging that there is some randomness in where crashes occur, a systemic approach may be more effective at guiding resources for mitigation through countermeasures to reduce risk and therefore be proactive in focusing on the *next* crash, rather than the *last* crash. Further, when it comes to pedestrian crashes, poor data on exposure and the relative rarity of crashes (although of higher severity) translates into a higher likelihood that a hot spot approach would underestimate the risk at other locations. However, looking at the characteristics of the places where crashes have occurred can inform assessments of the types of infrastructure, traffic conditions and environment where crashes are likely to occur.

2.1.1 Highway Safety Manual

First released in 2010, and updated with a supplement in 2014, the *Highway Safety Manual* (HSM) was developed to be an authoritative guide on estimating safety and risk on roadways (AASHTO, 2010, 2014). The HSM distinguishes between "descriptive analyses," which were more traditionally used by transportation professionals to describe crash frequencies, rates, types and severities. In contrast, the HSM seeks to advance "quantitative predictive analyses," which "are used to calculate an expected number and severity of crashes at sites with similar geometric and operational characteristics for one or more of the following: existing conditions, future conditions, or roadway design alternatives" (AASHTO, 2014, pp. 1–2). The HSM notes that there is generally no single "cause" of a crash, but rather that "crashes are the result of a convergence of a series of events that are influenced by a number of contributing factors (time of day, driver attentiveness, speed, vehicle condition, road design, etc.)" (AASHTO, 2014 p. 3-6).

More recent efforts to improve on the efficacy of the HSM methods have sought to incorporate more detailed roadway, land use, sociodemographic, and exposure data to improve upon risk prediction, particularly for vulnerable road users for whom crash data alone provides less confidence in predictive value (Thomas, Sandt, et al., 2018; Kumfer et al., 2019).

Safety performance functions, or SPFs, are one of the key analytical concepts detailed in the HSM, and are defined as "regression equations that estimate the average crash frequency for a specific site type (with specified conditions) as a function of annual average daily traffic (AADT) and, in the case of roadway segments, the segment length (L). Base conditions are specified for each SPF and may include conditions such as lane width, presence or absence of lighting, presence of turn lanes, etc." (AASHTO, 2014 p. 3-17). They are discussed in more detail later in this document.

2.1.2 Alternative Approaches to Measuring Safety Performance

Aside from the HSM method, there are a number of other approaches to analyzing safety and implementing crash mitigation strategies.

The **Safety Priority Index System (SPIS)** is essentially a hot spot-based tool that calculates a score for each roadway segment based on crash frequency, crash rate, and crash severity. SPIS was developed by the Oregon Department of Transportation (ODOT) to evaluate highway approach applications and provide heightened scrutiny to those on segments with the greatest history of crash activity (FHWA Roadway Safety Data Program, n.d.; Oregon DOT, Safety and Operations, n.d.).

Crash trees are a method or tool used by agencies to visualize the distribution of crashes across various distinctions, perhaps starting first with rural vs urban, then divided vs undivided, then by factors such as number of lanes, intersection vs driveway vs segment, etc. While essentially relying on empirical descriptive data, the crash tree can also help identify the types of roadway conditions that may pose greater risk, or the particular types of roadways on which an analysis will focus. However, the method is dependent upon the branch structure and order chosen by the agency / analyst. Perhaps more importantly, crash trees generally do not take into account exposure and other factors influencing risk, such as miles of roadway (Federal Highway Administration, 2021). As such, crash trees may be useful as an exploratory tool, but should be tested and reviewed.

In introducing the **Road Safety Analysis Program (or RSAP),** NCHRP Report 492 described various tools and procedures that engineers use to weight cost of a project against the expected safety benefit in order to inform types and locations of roadside safety decisions (Mak & Sicking, 2003). The report notes that while the existing procedures were of "varying degrees of usefulness", navigating the various suggested or required analyses resulted in a "time-consuming and sometimes inaccurate mix of tools for making decisions about roadside safety" (Mak & Sicking, 2003, pp. 1–2). Thus, the Road Safety Analysis Program (RSAP) was designed to be an improved cost-effectiveness procedure, which has modules to predict estimated encroachment frequency, the likelihood that an encroachment will result in a crash, the expected severity, and finally the expected cost/benefit. An "Engineer's Manual", describing the analysis contained within the modules, and a "User's Manual," describing the interface developed for the program (Mak & Sicking, 2002, 2003), which was viewed as making the program much easier to use than earlier assessment programs (Ray et al., 2022). Both the RSAP and its predecessor the

"ROADSIDE" program has been included in the AASHTO Roadside Design Guide dating back to 1989. A 2022 update, called RSAPv3, sought to further improve the approach, updating the code to function as an (easily updateable via lookup tables) MS Excel macro-enabled workbook, and switching from a Monte Carlo method to a deterministic method to address "convergence issues" (Ray et al., 2022).

In preparing the 2022 RSAP update, a survey was sent to 2100 road safety professionals asking about if and how they use RSAP. Of those, 122 took the survey, and 65% of those said they never use RSAP; though those who had used the program had generally positive feedback about their experience (Ray et al., 2022).

The **United States Road Assessment Program (usRAP)** seeks to deploy geographic data to develop risk maps and star ratings for roadway segments using proprietary software (ViDA) in order to help agencies incorporate risk assessment into their decision making (FHWA Office of Safety, 2016; usRAP, n.d.-b). The program method was developed by the AAA Foundation for Traffic Safety (AAAFTS). The software, available to highway agencies, has been used by at least eight states (usRAP, n.d.-a). Pilot studies of usRAP were conducted in Florida, Iowa, Michigan and New Jersey (Phase I and II pilots) and in Illinois, Kentucky, New Mexico and Utah (Phase III pilots).

2.1.3 Pedestrians and Systemic Safety

The 2018 NCHRP Report 893 "Systemic Pedestrian Safety Analysis" (Thomas, Sandt, et al., 2018) and the accompanying contractors technical report (Thomas, Kumfer, et al., 2018) outline much of the state of the practice in applying the systemic safety approach to pedestrian safety. NCHRP Report 893, a guidebook to help US transportation agencies to implement a systemic approach to pedestrian safety, describes a seven-step process to do so (see Figure 2.1).

Figure 2.1 Steps in a systemic pedestrian safety analysis process" Figure 3 in (Thomas, Sandt, et al., 2018)

The accompanying contractor's technical report provides more insight into the state of systemic pedestrian safety application in the US today, including noting that the desire by state DOTs to employ more pedestrian systemic safety approaches was stronger than the actual application of such methods. The "Key Takeaways" section of chapter 1 "State of Systemic Pedestrian Safety Practices in the U.S." (Thomas, Kumfer, et al., 2018, pp. 13–15) summarizes some reasons for the gap between desire to use systemic methods and actual use. Those reasons include limited existing crash and facility data, analytical approaches that neglect to include key factors such as pedestrian exposure, limited funds to implement costly countermeasures, and in some cases, largely pre-determine outcomes – e.g., searching for a place to implement a pre-identified countermeasure rather than examining what type of countermeasure was needed in what context.

2.2 SAFETY PERFORMANCE FUNCTIONS (SPFS) IN THE HIGHWAY SAFETY MANUAL

As noted above, SPFs estimate average or expected crash frequency for a site based on AADT and segment length, given a set of base conditions. They are developed using regression techniques on "observed crash data collected over a number of years at sites with similar characteristics and covering a wide range of AADTs," and using the assumption of an overdispersion and therefore a negative binomial distribution (AASHTO, 2014 p. 3-18). Related and important in assessing the expected number of crashes at a given location, crash modification factors (or CMFs) are used to account for "geometric or geographic differences between the based conditions of the model and local conditions of the sit under consideration" (AASHTO, 2014 p. C-15). Thus, SPFs must be accompanied by an AADT and assessed for whether a CMF is needed in order to be applied.

SPFs are described in Chapter 3 (section 3.5.2) and in Section C.6.3 of the HSM as one of the key components of predictive method for estimated the expected crash frequency of a site, facility or network. Part C of the manual further details the predictive method, with separate chapter fur rural two-lane, two-way roads (Chapter 10), rural multilane roads (Chapter 11), and urban and suburban arterials (Chapter 12). Section 3.5.4 describes how SPFs should be calibrated to local conditions, with further detail in Appendix A of Part C. The process of deriving SPFs is described in Appendix 3B.

2.2.1 Calibration and development of local SPFs

The HSM notes that SPFs are developed for specific roadway segment and intersection types, and that although they sought to use the best available data for the SPFs presented in the Manual, variation would be expected from one place to another.

2.2.1.1 Key resources

Since the publication of the HSM in 2010, several key reports have been published focusing on the calibration or development of local SPFs. Among them are:

Srinivasan, Carter, and Bauer (2013): "Safety Performance Function Decision Guide: SPF Calibration vs SPF Development"

- "This guidebook is intended to provide guidance on whether an agency should calibrate the safety performance functions (SPFs) from the Highway Safety Manual (HSM) or develop jurisdiction-specific SPFs. The guidebook discusses the factors that need to be considered while making the decision. It is intended to be of use to practitioners at state and local agencies and to researchers" (2013, p. 6).

Chapter five of the report details the decision process an agency should undergo when deciding if they should calibrate SPFs or develop local-specific SPFs, which includes:

- "Step 1. Determine intended use of SPF
- Step 2. Determine facility type

- Step 3. Identify existing SPF
- Step 4. Consider sample size necessary for calibrating SPF
- Step 5. Consider roadway data necessary for calibrating SPF
- Step 6. Calibrate existing SPF
- Step 7. Assess quality of calibration factor
- Step 8. Consider statistical expertise necessary for developing SPF
- Step 9. Consider sample size necessary for developing SPF
- Step 10. Determine crash type to be addressed by SPF
- Step 11. Develop SPF" (2013, p. 6).

Srinivasan and Bauer (2013) "Safety Performance Function Development Guide: Developing Jurisdiction-Specific SPFs."

- "This is a "how-to" guidebook for states that are developing jurisdiction-specific safety performance functions (SPFs). The guidebook discusses the issues associated with the development of jurisdiction-specific SPFs and provides a step-by-step procedure that states can use to develop jurisdiction-specific SPFs" (2013, p. 5).

The step-by-step procedure described includes:

- "Step 1 – Determine use of SPF
- Step 2 – Identify facility type
- Step 3 – Compile necessary data
- Step 4 – Prepare and cleanup database
- Step 5 – Develop the SPF
- Step 6 – Develop the SPF for the base condition
- Step 7 – Develop CMFs for specific treatments
- Step 8 – Document the SPFs" (2013, pp. 5–6)

Lyon, Parsaud and Gross (2016): "The Calibrator—An SPF Calibration and Assessment Tool User Guide"

- "The focus of The Calibrator is to help users assess SPF compatibility and applicability. …This research developed a spreadsheet-based tool (i.e., The Calibrator) to assess compatibility and applicability of SPFs and CMFs for application in a different time or place….The audience for the tool and this user guide are road safety practitioners responsible for developing new SPFs or calibrating existing SPFs to data from their jurisdiction" (2016, p. 1).

Lyon et al. note that a number of factors may differ from place to place that would affect the utility of an SPF. These range from crash reporting practices, socio-demographic characteristics, weather, roadway maintenance, among other factors; therefore "it is recommended to replace any default crash distributions and adjustment factors, such as those found in the HSM, using jurisdiction-specific data for the same years as the jurisdiction-specific SPF calibration factors" (Lyon et al., 2016, p. 3). They outline a process for calibrating SPFs, which result in a calibration factor (or multiplier for the original SPF estimate) and a dispersion parameter. The process is described in Figure 1 of their report (Figure 2 Figure 2.2here).

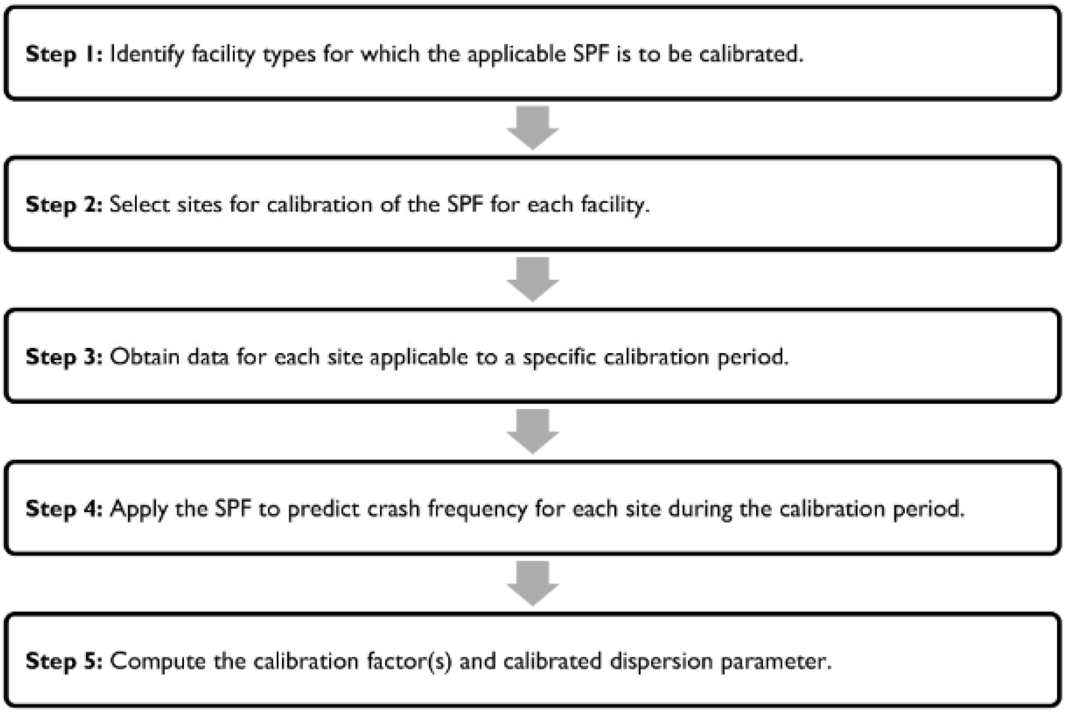


Figure 2.2 Calibration Process proposed in Lyon et al., 2016 (p. 4)

2.3 APPLICATIONS OF SPFS

The HSM outlines a set of key applications of SPFs, including:

- network screening;
- diagnosis;
- selecting countermeasures;
- economic appraisal;
- Prioritize projects; and
- safety effectiveness evaluation.

A 2018 report on best practices in the application of the HSM approach, including use of SPFs, by state DOTs (Milton et al., 2018) focuses on seven areas of interest, including: Status/Policy; Training; Technical Functions; Data; Cultural; Information Dissemination; and, Achieving Performance. The report pulls lessons from ten state departments of transportation which they state are leaders in HSM implementation, including DOTs from Alabama, Florida, Illinois, Louisiana, Maine, Michigan, Missouri, Ohio, Virginia, and Washington. Calibration is discussed in the "technical functions" chapter. The report notes that each of the ten state DOTs "developed calibration factors and crash distributions specific to their local conditions for use in network screening and/or the crash prediction models in HSM Part C" (Milton et al., 2018 p. 4-18).

Although calibrating was generally preferred to developing state-specific SPFs due to the interest in moving quickly to implement HSM methods, several states including IDOT, MDOT and WSDOT did developed state-specific SPFs. Milton et al. (2018) also note that a number of state DOTs, including Florida, Louisiana, Maine, Virginia, and Washington State "have taken a systemic approach to address pedestrian and bicycle crashes" (p 4-19), although it was not specified how exactly they did so.

The NCHRP contractors report discusses several available tools for applying systemic safety approaches to pedestrian safety (Thomas, Kumfer, et al., 2018, pp. 7–9). These include the FHWA Systemic Safety Project Selection Tool, Pedestrian and Bicyclist Intersection Safety Indices developed for FHWA by the UNC Highway Safety Resource Center, and NCHRP's Active Trans Priority Tool.

2.3.1 Scan of state DOT SPF implementation

In order to understand more about how state DOTs are implementing systemic safety methods generally, and safety performance functions more specifically, we conducted an online scan to assess their use. The scan included conducting a Google search for each of the 50 states using the following method for each state:

- Google: state name + "transportation systemic safety analysis SPF"
- Review 1st page of results for each state, looking for relevant website, reports, or other documents, that discuss how the state is implementing systemic safety methods.

We identified relevant documents for 37 of 50 states, which were most commonly agency reports or websites, but also included university research reports or resultant journal articles (e.g., the state DOT contracted with a university to conduct a study). For 12 states, we identified reports or articles describing their development of local SPFs, including Arkansas; Colorado; Idaho; Illinois; Kansas; Michigan; New Jersey; North Carolina; Ohio; Oklahoma; Pennsylvania; and Virginia. For 8 states, we identified reports describing local calibration efforts, including Idaho; Illinois; Mississippi; Missouri; New Jersey; Ohio; Tennessee; and Vermont. The states and documents identified in the search are listed in Table 2.1.

Table 2.1 State systemic safety online scan – local SPF development and calibration documents State/Region

State / Region	Document Title	Local SPF dvpt.	Local calibr.	Document Link
Arkansas	Safety Performance Functions for Arkansas	Yes		http://www.ahtd.state.ar.us/TRC/Final_Report/TRC1503_Safety_Performance_Functions_for_Arkansas.pdf
Colorado	Safety Analysis Information	Yes		https://www.codot.gov/safety/traffic-safety/assets/safety-analysis-information
Colorado	Evaluating Effectiveness of Crash Type SPFs in Safety Management: Comparing the Effectiveness of Network Screening and Diagnostic Methods Using Aggregate SPFs and Test of Proportion with Crash Type SPFs	Yes		https://www.codot.gov/programs/research/pdfs/2020-research-reports/cdot-2020-12.pdf
Idaho	Calibration and Development of Safety Performance Functions for Rural Highway Facilities in Idaho	Yes	Yes	https://apps.itd.idaho.gov/apps/research/Completed/RP225.pdf
Illinois	State-Specific Highway Safety Manual and Systemic Safety Analysis in Illinois: Roadway Safety Data and Analysis Case Study	Yes	Yes	https://safety.fhwa.dot.gov/rsdp/downloads/fhwasa17014.pdf
Kansas	A comparative study of newly developed Kansas-specific safety performance functions with HSM models for rural four-lane divided highway segments	Yes		https://www.tandfonline.com/doi/abs/10.1080/19439962.2019.1622614
Michigan	Safety Performance Functions for Rural Road Segments and Rural Intersections in Michigan	Yes		https://mdotjboss.state.mi.us/SpecProv/getDocumentById.htm?docGuid=f758a84e-6933-4c82-92f7-5e13caeb2e3d

State / Region	Document Title	Local SPF dvpt.	Local calibr.	Document Link
New Jersey	Calibration/Development of Safety Performance Functions for New Jersey	Yes	Yes	https://c2smart.engineering.nyu.edu/wp-content/uploads/2022/03/CalibrationDevelopment-of-Safety-Performance-Functions-for-New-Jersey-Final-Report.pdf
North Carolina	Development of Safety Performance Functions for North Carolina	Yes		https://connect.ncdot.gov/projects/research/RNAProjDocs/2010-09FinalReport.pdf
Ohio	Freeway Segment Safety Performance Function (SPF) Development	Yes		https://rosap.ntl.bts.gov/view/dot/58750/dot_58750_DS1.pdf
Oklahoma	Enhanced Safety Performance Function for Highway Segments in Oklahoma	Yes		https://ascelibrary.org/doi/full/10.1061/%28ASCE%29IS.1943-555X.0000616
Pennsylvania	Safety Performance Functions	Yes		https://www.mautc.psu.edu/docs/PSU-2013-09.pdf
Pennsylvania	Regionalized Safety Performance Functions	Yes		https://gis.penndot.gov/BPR_PDF_FILES/Documents/Research/Complete%20Projects/Operations/Regionalized_Safety_Performance.pdf
Virginia	Developing and Using State-Specific SPFs in Virginia	Yes		https://safety.fhwa.dot.gov/rsdp/downloads/va_case_study.pdf
Illinois	AASHTO Highway Safety Manual - Illinois User Guide with Illinois Calibration Factor and Default Values		Yes	https://idot.illinois.gov/Assets/uploads/files/Transportation-System/Memos-&-Letters/Safety/HSM_IL_UserGuide_11062014.pdf
Mississippi	Safety Performance Function Calibration for Mississippi		Yes	https://mdot.ms.gov/documents/Research/Reports/Interim%20&%20Final/MDOT_SPFCalibration_FinalReport_FINAL.pdf
Missouri	Missouri Highway Safety Manual Recalibration		Yes	https://spexternal.modot.mo.gov/sites/cm/CORDT/cmr18-001.pdf

State / Region	Document Title	Local SPF dvpt.	Local calibr.	Document Link
Ohio	Safety Analysis Guidelines		Yes	https://www.transportation.ohio.gov/programs/highway+safety/highway-safety-manual-guidance/safety-analysis-guidelines-cf
Tennessee	Highway Safety Manual Safety Performance Functions & Roadway Calibration Factors: Roadway Segments		Yes	https://www.tn.gov/content/dam/tn/tdot/long-range-planning/research/final-reports/res2016-final-reports/Final%20Report_RES2016-27%20Part%201%20Phase%20II_TDOT.pdf
Vermont	Calibration of the Highway Safety Manual Predictive Models for Rural Two-Lane Roads for Vermont		Yes	https://vtrans.vermont.gov/sites/aot/files/planning/documents/research/publishedreports/2020-01%20Calibration%20of%20Highway%20Safety%20Manual.pdf

Each state is required to submit an annual Highway Safety Improvement Program (HSIP) report. Within in that report is additional data that can be helpful in understand each state's use of HSM and systemic safety methods. In particular, the report template asks "Does the State use the Highway Safety Manual to support HSIP efforts?" and if so, "Please describe how the State uses the HSM to support HSIP efforts." A scan of each state's report from 2020 resealed that 46 out of 50 states indicated that they do use the HSM manual.

A Georgia DOT report entitled "Bicycle and Pedestrian Safety in the Highway Safety Manual" describes a survey conducted of people working on bicycle and pedestrian safety at state DOTs, MPOs, counties, cities, and consulting firms (Watkins et al., 2016). The survey, which garnered 238 total responses, including 133 completions, concluded that "Most agencies valued safety as a key component of their decision to implement infrastructure, but most did not collect enough exposure and crash data to adequately assess the safety impacts" (pg. viii).

2.4 ESTIMATION OF PEDESTRIAN VOLUMES

Since there are very few pedestrian counters, those interested in knowing about levels of pedestrian activity, an important factor for contextualizing crash and injury data, have had to rely on various methods of estimating pedestrian volumes, including the use of direct demand models. This section looks to the literature related to direct demand models to identify factors related to pedestrian volume.

NCHRP Report 770 "Estimating Bicycling and Walking for Planning and Project Development: A Guidebook" discussed factors affecting rates of walking and bicycling, looking at studies in the areas of land use and built environment, facilities, natural environment, sociodemographic factors, and attitudes and perceptions. As a companion to NCHRP Report 770, the NCHRP 08-36, Task 141 report titled "Evaluation of Walk and Bicycle Demand Modeling Practice" went into further detail on the state of the art and state of the practice related to modeling walk trips (RSG & The RAND Corporation, 2019). Although much of the report focused on walk trips (e.g., person-based) rather than volume at particular point, the discussion of relevant factors influencing walk trips is useful. As shown in Table 2.2, the report organized key factors into groups of "travel characteristics / socio-economic variables", "infrastructure / network / facilities", and "land-use variables."

Table 2.2: From NCHRP 08-136 Task 141 Evaluation of Walk and Bicycle Demand Modeling Practice: "Key factors influencing the decision to walk or cycle that have been incorporated in models"

Traveler characteristics / Socio-economic variables	Infrastructure / network / facilities	Land-use variables
Age	Distance	Household / employment density
Gender	Travel times and costs (for other modes)	Mix of uses
Work / student status	Directness	Transit stop density / distance to the nearest transit stop
Income	Slope / gradient / hilliness	Accessibility (e.g., attractions of a given type within a given distance)
Vehicle ownership / household competition for vehicles / car availability	Traffic volumes / lanes / speed / road type	Urban / suburban / rural areas
Driver license holding	Number / type of intersections	
Presence of children	Number of turns / left-turns (especially with heavy traffic)	
Household variables, including household size, workers, number of cars, competition for cars	Cycle / walk facilities / facility continuity, including cycleway network, parking, and complementary infrastructure such as showers	

Education level	Parking on road (parallel or angle) / parking occupancy	

A 2017 meta-analysis of the "Use of Direct-Demand Modeling in Estimating Nonmotorized Activity" (Munira & Sener, 2017) details significant variables from at least 18 studies that estimated pedestrian volumes. For a list of the studies, see Table 1 on pages 13 to 25 in that report, and for a list of significant factors, see Table 3 on pages 34 and 35. Variables covered in the latter table, and their impact on pedestrian volumes, included:

- **Demographic**, including population density (positive, 5 studies), population (positive, 2 studies), percent of non-White residents (positive, 1 study), percent of residents with a college education (positive, 2 studies), percent of Black residents (negative, 1 study)
- **Socioeconomic**, including Household income (negative, two studies), Total employment (positive, one study); Employment density (mixed, two studies)
- **Network/interaction with vehicle traffic**, including Major roads length (positive, one study); Percent of major arterials (negative, 1 study); Number of street segment (positive, one study); Street lengths (positive, one study); Principal arterial street (of count location) (negative, 1 study); Arterial street (of count location) (positive, 2 studies); Collector street (of count location) (positive, 2 studies); Presence of four-way intersection (positive, one study); Number of lanes (positive, one study);
- **Pedestrian or bicycle-specific infrastructure**, including Sidewalk length (positive, one study); Off-street trail length (mixed, two studies); Pathway length (positive, one study); Presence of bike lane (positive, one study); Presence of sidewalk (positive, one study); Footway pavement width (positive, one study);
- **Transit facilities**, including Number of transit/bus stops (positive, six studies); Presence of a subway stop (positive, one study); Bus frequency (positive, one study); Accessibility to an underground station (positive, one study);
- **Major generators**, including, Distance from the CBD/downtown (negative, 2 studies); Distance from the nearest body of water (negative, 1 study); Proximity to a university campus (positive, 2 studies); Number of schools (positive, one study);
- **Weather and environmental**, including Recorded temperature (negative, 1 study); Very warm temperature (max. temperature >32°C) (negative, 1 study);
- **Land use**, including Residential land use (mixed, four studies); Land-use mix (area of retail, office, and commercial space per housing unit) (positive, one study);Retail area (mixed - mostly positive, 3 studies); Office space area (positive, 1 study); Industrial area (negative, 1 study); Cultural and entertainment space area (positive, one study); Storage and maintenance space area (positive, one study); Vacant space area (positive, one study); Open space area (positive, one study); Job accessibility (positive, one study); Dwell count (positive, one study); Commercial space (positive, three studies); Open space (negative, one study); Schools (positive, one study); high-activity zone intersection (positive, one study); Maximum/mean slope (negative, 2 studies); Traffic-signal-controlled intersection (positive, one study); Patch richness density (positive, one study); Single-family residential areas (negative, 1 study);

Average visibility within the street network (positive, one study); Tourist and downtown area (positive, one study);

Ongoing work by a team led by PSU conducted a review of factors related to pedestrian activity in existing studies (Kothuri et al., n.d.). Initial results of that review were generally consistent with the Munira and Sener findings. Beyond those listed above, several additional variables identified include:

- percentage of residents with a 4-year degree (positive, two studies);
- household size (positive, three studies);
- vehicle ownership (negative, three studies);
- presence of university (positive, two studies); and
- existence of slope on street (negative, four studies)

NCHRP Report 770 included a table of built environment factors, with elasticities for impact on walking activity, presented here as Table 2.3.

Table 2.3: From NCHRP Report 770, "Table 3-1. Weighted average elasticities of walking in relation to built-environment factors"

"D" Variable	Measure	Elasticity
Density	Residential density	0.07
Density	Employment density	0.04
Density	Commercial FAR	0.07
Diversity	Mix entropy	0.15
Diversity	Jobs / housing balance	0.19
Diversity	Distance to nearest store	0.25
Design	Intersection density	0.39
Design	Percent 4-way intersections	−0.06
Destination Accessibility	Distance to nearest transit stop	0.14
Destination Accessibility	Jobs within 1 mile	0.15

For discussion of methods of modeling pedestrian activity via direct demand, see:

- *Use of Direct-Demand Modeling in Estimating Nonmotorized Activity: A Meta-analysis*, starting at page 28, (Munira & Sener, 2017).
- *Guide for Scalable Risk Assessment Methods for Pedestrians and Bicyclists,* Step 6 (Turner et al., 2018).
- *Synthesis of Methods for Estimating Pedestrian and Bicyclist Exposure*
- *to Risk at Areawide Levels and on Specific Transportation Facilities*https://rosap.ntl.bts.gov/view/dot/36098/dot_36098_DS1.pdf, starting at page 29, (Turner et al., 2017) .

2.4.1 Applications of Data Fusion

Data fusion involves combining data from different sources to estimate volumes, and has been used in recent years to estimate bicycle volumes and, less frequently, pedestrian volumes. Input sources may include counts from pedestrian or bicycle counters, emerging and big data sources such as activity estimates or counts from fitness apps (e.g., Strava) or passively collected, and mode-unspecified, mobile device data (e.g., StreetLight), as well as the types of variables typically seen in direct demand models, such as land use, density, transportation infrastructure, and so on.

The use of data fusion to estimate pedestrian volumes is very limited. In fact, a recent review of the use of crowdsourced mobile data to estimate active transportation activity found that only two of 22 reviewed such studies included pedestrian volumes (Tao et al., 2024). Torbic et al. (2023) in NCHRP Research Report 1064: Pedestrian and Bicycle Safety Performance Functions, developed pedestrian SPFs on a range of facility types in Minneapolis and Philadelphia. Their models used a set of counts provided by each city, and used direct demand models to estimate annual average daily pedestrian and bicycle volumes as an input to the SPF models. The pedestrian exposure models, in addition to the, primarily short-term counts, included population density, walk commute share, employment square footage of uses with foot traffic, land-use entropy, number of transit stops, and some roadway characteristics – for example if the facility is an arterial, and if the speed limit is 25 mph or less (Torbic et al., 2023)

Kothuri et al (2024) fused land use, sociodemographic data and network data ("static" data) with Strava pedestrian activity data to estimate two-hour intersection counts at select locations in Portland, OR. The study found that static data combined with Strava data improved the pedestrian estimates compared to static only models, albeit only slightly - pseudo-R2 increased from 0.76 to 0.8 with the Strava data, and prediction improved (Kothuri et al., 2024). Data fusion studies to explore bicycle volume estimation have found that static data sources, Strava and Streetlight each improve model fit and prediction performance (Broach et al 2023).

Recent scholarship has also identified pedestrian push button actuations as a potential valuable source of pedestrian activity (Kothuri, Singleton et al 2024), although it has seen little documented application in data fusion models.

2.5 VARIABLES SIGNIFICANT IN PEDESTRIAN SAFETY ANALYSES, INCLUDING PEDESTRIAN SPFS

Moving beyond the more general literature on the safety systems approach and highway safety manual methods, we scanned for studies the specifically explored SPFs or crash risk factors for pedestrians. As previously noted, our separate literature review for Phase 1 of this research looked characteristics that could be examined at zonal or area level – typically block group or census tract level (Roll & McNeil, 2021a, 2021b).

In contrast, this scan focuses on studies examining characteristics at the network level – primarily segments or intersections. Details of studies included in this scan are presented in Table 2.2. These include 14 studies that proposed SPFs for pedestrians (Brüde & Larsson, 1993;

Dolatsara, 2014; Gates et al., 2016; Harwood et al., 2008, 2008; Kumfer et al., 2019, 2024; Kwayu et al., 2020; Lyon & Persaud, 2002; McArthur et al., 2014; Quaye et al., 1993; Singh et al., 2021; Thomas et al., 2017; Thomas, Kumfer, et al., 2018; Torbic et al., 2010, 2023), and others that examined pedestrian crash risk but did not develop SPFs (Clifton et al., 2009; Dai & Jaworski, 2016; Guerra et al., 2019; Hu et al., 2018; Kim, 2019; Lee et al., 2017; Miranda-Moreno et al., 2011; Monsere et al., 2017; Pulugurtha & Sambhara, 2011; Singleton et al., 2022; Xie et al., 2018; Yu, 2015).

This section discusses factors found to be related to pedestrian crash risk in these SPF and crash risk studies focuses on networks, specifically segment and intersection locations.

Table 2.4: Study details, Phase II Pedestrian SPF and crash risk scan

Study	Location	Sample	Analysis level Intersections	Analysis Level Segments	Analysis Level Tracts / zones	SPFs developed?	Statistical approach	Measure of fit / accuracy
Kumfer et al 2024	Montgomery County, MD	16,387 intersections (stop-controlled and signalized) and 29,715 segments (all functional classifications except freeways)	Yes	Yes		Yes	Negative binomial (NB) regression	AIC (Akaike information criterion); BIC (Bayesian information criterion) CURE plots
Torbic et al 2023	Ohio (rural roads); Minneapolis and Philadelphia (urban and suburban)	253 mi of urban and suburban roads; 271 urban and suburban intersections	yes	yes	only for direct demand exposure estimates	yes	NB regression	overdispersion parameter
Singh et al 2021	California	6,198 intersections in the state routes	yes	no	no	yes	Bayesian inference approach (integrated nested Laplace approximation (INLA)) was adopted for bivariate setting	DIC (deviance information criterion), D (posterior mean deviance), PD (effective number of parameters) and LPML (log pseudo marginal likelihoods) are employed.
Gates et al 2016	Lower Michigan	segments midblock crosswalks, signalized intersections	yes	yes	no	yes	NB distribution	overdispersion factor
Harwood et al 2008	Toronto, Charlotte, Minnesota	signalized intersections; arterial segments	yes	yes	no	yes	regression	overdispersion parameter; R2LR
Dolatsara 2014	Detroit, MI	signalized intersections	yes	no	no	yes	NB regression	t test
Kumfer et al 2019	Seattle, WA	midblock arterial segments:	no	yes	no	yes	NB regression; conditional random forest (CRF) regression	AIC

Study	Location	Sample	Analysis level Intersections	Analysis Level Segments	Analysis Level Tracts / zones	SPFs developed?	Statistical approach	Measure of fit / accuracy
Thomas et al 2017	Seattle, WA	all intersections of three or more legs	yes	no	no	yes	NB regression	AIC; BIC
Lyon and Persaud 2002	Toronto, ON	three- and four-legged urban intersections, with and without signal control	yes	no	no	yes	NB regression	k overdispersion parameter
Quaye et al.	Hamilton, ON	semi protected left-turn schemes; permissive left-turn schemes	yes	no	no	yes		
Brude and Larsson	Sweden	285 urban intersections in Sweden, including three- and four-legged signalized and non-signalized intersections and roundabouts	yes	no	no	yes	least squares regression	
Kwayu et al 2020	Michigan	stratified random sampling of all urban intersections in Michigan State that comprise of collector and arterial roads	yes	no	no	yes	Zero-Inflated Poisson	AIC; BIC; and Vuong test
MacArthur et al 2014	Michigan	crashes in vicinity of schools	no	yes	no	yes	Random Effects NB Model	Marginal Effect
Thomas et al 2018	Seattle, WA	segments	no	yes	no	yes	NB	AIC; BIC
Singleton et al, 2022	Utah	Segments, Non-signalized Intersections, Signalized intersections on State highways and local "federal aid" roads and streets. Not local streets	yes	yes	no	no	NB and Poisson models	McFadden's pseudo-R2
Xie et al 2018	Hong Kong	262 signalized intersections	Yes	no	no	no	Poisson; full Bayesian inference using the	deviance information criterion (DIC)

Study	Location	Sample	Analysis level Intersections	Analysis Level Segments	Analysis Level Tracts / zones	SPFs developed?	Statistical approach	Measure of fit / accuracy
							Markov Chain Monte Carlo algorithm	
Monsere et al 2017	Oregon	188 statewide segments and 184 intersections	yes	yes	no	no	binomial logistic regression	AIC
Lee, Abdel-Aty, and Cai 2017	Florida	intersections	yes	no	no	no	mixed-effects NB model	AIC; BIC, McFadden's 2, and adjusted 2.
Miranda Moreno et al 2011	Montreal	519 signalized intersections	yes	no	no	no	generalized NB model (GNB); a latent-class NB model; Bivariate Poisson model	Pearson correlation and normalized root mean squared deviation (NRMSD)
Pulugurtha and Sambhara 2011	Charlotte	176 randomly selected signalized intersections	yes	no	no	no	Wald chi-square	quasi-likelihood under independence criterion (QIC); corrected version of quasi-likelihood criterion (QICC)
Dai and Jaworski 2016	DeKalb County, Georgia	network: 100-meter segments	no	yes	yes	no	bivariate correlation; multivariate analysis based on NB regression	over-dispersion parameter
Hu, Zhang and Shelton 2018	Houston, TX	2,286 traffic light controlled-intersections, 18,882 stop sign controlled-intersections and 57,204 non-controlled intersections	yes	no	no	no	Global Colocation Quotient; Local Colocation quotient	Monte Carlo simulations
Kim 2019	Los Angeles county, CA	intersections (80,108)	yes	no	yes	no	multinomial logistic (MNL) regression	pseudo-R square; Cox and Snell and Nagelkerke; McFadden
Yu 2015	Austin, TX	140 census tracts + crash specific characteristics	yes	yes	yes	no	single-level ordered logistic model and the multi-level ordered logistic model (five	intra-class correlation coefficient (ICC); AIC

Study	Location	Sample	Analysis level Intersections	Analysis Level Segments	Analysis Level Tracts / zones	SPFs developed?	Statistical approach	Measure of fit / accuracy
							KABCO categories); single-level binomial logistic model and the multi-level binomial logistic model (fatal/high and no injury/low)	
Guerra et al 2019	Philadelphia region, PA	intersections and segments (~250,000); census tracts (998)	yes	yes	yes	no	multilevel NB models with the lme4 package	AIC; BIC
Clifton et al 2009	Baltimore City, MD	crash location (with 1/4-mile buffer)	yes	yes	yes	no	Ordered probit model	log likelihood statistics

2.5.1 Activity

Variables for measuring pedestrian activity, including some variables related to pedestrian activity, are presented in this section. In the determining risk factors section, NCHRP Report 893 notes that, while SPFs for motor vehicles require a measure of vehicle volume, or AADT, developing SPFs for pedestrian would need both an AADT plus a measure of pedestrian volume or activity (Thomas, Sandt, et al., 2018). That latter of which has not been readily available in past years for many agencies.

Increased pedestrian volume was typically associated with more pedestrian crashes. The exceptions to this trend were that Thomas et al. (2018) found that at locations with average annual daily pedestrian (AADP) volumes greater than about 9,000, the crash risk starts to decrease. Pulugurtha and Sambhara (2011) found that for intersections they defined as having higher pedestrian volumes (more than 40 pedestrian overs a 12-hour period in Charlotte, North Carolina), higher vehicle volumes were associated with a decrease pedestrian crash risk.

Pedestrian volume data came from a range of sources, often estimates or proxies for a lack of robust count data. Some studies collected count information for a set of sites (e.g., Gates et al., 2016; Harwood et al., 2008; Lyon & Persaud, 2002; Miranda-Moreno et al., 2011); others developed models for pedestrian or deployed pre-existing models (Kumfer et al., 2019; Singh et al., 2021; Thomas et al., 2017; Thomas, Kumfer, et al., 2018; Xie et al., 2018; Torbic et al., 2023); and some relied on proxies such as population density and commute walk rates (Clifton et al., 2009; Dai & Jaworski, 2016; Guerra et al., 2019; Lee et al., 2017; Yu, 2015). One used pedestrian volume estimates derived from pedestrian actuations at traffic signals (Singleton et al., 2022).

Other variables that may, at least in part, be proxies for pedestrian activity. In particular, the presence of or number of transit stops has been found to be associated with increased pedestrian crashes (Dai & Jaworski, 2016; Dolatsara, 2014; Harwood et al., 2008; Kim, 2019; Kumfer et al., 2019; Miranda-Moreno et al., 2011; Singleton et al., 2022; Torbic et al., 2023). One study found that while bus stops in general were associated with more pedestrian crashes, near side bus stops were actually associated with fewer pedestrian crashes at signalized locations (Singleton et al., 2022). A pair of studies looking at the number of daily bus arrivals was also associated with increased crashes at both segment and intersection locations (Thomas et al., 2017; Thomas, Kumfer, et al., 2018).

2.5.2 Roadway features

Various roadway segment and intersection features have been found to be significant in pedestrian crash risk models. Roadway segments and intersections with higher motor vehicle volumes (Dolatsara, 2014; Gates et al., 2016; Harwood et al., 2008; Kumfer et al., 2019; Kwayu et al., 2020; Lee et al., 2017; Lyon & Persaud, 2002; Miranda-Moreno et al., 2011; Singleton et al., 2022; Thomas, Kumfer, et al., 2018; Torbic et al., 2010; Xie et al., 2018) and classifications of arterials or highway (Singleton et al., 2022; Thomas et al., 2017) were associated with more pedestrian crashes in nearly all cases, although one study noted that limited access highways were associated with a decrease (Guerra et al., 2019). Other factors consistently found to be associated with increases in pedestrian crashes include locations with higher speed limits (Guerra et al., 2019; Kumfer et al., 2019; Monsere et al., 2017; Thomas, Kumfer, et al., 2018; Yu, 2015), longer segments (Gates et al., 2016; Guerra et al., 2019), more driveways (Gates et al., 2016; Kim, 2019), more lanes(Kim, 2019; Kumfer et al., 2019; Monsere et al., 2017; Thomas et al., 2017; Thomas, Kumfer, et al., 2018). One study did, however, find that the presence of a speed limit sign was associated with fewer pedestrian crashes (Dolatsara, 2014). Torbic et al also found that, for one-way roads, two- or three-lane roads were associated with fewer pedestrian crashes compared to one-lane one-way roads (Torbic et al., 2023). The presence of parking lanes was generally associated with increases in pedestrian crashes, though one study found a negative effect for uncontrolled locations (Gates et al., 2016).

Midblock crosswalks found to be associated with an increase in pedestrian crashes, and three studies (Kumfer et al., 2019; Singh et al., 2021; Thomas, Kumfer, et al., 2018) found the presence of streetlights to be associated with more crashes; although in both cases its possible these are partially proxies for pedestrian activity, and / or may have been places at locations with pre-existing elevated risk.

Two studies found slope to be associated with decreased pedestrian crash risk (Thomas et al., 2017).

For intersection specific variables, signalized locations and locations with more legs are generally associated with higher pedestrian crash risk. Torbic et al found that protected left turns and no-turn-on-red restrictions were associated with fewer pedestrian crashes (Torbic et al., 2023).

2.5.3 Built Environment and sociodemographic variables

Various factors related to the built environment and socioeconomic realm have been found to be significant in pedestrian crash risk. For example, more commercial land uses were commonly found to be associated with greater pedestrian crash risk (e.g., in Kumfer et al., 2019; Dai and Jaworski, 2016; Singleton et al., 2022). Population density generally associated with more pedestrian crashes (e.g., Kumfer et al., 2019; Singleton et al., 2022), which may be in part a proxy for activity. Income is consistently found to be negatively correlated with pedestrian crashes, as detailed in the Phase 1 report for this project (Roll and McNeil, 2021b)

3.0 DATA DEVELOPMENT

3.1 OVERVIEW

This research integrates many datasets to analyze pedestrian safety in Oregon urban areas. This chapter documents the input data and methods used to develop each of the analytic datasets for this research. The goal of the data development process is to create analytic datasets that represent traffic operations, geometric conditions, and built-environment factors for roadway segments, intersections, and transit stops. Each of the following subchapters describe the data development steps shown in **Figure 3.1**.

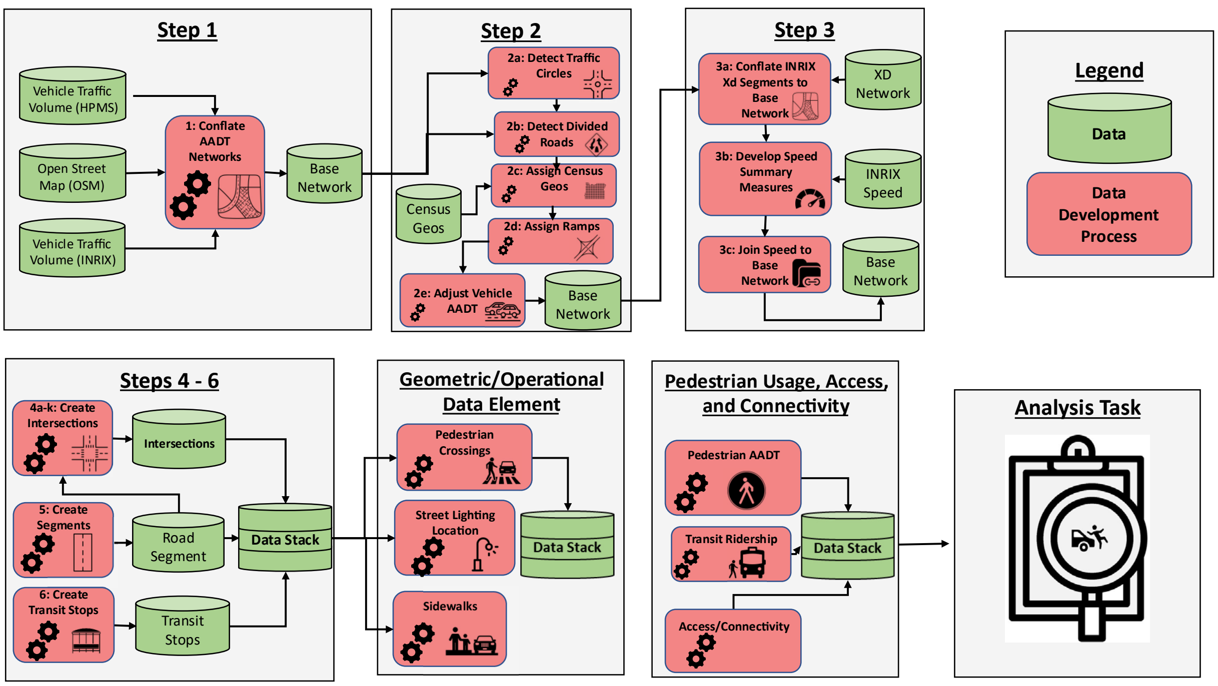


Figure 3.1: Data Development Flow Chart

All of the data available in this research can be explored in more detail using the R Shiny app data viewer available at the address below:

https://joshroll.shinyapps.io/Oregon_Pedestrian_SPF_Explorer/

3.2 NETWORK DATA

This research utilizes a number of high-quality network datasets, including a partial streets Open Streets Map (OSM) with estimated vehicle AADT from INRIX ®, an all-streets OSM network, and the Highway Performance Monitoring System (HPMS) network. Multiple network datasets are required because no one dataset contains all the necessary data elements. The INRIX® OSM data was used as the base network since it represented most roads classified as collector or higher and it has an estimate of vehicle volume necessary for analysis. The primary components of each network dataset are described below to help readers understand the motivations for including the various datasets.

1. OSM with INRIX® AADT Estimates (Base network)
    a. Vehicle AADT Estimates
    b. Detailed roadway geometries
    c. Eases integration with INRIX probe-based speed data (operational speeds)
2. All-streets OSM
    a. Road geometries for local/residential roads
    b. Pedestrian crossing (Portland UA)
    c. Driveways
3. Highway Performance Monitoring System (HPMS)
    a. Vehicle AADT estimates
    b. Lane Count
    c. Posted Speed

Combining network datasets occurs through a process called network conflation. In this process network geometries are compared using a variety of measures like heading, bearing, angle, and proximity, and if possible, attributes from the road themselves such as street name are compared to assess whether attributes from one network can be assigned to the other. This research uses network conflation to join attributes from the available datasets to the base network. These processes are described in more detail in the following subchapters.

3.3 INITIAL HPMS AND OSM NETWORK CONFLATION (STEP 1)

The spatial HPMS data was conflated to the bae OSM network using a custom network conflation algorithm. This algorithm crawled link by link in the base network data and searched for HPMS network links nearby and measured the nearness, bearing, and angle to determine the links that matched. **Figure 3.2** summarizes this process.

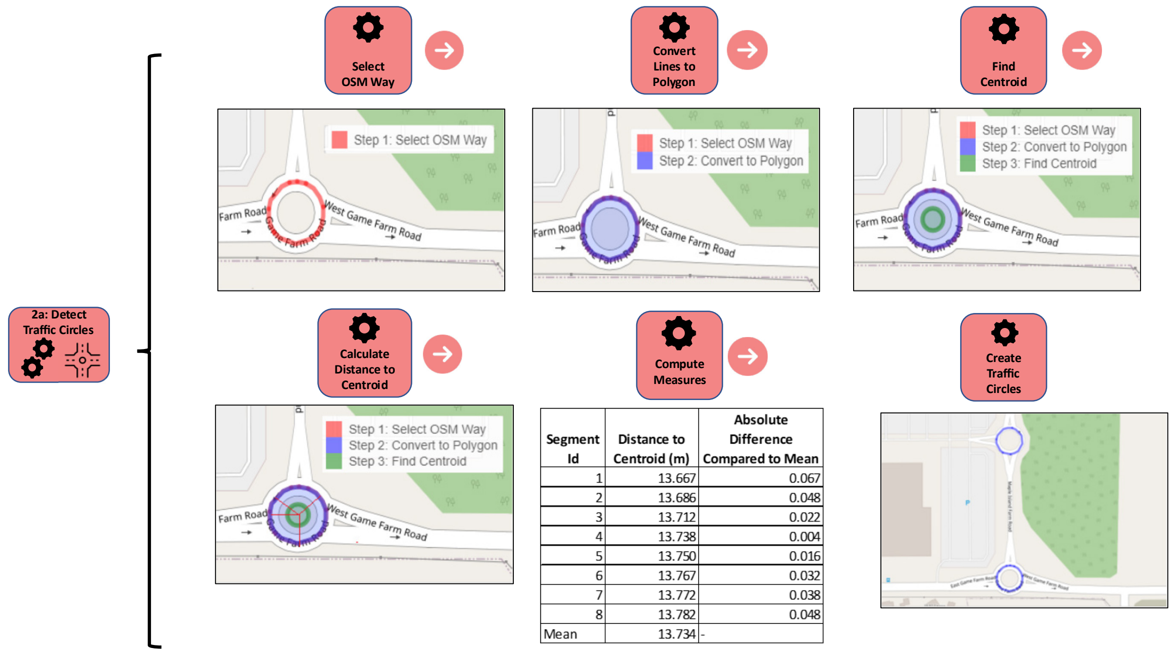
Measure	Value
Link_Id	12676
Mean_Distance	1.15
Tops_Bear	27.2
Osm_Bear	26.6
Osm_Angle	1.11
Tops_Angle	1.11
Prop_5m	1
Prop_10m	1

Figure 3.2: Step 1 – HPMS and Base Network Conflation Process

3.4 TRAFFIC CIRCLES, DIVIDED ROADS, CENSUS GEOGRAPHIES, INTERSTATE RAMPS, AND AADT ESTIMATION (STEP 2)

Step two of the data development process prepares a number of elements for later analysis including classifying base network segments as a part of traffic circles (roundabouts) or interstate ramps, defines roads as part of a divided roadway (with median separation), and assigns census geographies.

Traffic circle related segments are determined by looking at each OSM way ID, which is a group of OSM segments and then for that OSM way calculating a centroid. The process then measures the distance from each related OSM segment to the centroid, and if those measures are all within a small tolerance relative to the average, determines that those segments are part of a traffic circle. **Figure 3.3** summarizes this step.



Segment Id	Distance to Centroid (m)	Absolute Difference Compared to Mean
1	13.667	0.067
2	13.686	0.048
3	13.712	0.022
4	13.738	0.004
5	13.750	0.016
6	13.767	0.032
7	13.772	0.038
8	13.782	0.048
Mean	13.734	-

Figure 3.3: Step 2a – Classifying OSM Segments as Traffic Circles Process

3.4.1 Detecting Divided Road Segments

Another important step in the data development process was detecting divided road segments. This process was done in a step-wise fashion, crawling through the network data one segment at a time. The process used some of the conflation measures, as well as evaluating the street name for a matching name. A summary of this process is described for a sample segment in Figure 3.4.

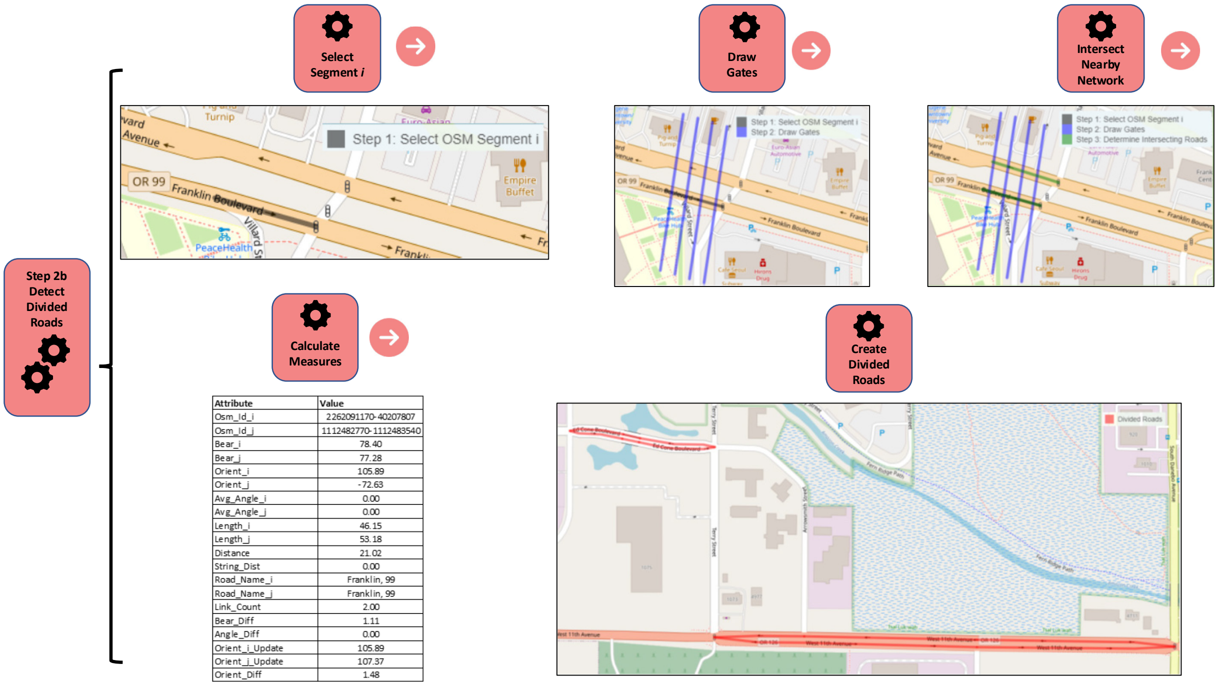
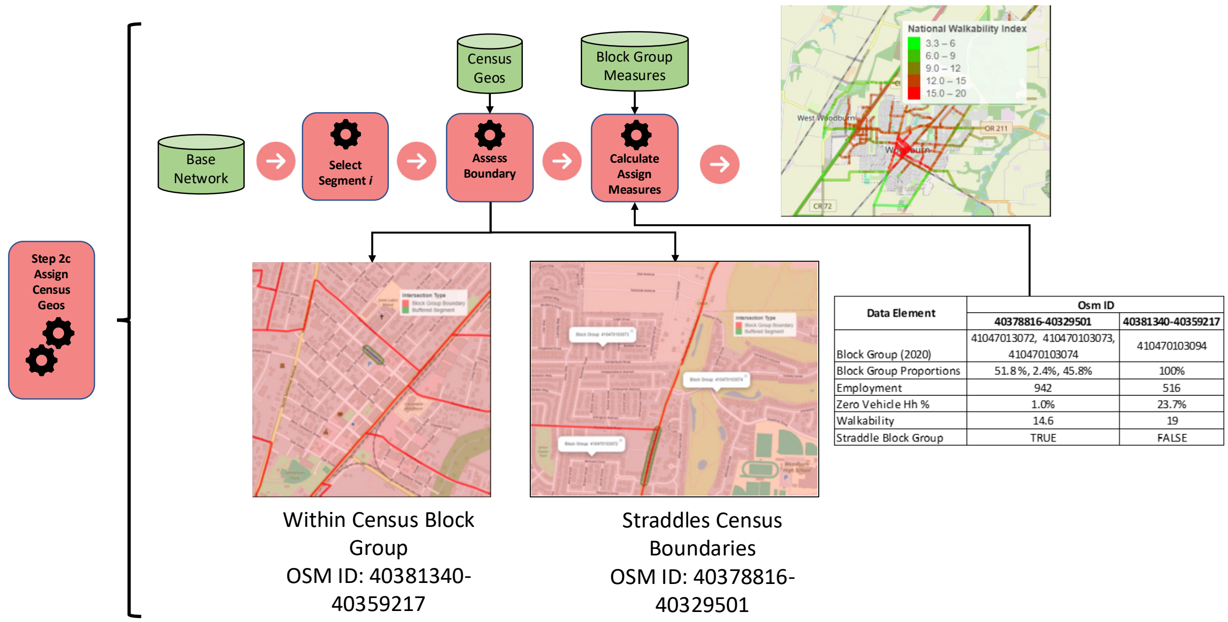


Figure 3.4: Step 2b – Workflow for Detecting Divided Roads in Base Network data

3.4.2 Assigning Census Geographies to Segments

This research will use information about sociodemographic, population, and employment from US Census and requires the segments to be linked to Census geographies. Since many roads cross Census boundaries, a process was developed to apportion Census data to segments based on how much of the segment is present in each Census block and block group when the segment crosses one of these Census boundaries. **Figure 3.5** summarizes this process for a sample segment.

Once the segment and related intersections have the Census geography assigned, Census data is appended to those elements, including information about the population and employment, as well as supplemental datasets such as the Smart Location Database (SLD) developed by Environmental Protection Agency (EPA 2021).



Data Element	Osm ID	
	40378816-40329501	**40381340-40359217**
Block Group (2020)	41047013072, 410470103073, 410470103074	410470103094
Block Group Proportions	51.8 %, 2.4%, 45.8%	100%
Employment	942	516
Zero Vehicle Hh %	1.0%	23.7%
Walkability	14.6	19
Straddle Block Group	TRUE	FALSE

Figure 3.5: Step 2c – Process Example for Assigning Census Boundaries to Segments

3.4.3 Vehicle AADT Adjustment and Estimation (Step 2e)

In some instances, the OSM network segments do not have a matching HPMS segment and therefore do not have an associated AADT value. This happens for a few reasons, including the OSM segment is of lower functional classification then what is monitored in the HPMS data but many segments are divided roads represented by two segments in the OSM network but only one in the HPMS data which when conflating leads to no matches. In many cases, the segments with missing volumes have an estimate derived from INRIX ®, which utilizes probe counts to estimate AADT. ODOT has previous evaluated this data product finding that these AADT estimates are accurate when compared to observed traffic counts (Roll 2023). In some cases, there is neither an HPMS or INRIX estimate of AADT. When this occurred, a data fusion model is constructed using available data, such as the volume of connected network segments, functional classification, access to jobs, network centrality, and the urban area in which the segment resides. Nine models are developed based on data availability so that no segment is discarded due to one or more variables missing when the model is applied.

Multiple modeling frameworks were estimated, where a random forest model was found to perform best in terms of reducing estimation error based on a 10-fold cross validation exercise. The median absolute percent error from the cross-validation process is presented in **Figure 3.6**, along with results from a negative binomial model for comparison. **Figure 3.6** shows the median error for each of the models and demonstrates how the random forest model is able to produce AADT estimates with error less than 20% in most models, and as little as <1% error.

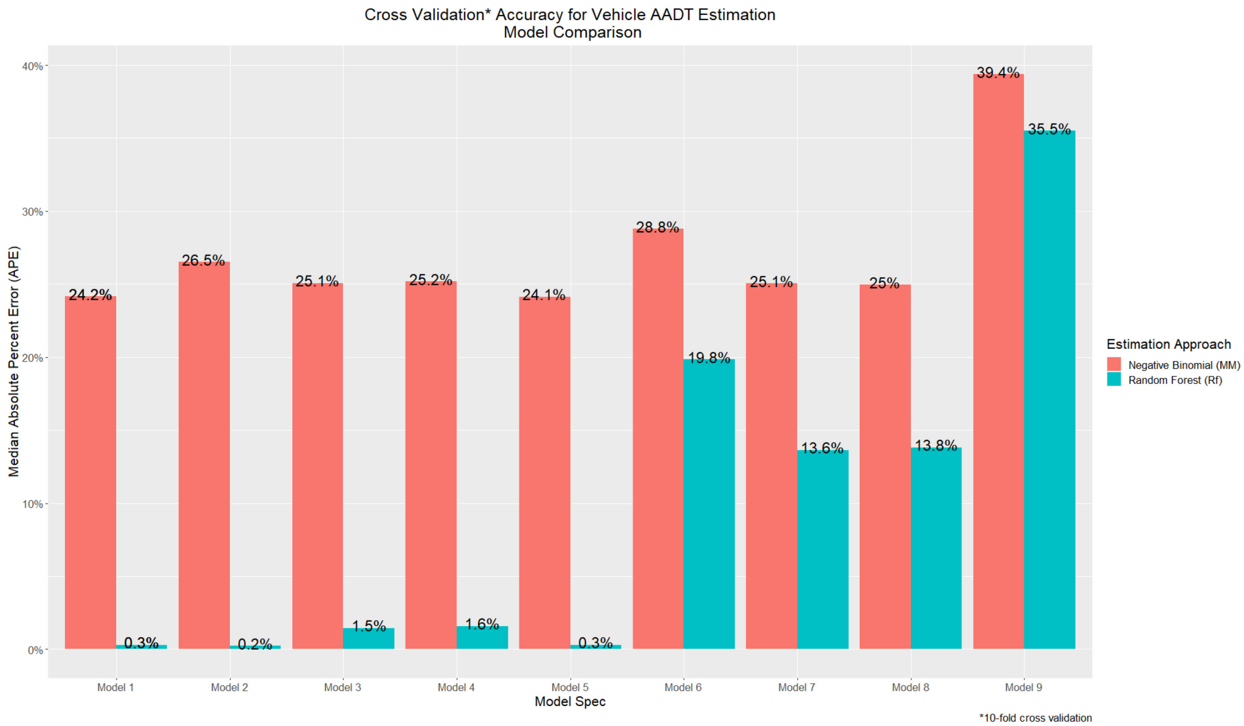


Figure 3.6: Median Absolute Percent Error from 10-fold Cross-Validation for AADT Data Fusion Model

Ultimately, HPMS is used for about 76% of both the count and length of segments, with around 8% derived from INRIX. The remaining 16% is derived from the developed models.

Table 3.1: Summary of Segment Count and Length by AADT Data Source

AADT Source	Segment Count	Segment %	Length Miles	Length %
HPMS (Observed)	68191	75.5%	4216	75.7%
INRIX	7566	8.4%	465	8.4%
Model 1	1584	1.8%	88.3	1.6%
Model 2	3	0.0%	0.06	0.0%
Model 3	1937	2.1%	150	2.7%
Model 4	2076	2.3%	163	2.9%
Model 5	4	0.0%	0.15	0.0%
Model 6	8329	9.2%	454	8.2%
Model 7	57	0.1%	3.43	0.1%
Model 8	78	0.1%	7.25	0.1%
Model 9	273	0.3%	14.7	0.3%

3.5 OPERATIONAL TRAFFIC SPEED NETWORK CONFLATION

The operational speed data comes from INRIX® probe-based speed measures which are similarly on a version of the OSM. Unfortunately, the OSM IDs change, and without knowing the *exact* vintage that someone used, matching on that index can produce a mis match of segments that might not be obvious. To alleviate this issue, conflation is used to match the information from the INRIX based OSM to the base network. The base network is also based on OSM, so joining them using conflation works well.

Speed measures that were appended to segments and intersections are summarized in Table 3.2. Table 3.2 summarizes the average, median, and maximum of processed INRIX® speed data. The trend generally shows that the lower the AADT the lower the speeds for both intersections and segments. This data can be explored in more detail on the data viewer.

Table 3.2: Summary of Speed (mi/h) Measures Appended to Segments and Intersections

Measure	<5K AADT	5–10K AADT	10–20K AADT	20–40K AADT	40–80K AADT
Intersection – Mean Speed (mph)	25.8	25.8	27.5	29	30.3
Intersection – Median Speed (mph)	26	26	27.5	28.9	30.2
Intersection – Maximum Speed (mph)	36	35.7	38.8	42.7	46.4
Intersection – Observations	44,220	23,099	19,070	8,170	867
Segment – Mean Speed (mph)	25.6	26.9	29.4	31.5	51.3
Segment – Median Speed (mph)	25.7	26.9	29.4	31.4	53.6
Segment – Maximum Speed (mph)	35.9	38.8	43.4	47.3	63.2
Segment – Observations	14,263	14,926	15,308	12,784	6,448

3.6 INTERSECTION DATA (STEP 4)

Intersections and segments are the traditional units of analysis in a systemic safety analysis. This section documents the development of the intersection data used in the safety analysis chapter. The intersection data development benefits from data elements from the base network segments, but includes additional data from local agencies where possible. The following subchapters describe the steps featured in **Figure 3.7**.

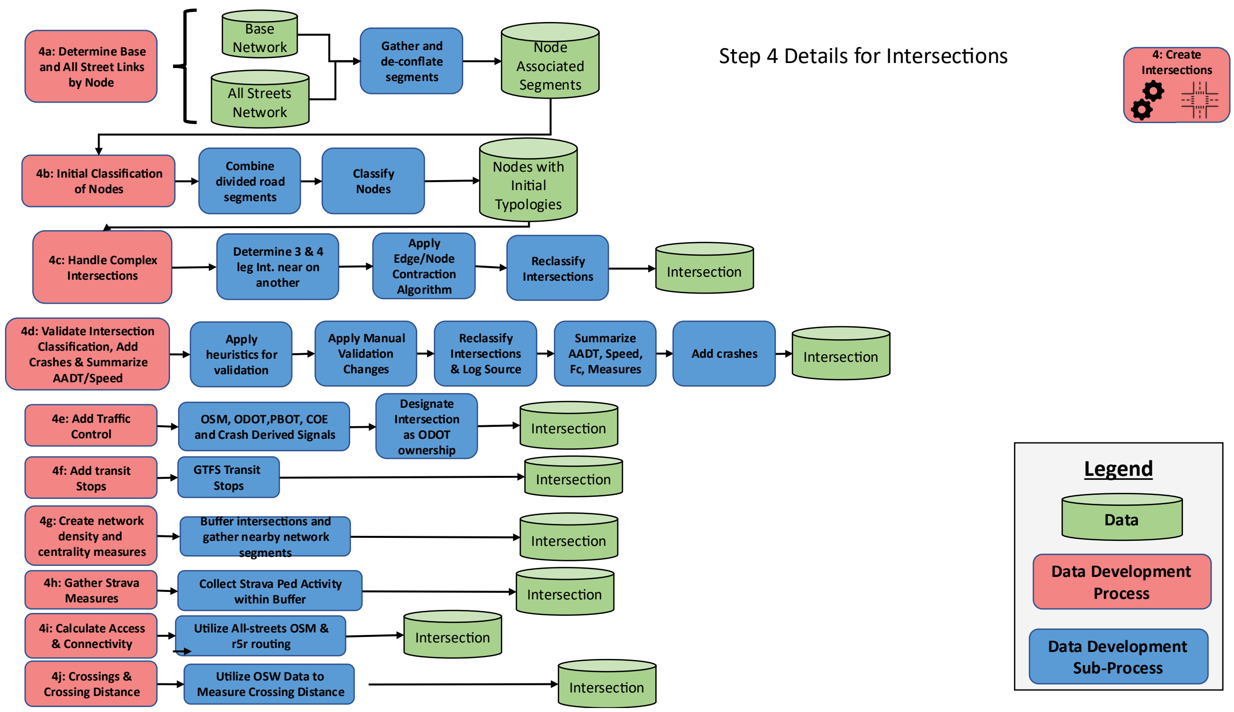


Figure 3.7: Intersection Data Development Processes and Sub-Processes

3.6.1 Intersection Typology, Validation, and Entering AADT Calculation (Step 4a through Step 4d)

The initial step in the intersection data development defines intersections by identifying locations where multiple segments come together and are defined as driveways, three-leg, four-leg, or five-leg or more intersections. To do this, the base network of segments is combined with the all-streets network data and the number of entering legs are assessed to classify the intersections as one of the defined classifications (driveway, three-leg, four-leg, etc.). Once initial classifications are defined in Step 4a and Step 4b, the intersections have a contraction algorithm applied to handle complex network situations where individual lanes are defined by separate network segments. These complex network geometries need to be collapsed to properly assess the number of legs entering the intersection. **Figure 3.8** shows this process for a set of initial intersection nodes and their final classification after applying the network and node contraction algorithm.

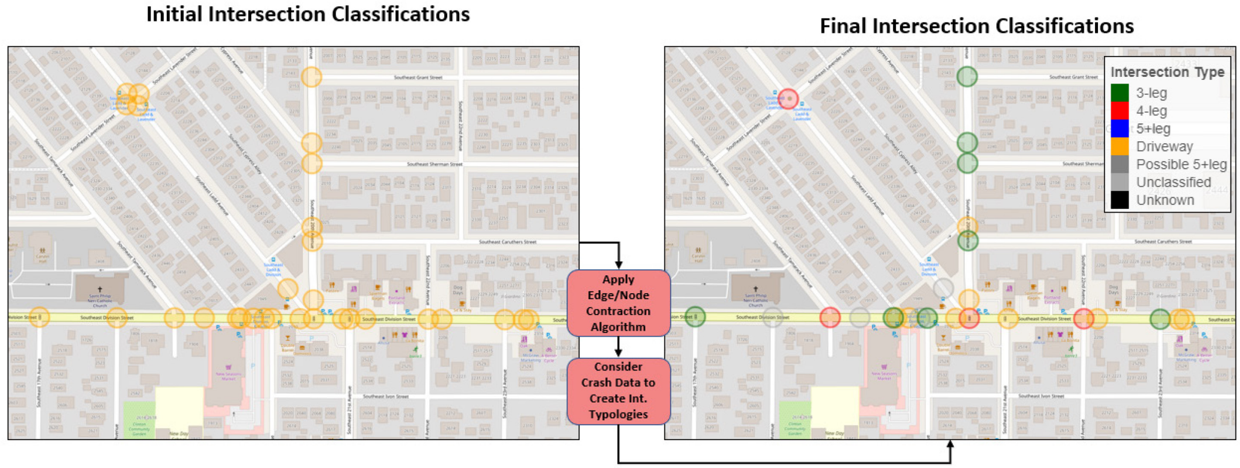


Figure 3.8: Example of Intersection Node Simplification Following Contraction Algorithm Application and Typology Assignment

To validate and establish confidence in the final intersection node typologies, intersection descriptions and number of legs from ODOT crash data was utilized. Lastly, a manual review of intersection typologies was performed, where 908 intersections had to be manually classified due to improper classification. These 908 intersection nodes represent 1.5% of the total (64,000) intersection nodes developed for this analysis. After the intersection typology is assigned and validated, the intersections' entering vehicle AADT are calculated. Entering AADT is calculated by summing each of the entering legs.

The final result of this step is an intersection dataset that has a defined number of legs and an estimate of the number of entering vehicles. **Figure 3.9**

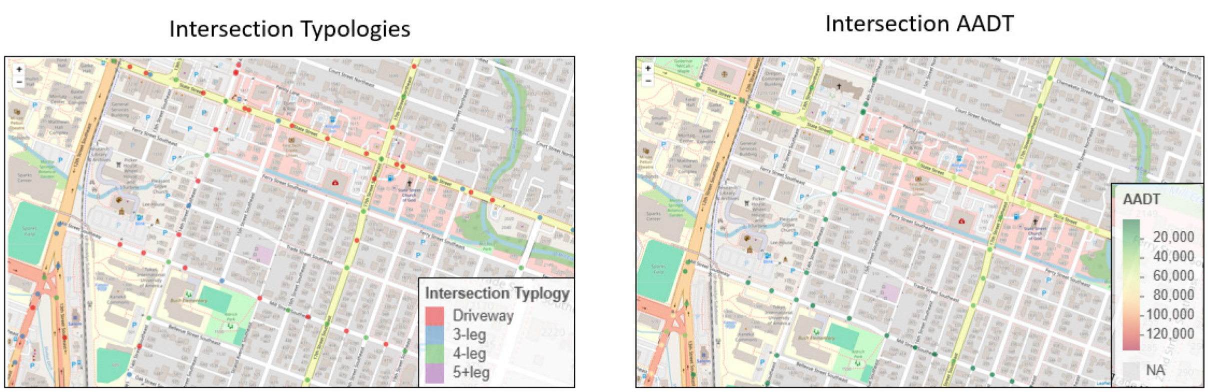


Figure 3.9: Example of Step 4a through Step 4d – Intersection Typology and Entering AADT

3.6.2 Traffic Control Assignment Traffic Signal Location (Step 4e)

There is no single database that records the location of traffic control throughout Oregon, therefore to determine if traffic signals are present, this research utilized existing local data, OSM, and a novel approach that utilizes crash data to inform the presence of traffic signals. This section describes the data integration process used to identify intersections with a traditional traffic signal. Other traffic signal devices, like rectangular rapid flashing beacons, emergency vehicle signals, or other non-motorized signal equipment are not included in this process.

Traffic signal location data for signals owned or managed by ODOT, City of Portland, and City of Eugene were available to this research and were used for establishing with confidence the location of traffic signals. OSM has traffic signal location data as well but is known to be incomplete due to the crowd sourced nature of those data and so has less confidence. These two sources together do not fully account for all the traffic signals in the state so a novel approach to establishing traffic signal locations was developed using crash data and is described in the next section.

3.6.2.1 Using Crash Data Mining to Inform Traffic Signal Location

ODOT maintains crash data with precise latitude and longitude coordinates of the crash for years 2007 to 2022, and though crashes are rare, in any one year the agency records roughly 38,000 to 60,000 crashes a year with over 771,000 crashes recorded over the last 21 years. These data record many details of the crash, including the traffic control device for intersection crashes. In order to fill gaps in knowledge about where traffic signals exist, crash data was joined to the intersection points. To test the reliability of this process it was tested against known locations of traffic signals from the ODOT, City of Portland, and City of Eugene datasets. The number of signals available for testing is summarized in Table 3.3.

Table 3.3: Traffic Signals by jurisdiction

Jurisdiction	Signal Count
Oregon DOT	735
City of Portland	984
City of Eugene	250

For each intersection, including intersections with a known traffic signal, intersection crashes were assigned to the intersection using a spatial join with a maximum distance of 18 meters (59 feet) (derived from some trial and error). **Figure 3.10** shows an example of an intersection in Oregon and crash data locations associated to that intersection as well as the traffic control recorded for each crash. It should be noted that all of the crash locations shown in **Figure 3.10** were randomly jittered to be able to show all the crashes, but the actual crash data spatial coordinates had them located right on top of the intersection.

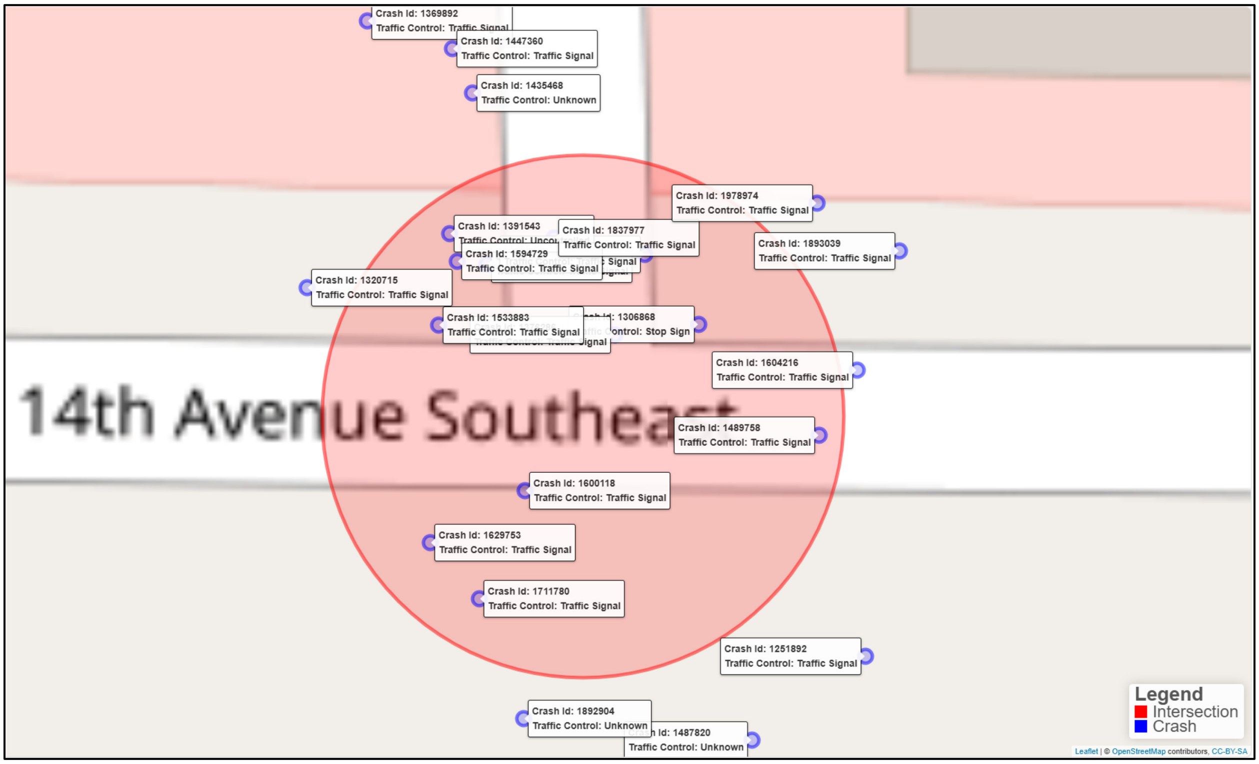


Figure 3.10: Example of crash data locations near selected intersection and their traffic control

The crashes associated with the intersection shown in **Figure 3.10** are summarized in Table 3.4. Summarized is the information on the recorded traffic control device noted by police in the crash report. For this intersection, there were 24 crashes recorded from 2007-2022, with 19 of those indicating the intersection had a traffic signal, while one crash noted a stop sign, one crash noted that the intersection was uncontrolled, and three crash records indicated the traffic control was unknown. This location is known to have a traffic signal based on City of Portland data, as well as visual inspection using Google Street View; it is unclear why some crash records indicate otherwise. Review of other intersections with known traffic signals indicates that this issue is pervasive; therefore, it needs to be accounted for when informing traffic signal location from crash data.

Table 3.4: Crashes associated with intersection shown in Figure 3.10 by Traffic Control Type

Year	Traffic Signal	Stop Sign	Uncontrolled	Unknown
2007	1	0	0	0
2008	1	1	0	0

Year	Traffic Signal	Stop Sign	Uncontrolled	Unknown
2009	1	0	0	0
2010	3	0	1	0
2011	1	0	0	1
2012	1	0	0	1
2013	1	0	0	0
2014	3	0	0	0
2015	2	0	0	0
2016	1	0	0	0
2019	2	0	0	0
2020	1	0	0	1
2022	1	0	0	0
Total	19	1	1	3

To handle this misreporting, a proportion of the total crashes assigned to each intersection is calculated, ignoring the unknowns, so that for each intersection with a known traffic signal in the cities of Portland and Eugene the number of crashes noting a traffic signal was present are counted. Then, the proportion of total crashes that indicated traffic signal traffic control is calculated. To determine the proper cutoff for the number of crashes indicating the presence of a traffic signal and the proportion of the total crashes, a recursive portioning algorithm was deployed using a classification model. The model tree is shown in **Figure 3.11**, and helped to inform the cutoff in the number of crashes and the proportion of the total crashes that were noted as traffic signal controlled.

The recursive portioning model was reasonably accurate with balanced accuracy of 91.7%, sensitivity of 97.5%, and specificity of 86%. A confusion matrix is shown in **Figure 3.12**. **Figure 3.12** shows the model is more likely to predict that a traffic signal exists where it does not than the converse. Because of the risk of false positives (assign intersections a signal where none exists) a higher threshold than what the recursive partitioning model used was employed.

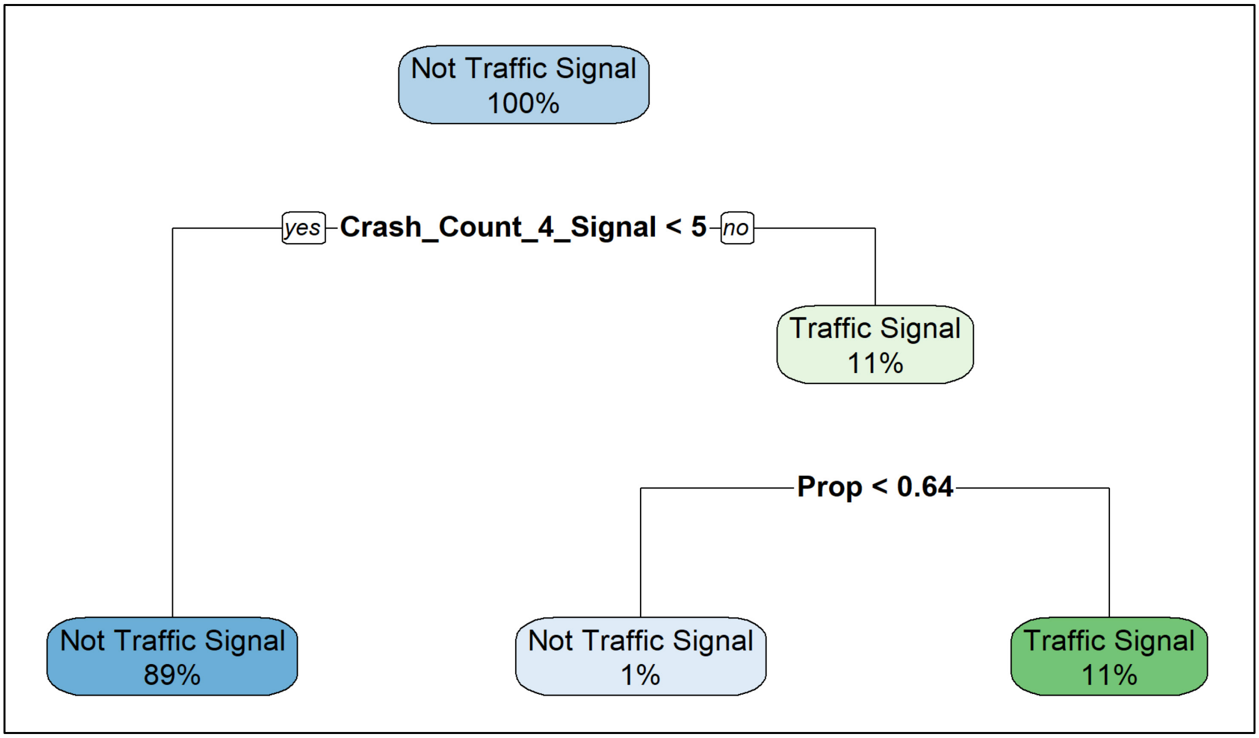


Figure 3.11: Example of crash data locations near selected intersection and their traffic control

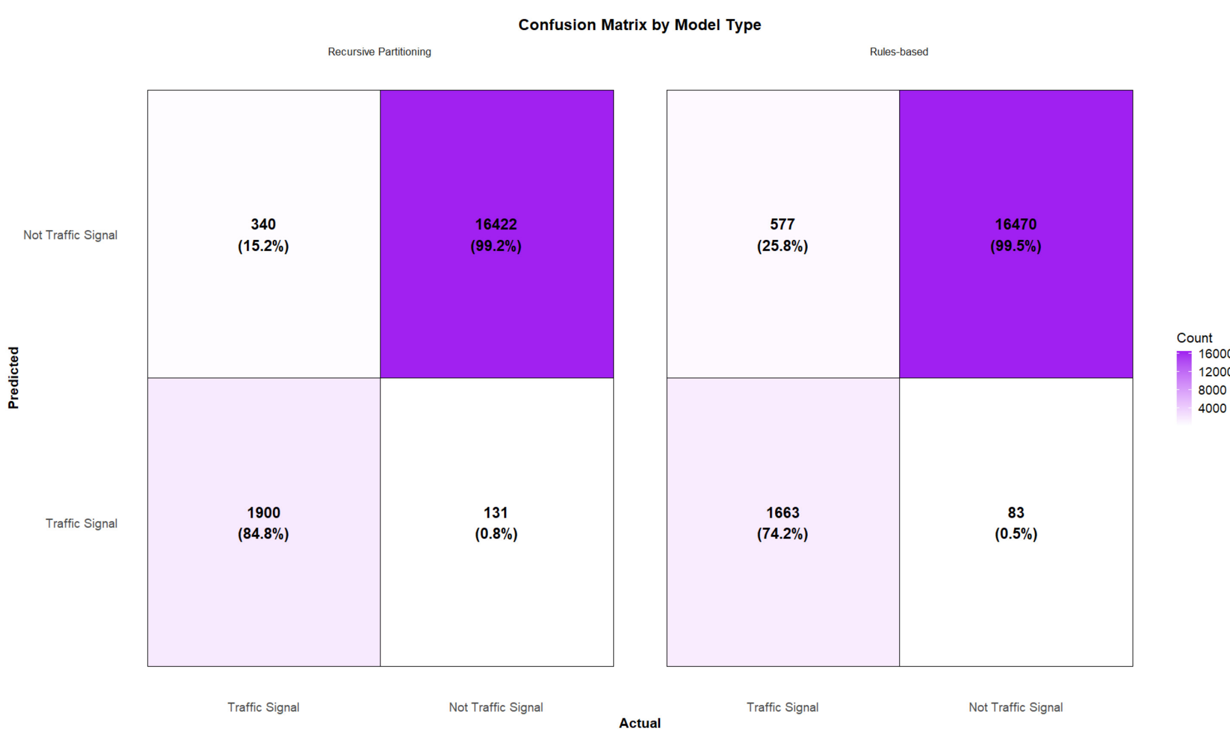


Figure 3.12: Confusion matrix from 10-fold cross validation of recursive partitioning model

To test the thresholds established by the recursive partitioning model, rules established from the recursive partitioning were used on the entire intersection inventory and reviewed. Results found that many intersections were designated as signalized, but upon review in Google Streetview no traffic signal was present. To address this, the rules were adjusted to be more conservative. Instead of assigning a traffic signal to intersections with at least five crashes and at least 64% of those crashes indicating traffic signal control, there had to be 15 crashes stating a traffic signal control or 10 crashes with 75% or more indicating a traffic signal control. To validate these rules, 50 intersections where the rules indicated a traffic signal were manually reviewed. Of the 50 intersections reviewed, 47 had traffic signals, while two previously had signals but had been removed in the past three years and one was very close to a signalized intersection, helping to explain the misclassification.

The accuracy of this smaller validation using the more conservative rules was 94%, however, the two intersections where signals had been removed were removed after the study period established in this research study . Also, in both cases the crash data indicated a traffic signal up until 2022 with the signal changes appearing after that period. If the 2023 and 2024 crash data were timelier, it is possible these data would be utilized to improve the accuracy in cases where signals were removed.

After vetting the crash data mining approach and determining the proper ruleset, data was integrated from Open Street Map, City of Portland, City of Eugene, and Oregon Department of Transportation managed and owned traffic signals into a unified dataset for all the urban areas in Oregon. This data integration resulted in 3,144 unique traffic signals throughout Oregon, as shown in Table 3.5.

Table 3.5: Traffic Signal Source

Data Source	Intersections with Traffic Signals Count	Intersections with Traffic Signals %
ODOT Managed	1247	40%
ODOT Owned	550	17%
City of Portland	995	32%
City of Eugene	250	8%
Open Streets Map (OSM)	1695	54%
OSM (Not in known locations*)	795	25%
Crash Data Derived	2644	84%
Crash Data Derived (Not in known locations*)	293	9%
All Sources - Unique Only	3144	100%

*locations where no other data source indicated a traffic signal

3.6.3 Transit Stop Proximity (Step 4f)

This research incorporates proximity to transit stops by calculating the number of transit stops within a 50- and 100-meter buffer (164- and 328-feet). The source of the transit stop location data is the General Transit Feed Service (GTFS) and the year of data is 2019.

3.6.4 Network Density and Centrality Measures (Step 4g)

This research aims to understand the role of the built environment and pedestrian exposure in pedestrian safety and has created detailed network measures including network density and centrality using the base and all-streets networks.

Network density measures the density of network links around the intersection. For this study, different aggregations of network links near the intersection were created. Examples include the miles of local residential streets, or interstates within a 25-meter (82 feet) buffer. These measures are meant to characterize connectivity and the character of the area, where an intersection with more freeway miles around it is likely to be more disconnected from key destinations and be less attractive to pedestrians.

For this study, centrality betweenness is used, which measures how often a node (or intersection) appears on the shortest path between other nodes in a network, indicating its importance in facilitating movement or connectivity, assuming a shortest distance path but using AADT as an impedance. This research utilized the igraph R package (Csárdi et al 2025).

3.6.5 Gather Strava Measures (Step 4h)

This research utilizes Strava walk, jog, and run trip counts in the pedestrian data fusion models (see Chapter 3.7). The Strava smartphone application is a popular fitness app used by people to track their bicycle, run, jog, and walk trips. These data are a measure of pedestrian activity that likely skews towards recreational and exercise activities, but has been shown to be useful for understanding where people walk or bike (Lee & Sener 2020, Tao et al. 2024). Almost all previous studies have used Strava to understand bicycle activity, so the use of Strava to understand pedestrian volumes is a known gap in the literature. For this research, the count of walk, jog, and run trips were aggregated for all segments within 10- and 25-meters (33- and 82-feet).

3.7 PEDESTRIAN TRAFFIC VOLUMES

This research developed pedestrian volume estimates to be used in the safety analysis chapter by combining data from traffic signal push button derived pedestrian counts and short-duration counts collected via video. These counts are used directly in a subset of crash models, but also used to inform a data fusion model that is used to estimate pedestrian exposure for all intersections in Oregon's urban areas. Details about the methods used to develop the pedestrian data are below.

3.7.1 Traffic Signal Derived Pedestrian Volumes

The traffic signal push button derived counts are created using methods initially developed by Singleton et al (2021) and further tested by Singelton and Kothuri Singelton and Kothuri et al. 2024). This method relies on the correlation between crosswalk signal button pushes and the actual number of pedestrians using the intersection formalized by a quadratic factoring equation that uses traffic signal push buttons as the primary input. For the Oregon specific research (Singelton & Kothuri 2024), observed pedestrian counts were collected and compared with filtered push button data to develop the quadratic equation. The results indicated a high level of accuracy with an average of ± 2.4 pedestrians per hour. For the application of the model in Oregon, this research utilizes data from ODOT traffic signals spread across the entire state, as shown in **Figure 3.13**. For the purposes of this research report, the signals are considered as continuous or permanent pedestrian traffic counters. After cleaning the data and ensuring that at least 350 days of data are available for calculating an Annual Average Daily Pedestrian Traffic Volume (AADPT), there were 558 intersections with an estimate of pedestrian volume.

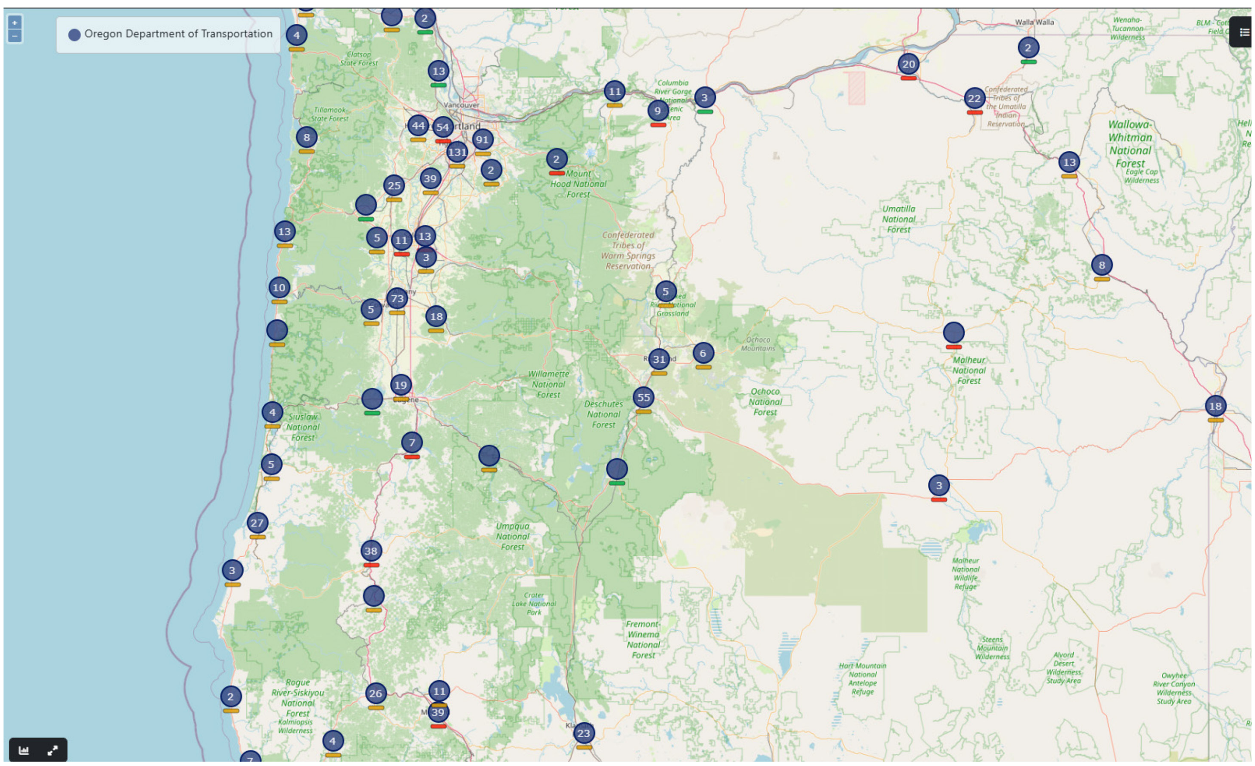


Figure 3.13: Location of Traffic Signals with Push Button Derived Pedestrian Volumes

3.7.2 Short-Duration Pedestrian Traffic Counts

This research utilizes intersection pedestrian traffic counts collected over seven years from across Oregon. These counts are collected via video, manually reduced by ODOT staff or private contractors, and then entered into ODOT's the Oregon Traffic Monitoring System (OTMS). After data cleaning and processing, including expanding the short-duration counts (SDC) to AADPT, there were 340 intersections with a pedestrian traffic volume derived from an SDC.

These data were collected from 2017 through 2023 and included at least 16 hours, but up to 48-hours.

3.7.3 Short-duration Factoring Methods

To expand the SDCs to represent an annual average two factoring methods were tested, including the AASHTO method and the day-of-year (DOY) method. AASHTO factors are recommended by FHWA (TMG 2022), though researchers have found that DOY factors can produce lower error (Nosal et al. 2014). Day-of-year factors are created by dividing the daily volume by the annual volume, then applied to SDCs based on the specific day that the SDC was collected. AASHTO factors are calculated by averaging the daily counts for each day of the week and for each month of the year, resulting in 84 factors for each year. One advantage of AASHTO factors is that if an SDC was collected on a day when the permanent counters were missing data, they can still be applied because they use an average measure using all the days in a month, so one missing day does not preclude the use of these factors. Conversely, with DOY factors, if the day in which the SDC was collected is missing from the permanent count sites then the count at that site cannot be factored. This context also applies to using factors created in one year and applying them to SDCs collected in another year, which was necessary since the permanent counters only had data in years 2022 and 2023.

The SDCs available for this research come from 16-hour and 24-hour counts, instead of 1- or 2-week counts as recommended by the Traffic Monitoring Guide (TMG 2022). This research also tried to match the guidance on the number of permanent sites to use for developing the expansion factors, aiming to use at least five permanent counting sites (TMG 2022). To match the SDC to permanent counters a simple distance-based approach was used where the five nearest permanent counters to the SDC site were used.

To understand the potential error from these temporal expansion factoring methods a series of tests were developed. In these tests, the permanent counter data was used where the annual and AADPT measures are available so that the expansion factoring results can be compared with an empirical measure. **Figure 3.14** summarizes the validation process, where daily counts are extracted from the permanent count data then each synthetic SDC_i is expanded using both the DOY and AASHTO factors. These expanded counts use the five nearest permanent count sites which can be from intersection nearby or in other urban areas. Once the two sets of factors are applied, AADPT is calculated and compared with the observed AADPT for that site. Different accuracy measures are computed, including percent error, absolute percent error, and the nominal difference. These accuracy measures are presented below to highlight the likely error associated with the annual expansion factor process.

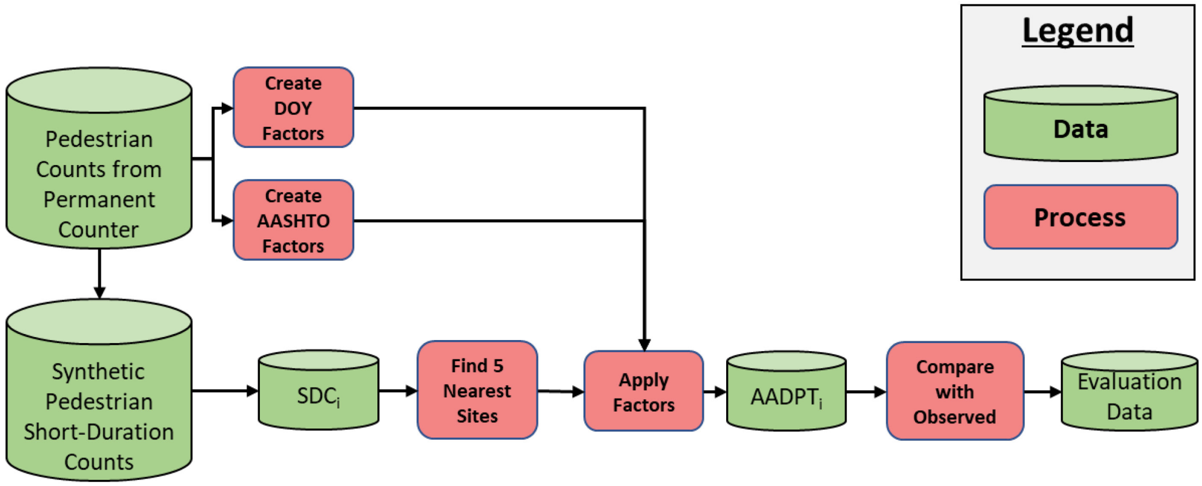


Figure 3.14: Annual Factoring Method Testing Flowchart

The charts in **Figure 3.15** show the mean absolute percent error by expansion method, month, and data year for 548 permanent count sites with enough data to perform the tests. The AASHTO method consistently produced the lowest amount of error with summer and spring months performing better than winter and fall months. This was the case when using 2022 data or 2023 data in the tests.

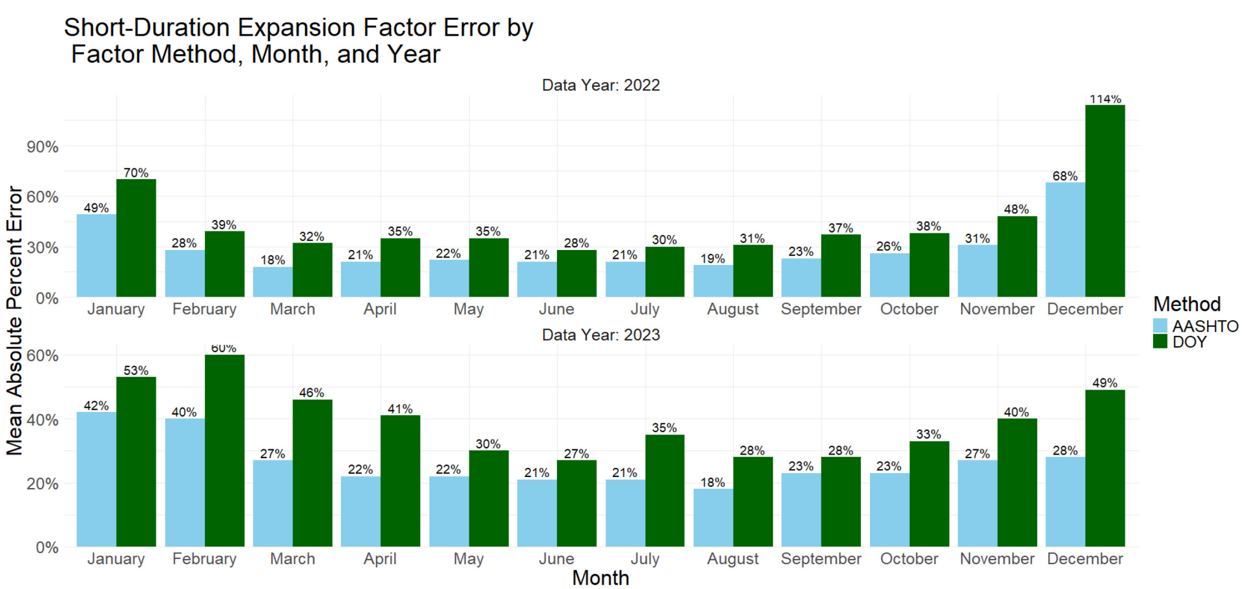


Figure 3.15: Mean Average Error from Annual Factoring Method Testing

To assess how the distance between the SDC and the permanent count stations impact estimation error, **Figure 3.16** presents the MAPE from the application of AASHTO factors by month and distance between stations. For most months, the distance between SDC and permanent counter can impact the MAPE, though it's typically a marginal effect. For spring and summer months, the difference between the error from SDCs factored by closer permanent sites (500 meters, 1,640 feet, or less) and permanent sites further away can range from 1% to 10%. The largest difference for those seasons were observed in August, where the MAPE is 17.1% when sites are 500 meters (1,640 feet) or less from one another and 26.2% when the sites are more than 8001 meters (about five miles) away. The error differences between distance thresholds are much greater for winter and fall months, where the MAPE varies from 6% to 25%, with September appearing to show the largest difference with 47.2% error at more than 8001 meters (about five miles) compared to 22.4% when sites are 500 meters (1,640 feet) or less from one another.

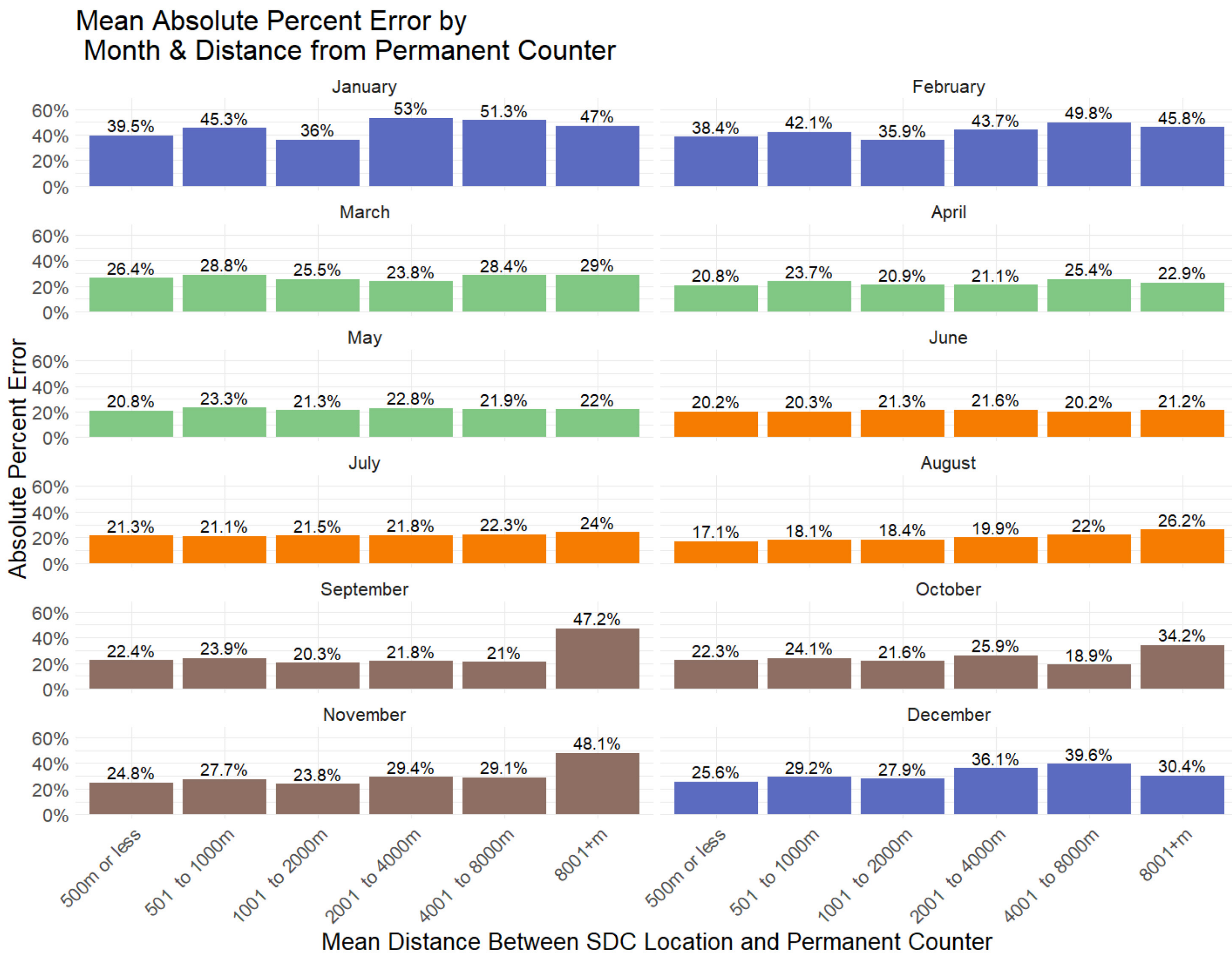


Figure 3.16: Mean Average Error from Annual Factoring Method Testing

These figures highlight a relatively low amount of error from the expansion factoring methods, with much of the expected error under 25% for most of the factored SDCs. The factored SDCs along with the permanent count locations (n = 989) are shown in **Figure 3.17** where the locations are plotted on a map of Oregon. These pedestrian count data will be used to train a model using available variables from the data fusion process, which will then be used to estimate pedestrian volumes in locations where no direct observation of pedestrian counts is available.

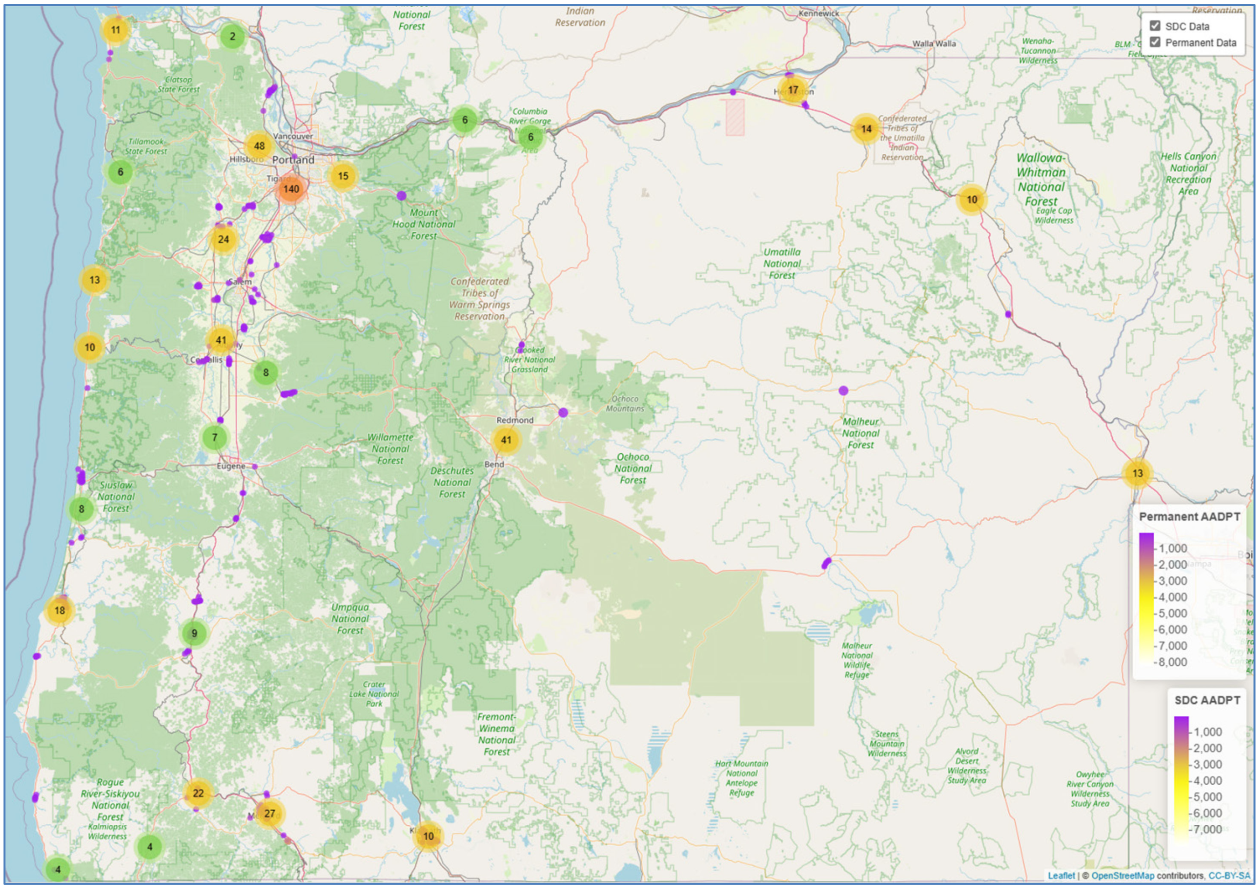


Figure 3.17: Statewide Locations of SDC and Permanent Count Locations

3.7.4 Data Fusion Model for Estimation of Pedestrian Volumes at Intersections

The previous sections document how this research effort developed pedestrian counts using traffic signal derived push buttons and short duration counts. These direct observations of pedestrian volumes were available at 889 intersections, but to understand pedestrian safety systemically, estimates of pedestrian volumes at all intersections are necessary. To attain these estimates, a number of data fusion modeling techniques are tested to determine the modeling framework that performs the best in terms of accuracy, error, and application. The modeling frameworks tested included the negative binomial with random effects, random forest, XGBoost, and neural networks.

Data fusion modeling was done in the R open-source statistical software using the following packages: glmmTMB (Brooks et al. 2017), caret (Kuhn 2008), ranger (Wright and Ziegler 2017), and randomForest (Liaw and Wiener 2002). Performance measures used to determine the best model include mean absolute percent error, root mean squared error (RMSE) and R^2. In addition to these performance measures, application of the models to the study area dataset were performed to gauge the reasonableness of the final estimates compared with the range of the observed data.

3.7.4.1 Data Fusion Input Data

As described previously, this research benefits from 889 pedestrian traffic counts from both permanent and SDC sites. These count observations are plotted in **Figure 3.18** to show the distribution of AADPT. Overdispersion was tested for using a Pearson Chi-square overdispersion test on the final Poisson model specification. The R performance package was used (Lüdecke et al. 2021), which revealed a dispersion ratio of 156.395 ($p < 0.001$), indicating overdispersion was present.

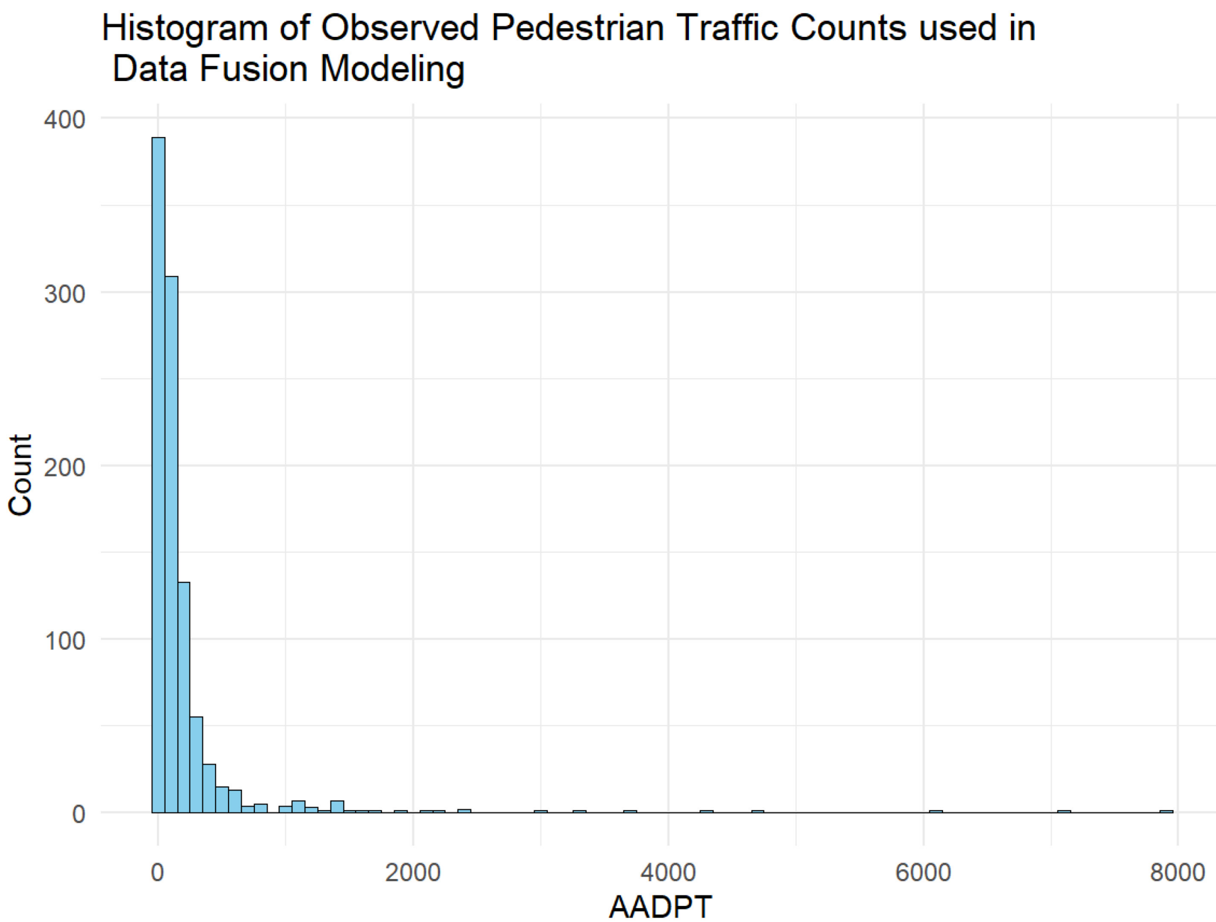


Figure 3.18: Distribution of AADPT Measures

To predict pedestrian traffic volumes this research explored many potential variables. These variables can be categorized into the following groups:

- Vehicle traffic and operational conditions
- Transportation network measures (centrality and accessibility)
- Socioeconomic characteristics
- Built environment
- Transportation and proxies for pedestrian activity

3.7.4.1 Data Fusion Methods

A number of statistical and machine learning approaches were tested to estimated AADPT as a function of explanatory variables (e.g., vehicle traffic and operational conditions, transportation network measures, etc.). For the statistical approach, a mixed-effects negative binomial model was applied. The negative binomial model was used due to significant overdispersion in the count data and the mixed-effects model was used due to the panel structure of the data.

A mixed-effects model is a combination of fixed- and random-effects, where the random-effects model between-group differences and the fixed-effects model within-group differences. By adding the random-effects, through a random intercept (based on a panel structure variable), the model allows for additional correlation among groups. The negative binomial model assumes pedestrian counts follow a negative binomial distribution, such that:

$$\boldsymbol{y_{iu}} \sim \mathbf{NB}(\boldsymbol{\mu_{iu}} = \mathbf{E}[\boldsymbol{y_{iu}}], \boldsymbol{\alpha}) \tag{3-1}$$

where NB denotes the negative binomial distribution, y_{iu} is the pedestrian volume count at intersection i in urban area u, and α is the dispersion parameter with $\mathrm{Var} = \mathrm{E}[y_{iu}] + \alpha\mathrm{E}[y_{iu}]^2$.

Under these assumptions, the conditional mean of y_{iu} relying on both fixed- and random-effects can be represented as:

$$\mathbf{log}[\boldsymbol{\mu_{iu}} = \mathbf{E}(\boldsymbol{y_{iu}})] = \boldsymbol{X_{iu}\beta} + \boldsymbol{z_{iu}b_u} + \boldsymbol{\varepsilon_{iu}} \tag{3-2}$$

where $\boldsymbol{X}_{iu}$ is a vector of explanatory variables (e.g., AADT, speed, intersection type, centrality, demographics), $\boldsymbol{\beta}$ is a vector of fixed-effects parameter estimates (includes intercept β_0), $\boldsymbol{z}_{iu}$ is a vector of variables with random-effects (the panel variable, such as urban area), $\boldsymbol{b}_u$ is a vector of random-effects parameter estimates for variables $\boldsymbol{z}_{iu}$ (e.g., urban area), and ε_{iu} is a gamma-distributed error term with mean 1 and variance α. Random-effects are used to assess between-urban area differences by estimating random intercepts. Through this approach, they avoid biased inferences on the fixed-effects.

Lastly, the random-effects are assumed to have a multivariate normal distribution, such that:

$$\boldsymbol{b_u} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\varphi}) \tag{3-3}$$

where N denotes the normal distribution and φ is a positive-definite variance-covariance matrix that accounts for correlation of the random-effects.

3.7.4.2 Statistical Model Results

This section summarizes the results from the model development process and includes the statistical model results for the negative binomial model as well as results from the cross-validation analysis. The statistical model was developed using a step-wise approach that specified different models based on different combinations of variables available to the research. To determine the final model, traditional statistical model diagnostics and cross-validation results were used. The cross validation uses a 10-fold partitioning process where training data was partitioned into 10 equal sized partitions. Then nine of the partitions were used to develop a model and estimate AADPT in the remaining partition. This was done 10 times so that each partition was held out once to determine each of the model's predictive capability.

The statistical model results using the negative binomial model with mixed effects are presented in Table 3-5. The model results show the estimated coefficient and the p-value for each variable selected for use in the model. At the bottom of the table, the marginal and conditional

R^2 values, using methods explained in Nakagawa et al. (2017), are presented. The marginal R^2 represents the variance explained by the fixed effects variables while the conditional R^2 represents the variance explained by the entire model including both fixed and random effects. In addition to the model specification table, the discussion features charts showing the marginal effects of select variables. These charts show the effects of select variables with all other model variables held constant (their mean values) to show how the selected variable impacts AADPT, all else being equal.

Table 3.6: Model Results from Negative Binomial with Mixed Effects

Variable Category	Predictors	Model 1 Estimate	Model 1 p	Model 2 Estimate	Model 2 p	Model 3 Estimate	Model 3 p	Model 4 Estimate	Model 4 p	Model 5 Estimate	Model 5 p
Vehicle traffic & operational conditions	Intercept	4.4833	**<0.001**	2.8412	**<0.001**	6.0383	**<0.001**	4.9529	**<0.001**	4.866	**<0.001**
Vehicle traffic & operational conditions	AADT (thousands)	0.0137	**<0.001**			0.0193	**<0.001**	0.0135	**<0.001**	0.013	**<0.001**
Vehicle traffic & operational conditions	Max Operating Speed	-0.0371	**<0.001**			-0.0593	**<0.001**	-0.0401	**<0.001**	-0.04	**<0.001**
Vehicle traffic & operational conditions	Is 4+leg Intersection (reference 3-leg)	0.3439	**<0.001**			0.3376	**<0.001**	0.3832	**<0.001**	0.383	**<0.001**
Transportation network measures (centrality & accessibility)	Network Centrality	0.0713	**<0.001**	0.0925	**<0.001**			0.0799	**<0.001**	0.079	**<0.001**
Transportation network measures (centrality & accessibility)	Population w/in 10-min Walk	0.0008	**<0.001**	0.001	**<0.001**			0.0008	**<0.001**	0.001	**<0.001**
Transportation network measures (centrality & accessibility)	Total Jobs w/in 5-min Walk)	0.0005	**<0.001**	0.0008	**<0.001**			0.0005	**<0.001**	0.001	**<0.001**
Transportation network measures (centrality & accessibility)	Arts or Entertainment Jobs w/in 5-min Walk	0.0093	**0.001**	0.0088	**0.004**			0.0091	**0.003**	0.01	**0.001**
Transportation network measures (centrality & accessibility)	City Park Area (m^2) w/in (5-min Walk	0.0001	**0.015**	0.0001	**0.04**			0.0001	**0.002**	0.0001	**0.001**
Transportation network measures (centrality & accessibility)	School Area (m^2) w/in 5-min Walk	0.0003	**0.014**	0.0002	**0.038**			0.0003	**0.01**	0.0003	**0.009**

Variable Category	Predictors	Model 1 Estimate	Model 1 p	Model 2 Estimate	Model 2 p	Model 3 Estimate	Model 3 p	Model 4 Estimate	Model 4 p	Model 5 Estimate	Model 5 p
Socioeconomic characteristics	Median income ($000)	-0.0055	**<0.001**	-0.0067	**<0.001**	-0.009	**<0.001**	-0.0059	**<0.001**	-0.004	**0.008**
Socioeconomic characteristics	National Walkability Index	0.0347	**0.002**	0.0693	**<0.001**			0.0571	**<0.001**	0.049	**<0.001**
Socioeconomic characteristics	Proximity to Interstate Traffic (1st decile)	-0.6873	**0.002**	-0.8328	**0.001**			-0.6074	**0.006**	-0.645	**0.003**
Socioeconomic characteristics	Proximity to Interstate Traffic (2nd decile)	-1.0816	**0.001**	-1.2871	**<0.001**			-0.9516	**0.004**	-1.011	**0.002**
Socioeconomic characteristics	Proximity to Interstate Traffic (3rd decile)	-0.6501	**0.028**	-0.7733	**0.016**			-0.5928	**0.049**	-0.624	**0.037**
Socioeconomic characteristics	Proximity to Interstate Traffic (4th decile)	-1.3042	**<0.001**	-1.392	**<0.001**			-1.3509	**<0.001**	-1.445	**<0.001**
Built environment	Proximity to Interstate Traffic (5th decile)	-1.477	**0.001**	-1.4372	**0.004**			-1.6112	**0.001**	-1.624	**0.001**
Built environment	Proximity to Interstate	-0.903	**0.023**	-1.3005	**0.002**			-0.767	0.055	-0.818	**0.041**

Variable Category	Predictors	Model 1 Estimate	Model 1 p	Model 2 Estimate	Model 2 p	Model 3 Estimate	Model 3 p	Model 4 Estimate	Model 4 p	Model 5 Estimate	Model 5 p
	Traffic (6th decile)										
Built environment	Proximity to Interstate Traffic (7th decile)	-0.7477	0.086	-1.2586	**0.006**			-0.5296	0.238	-0.552	0.214
Built environment	Proximity to Interstate Traffic (8th decile)	-1.2686	**<0.001**	-1.5381	**<0.001**			-1.3399	**<0.001**	-1.366	**<0.001**
Built environment	Proximity to Interstate Traffic (9th decile)	-1.3691	**<0.001**	-1.4788	**<0.001**			-1.3113	**<0.001**	-1.347	**<0.001**
Built environment	Proximity to Interstate Traffic (10th decile)	-1.7484	**0.035**	-1.2248	0.168			-1.8069	**0.035**	-1.839	**0.031**
Transportation and proxies for pedestrian activity	No-car households (%)	1.011	**0.039**	0.7614	0.15	1.7855	**0.001**			1.135	**0.023**
Transportation and proxies for pedestrian activity	Transit Stops within ≤100 m	0.0964	**0.008**	0.0846	**0.028**	0.1669	**<0.001**			0.102	**0.006**
Transportation and proxies for pedestrian activity	Commute by Transit (%)	2.2121	**0.021**	3.2476	**0.001**	2.6938	**0.014**			2.037	**0.031**
Transportation and proxies for pedestrian activity	101-1500 Strava Ped Activity w/	0.5895	**<0.001**	0.7792	**<0.001**	0.7429	**<0.001**				

Variable Category	Predictors	Model 1 Estimate	Model 1 p	Model 2 Estimate	Model 2 p	Model 3 Estimate	Model 3 p	Model 4 Estimate	Model 4 p	Model 5 Estimate	Model 5 p
	25 m (reference 0-100)										
Transportation and proxies for pedestrian activity	1500+ Strava Ped Activity w/ 25 m	1.0079	**<0.001**	1.28	**<0.001**	1.3341	**<0.001**				

Random Effects

σ^2	0.48	0.53	0.56	0.51	0.5
τ_{00}	0.14 $_{Urban\_Area}$	0.17 $_{Urban\_Area}$	0.16 $_{Urban\_Area}$	0.15 $_{Urban\_Area}$	0.13 $_{Urban\_Area}$
ICC	0.22	0.24	0.23	0.22	0.21
N	54 $_{Urban\_Area}$	54 $_{Urban\_Area}$	54 $_{Urban\_Area}$	54 $_{Urban\_Area}$	54 $_{Urban\_Area}$
Observations	868	889	868	868	868
Marginal R^2 / Conditional R^2	0.633 / 0.714	0.576 / 0.678	0.561 / 0.660	0.581 / 0.675	0.599 / 0.682

The models presented perform generally well in terms of the variation explained by the model and measured in the R^2 measures. The model parameters also generally perform as expected based on the literature. Five models are presented in the table to instruct the reader on the relative importance of each group of variables. Model 1 is the full model with measures from each category of variables, while Model 2 does not include vehicle traffic and operational condition variables. Model 3 and Model 4 do not have transportation network measures and built-environment measures, respectively. Model 5 has the Strava pedestrian measures removed to highlight the model performance change without this data input.

For the variables in the vehicle traffic and operational conditions category, coefficient signs are consistent across model specifications. Vehicle AADT is positively associated with AADPT, while measures of max operational speed are negatively associated with the response variable. The impact of max operational speed on AADPT is shown in **Figure 3.19**. Holding other factors constant, 4-plus-leg intersections have higher pedestrian volumes than 3-leg intersections, likely because they offer more approaches and crossing movements through the intersection.

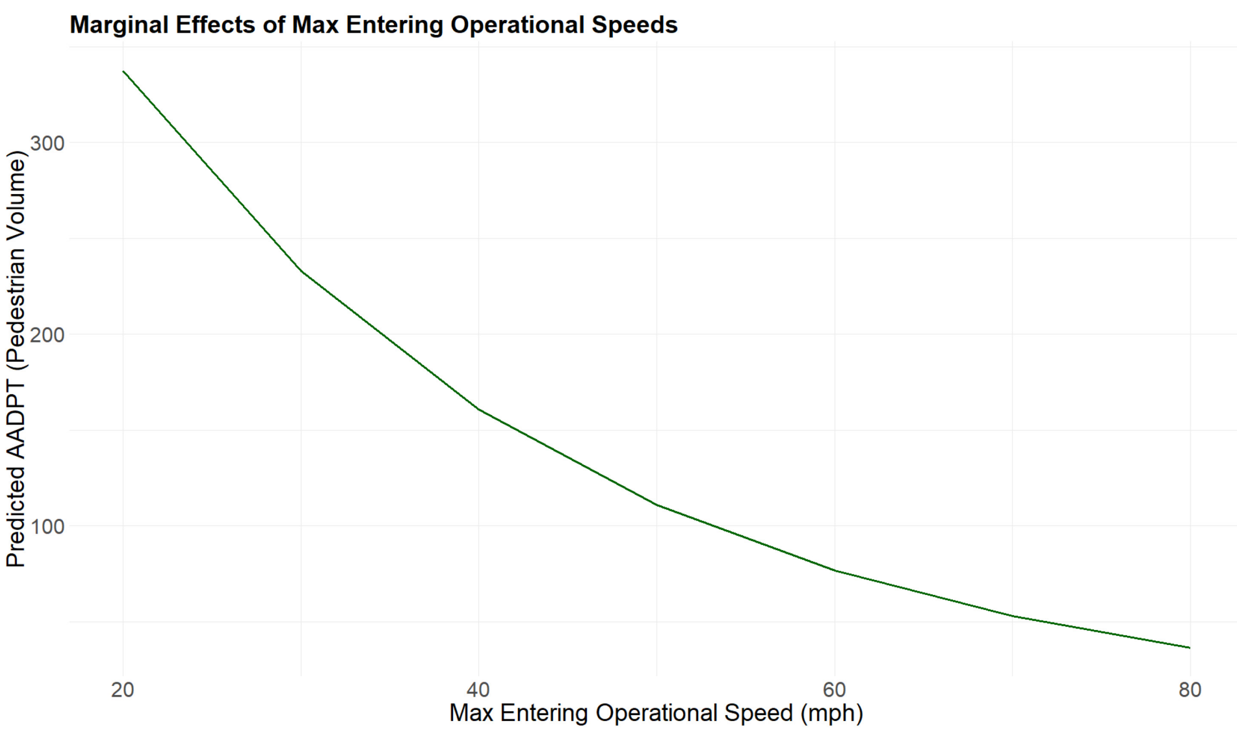


Figure 3.19: Max Operational Speed Marginal Effects on AADPT

For the variables in the transportation network measures, which include measures of network centrality and accessibility, six variables were used, including overall network centrality, total population within a 10-minute walk, total jobs within a five-minute walk, arts and entertainment jobs within a five-minute walk, city parks within five-minute walk, and schools within a five-minute walk. As mentioned previously, network centrality is a measure of the intersection's importance when traversing the network. High betweenness centrality means the intersection is used often when traversing the network from all nodes of the network to all other nodes. In the final model, the centrality variable is positively associated with pedestrian volume, as would be expected. Access to jobs, schools, and parks are also positively associated with AADPT, as would be expected. Marginal effects for access to schools within a five-minute walk are shown in **Figure 3.20**.

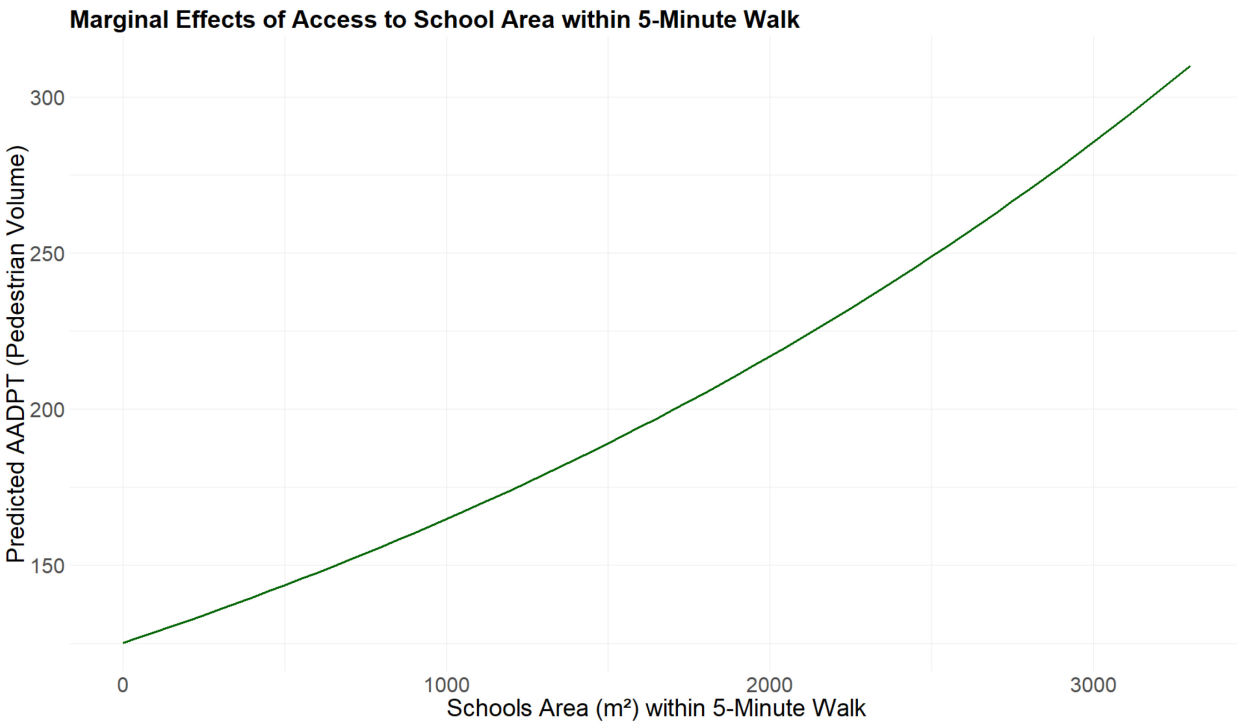


Figure 3.20: Access to School Area (m^2) Marginal Effects on AADPT

Though many different measures of socioeconomics were tested, including poverty rate, percentage of population using supplemental nutritional assistance program benefits, among others, median income was chosen as it resulted in a better performing model. The effect of median income on AADPT is demonstrated in **Figure 3.21**, which shows that as median income in the block group where the intersection resides increases, AADPT decreases.

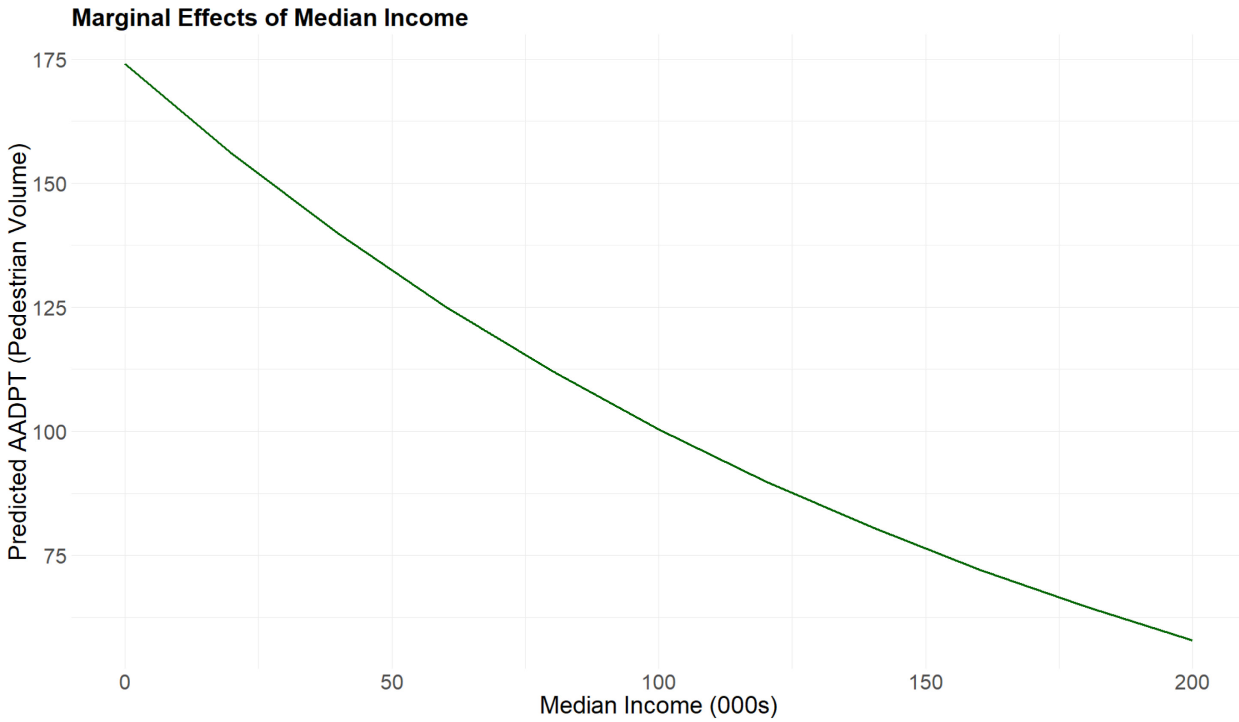


Figure 3.21: Median Income Marginal Effects on AADPT

The next category of characteristics used in the final pedestrian data fusion model include built-environment variables. Built-environment variables include the National Walkability Index from the Smart Location Database developed by the EPA. The National Walkability Index is not a measure of safety or comfort but rather a regionally indexed score that uses information about the diversity of land uses, including employment, residential, and the density of intersections and transit. These index values are relative to the urban area and should not be necessarily compared across states or urban areas. In the final model, this variable is associated with higher AADPT, all else being equal. The second variable used from the built-environment category is the proximity to interstates. Proximity to interstates uses walk sheds to interstate segments, is normalized for each urban area using deciles, and is operationalized in the model as a factor variable. The results somewhat vary by decile, but generally the higher the proximity to the interstate, the lower the AADPT.

The last set of characteristics used in the model aim to reflect measures of transportation and are likely proxies for pedestrian activity. These measures include the percentage of households in the associated Census block group that do not own a vehicle, the number of transit stops within 100 meters (328 feet), the percentage of workers in the Census block group that use transit to commute to work, and the number of walk, job, and run activities recorded by the Strava smart phone application.

The percentage of households without a vehicle are associated with higher AADPT, as is the number of transit stops within 100 meters (328 feet) of the intersection. When intersections are in locations with higher percentage of workers commuting by transit, the model shows that AADPT is also higher in those areas. And lastly, the measures of Strava walk, job, and run trips are also associated with higher AADPT. Compared to intersections with 0-100 Strava trips, intersections with 101-1500 Strava trips have higher AADPT and intersections with 1500 or more Strava trips have an even higher value of expected pedestrian volume.

This research took care to assess collinearity between variables used in the statistical model and uses a threshold of 0.70 for the Pearson correlation coefficient to determine whether variables should be dropped (Dorman et al 2013). **Figure 3.22** shows the Pearson correlation coefficients for all the variables in the final model.

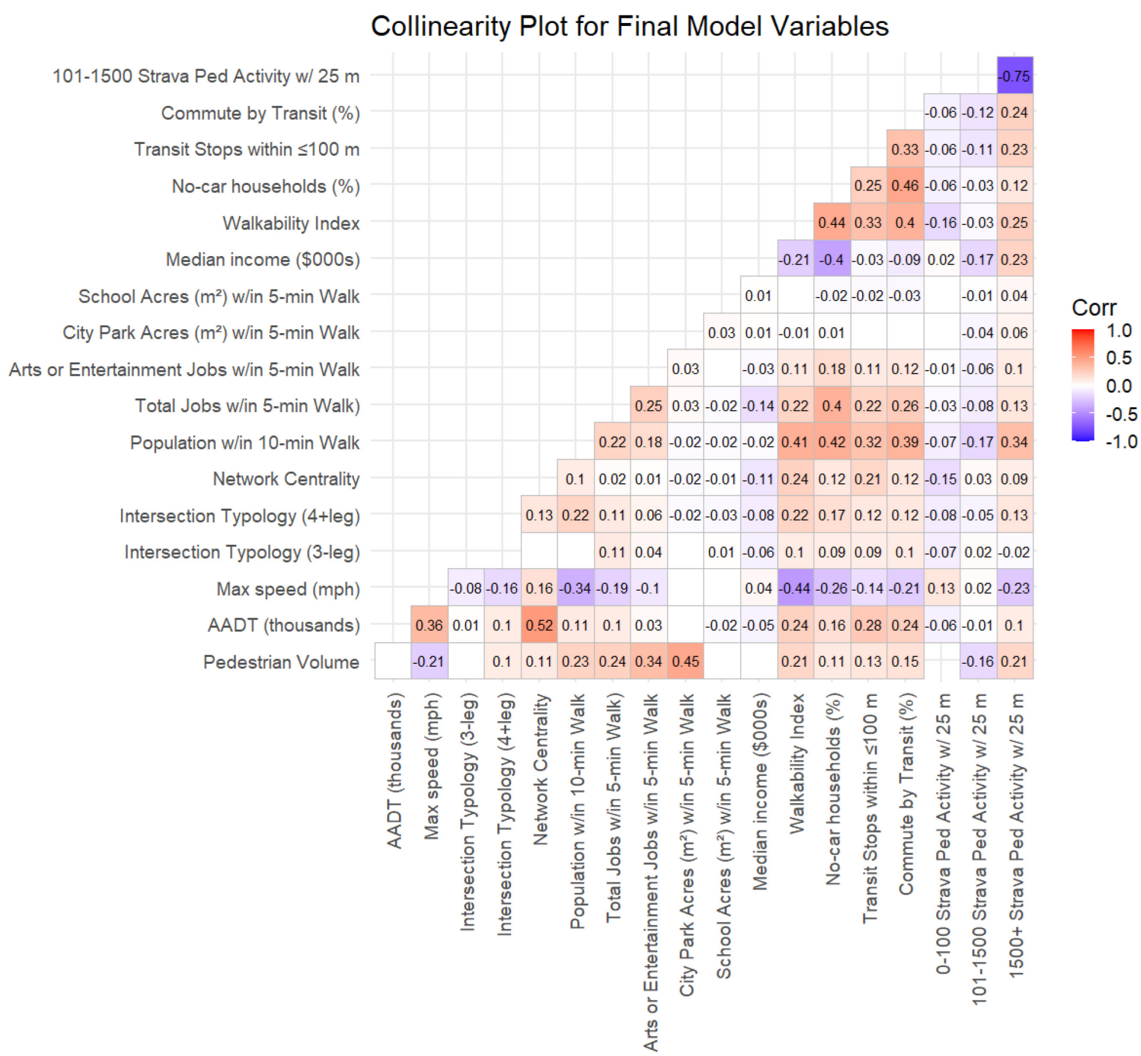


Figure 3.22: Collinearity Plot for Final Model Variables

The statistical models presented aim to show the effects of a variety of independent variables available to the research team and show that AADPT is a predictable quantity. The models presented appear to perform consistently across specifications and have a relatively high pseudo R^2. This research also tested machine learning approaches to predict AADPT, which is covered in the next subsection.

3.7.5 Machine Learning Model Specification

Statistical models are useful for demonstrating the effects of independent variables on the response variable, which are helpful for inference but may not be as durable for prediction compared to machine learning approaches. Therefore, this research develops machine learning models using random forest and XgBoost configurations, as these machine learning techniques have been shown to do better at predicting active transportation counts compared to statistical models (Kaiser et al. 2025; Roll 2023; ODOT 2021) due to their ability to handle nonlinearity and high-order effects in a way that statistical models cannot.

The machine learning training was performed in R and used two different packages for the XGBoost and random forest models, including the caret (Kuhn 2008) package and the ranger (Wright and Ziegler) package. These machine learning models used RMSE and mean average error to test against outliers. The XGBoost regression model's hyper parameters were tuned using repeated 5-fold cross validation over a prespecified grid with boosting rounds from 50 to 250 at steps of 50. The learning rate (eta) was 0.025, 0.05, 0.10, 0.30 and tree depth was 2 to 6. Two random forest packages were tried with the ranger implementation using a maximum of 2,500 trees while the caret package implementation of the random forest also used a maximum of 2,500 trees and a grid search to tune the hyper parameters.

Variables offered to the machine learning algorithms included those used in the statistical models, as well as also other variations of those measures. Additional measures included different walk and drive-shed distances to amenities like jobs, parks, and schools, in addition to various measures of operational speeds and additional Census and SLD data elements. After initial model runs, variables were trimmed and new models estimated for final specification.

The following subsection features a presentation of variables used in the machine learning model in terms of variable importance. Since machine learning models do not produce parameter estimates in the same way that traditional statistical models do, many readers are left with questions about how a particular machine learning model is working. To fill this gap, variable importance is used to explain how the model is working by showing which variables are 'most important' to the model's predictive capability.

For tree-based machine learning methods like random forest and boosted models, variable importance is measured using a measure of Gini impurity, which is a measure of the number of times a feature is used to make a node split for a given tree in a given forest (Guyon 2003). In most calculations of Gini impurity, the sum of the Gini decreases for each tree in the forest that is aggregated for each time that feature is chosen as a splitting variable. The sum of this aggregation is then divided by the count of trees in the forest to give an average measure. The scale of the final measure can be ignored and instead a focus on the comparison to other variables is what is most relevant. The next section documents feature importance as a way to diagnose how the model utilizes input features.

3.7.6 Machine Learning Model Results

Once the variable selection process was completed, a separate independent 10-fold cross validation was performed where the 868 AADPT observations were partitioned into 10 chunks. In an iterative fashion, nine partitions were used to train a model and then used to predict AADPT in a 10th partition, or test data. This was performed 10 times using different partitions and error calculations for all modeling approaches.

Error was calculated using two measures, including absolute percent error and root mean squared error (RMSE). These error measures were calculated using the following equations:

1. Absolute Percent Error (APE)

$$\textbf{Absolute Percent Error} = \left|\frac{(AADPT_{Est} - AADPT_{Obs})}{AADPT_{Obs}}\right| \tag{3-4}$$

2. Root Mean Squared Error (RMSE)

$$\textbf{RMSE} = \sqrt{(AADPT_{Est} - AADPT_{Observed})^2} \tag{3-5}$$

In addition to the error measures, R^2 was also calculated to show correlation and how well the estimates generally align with the observed. The results of this cross-validation are presented in Table 3-6. The measures of error and model performance show that the random forest models performed best when considering RMSE and R^2, but that the XgBoost specification performed better when considering median APE. The absolute error measures appear large but many of the AADPT measures were very small which makes percent error measures appear very large.

Table 3.7: 10-Fold Cross-Validation Results for Machine Learning and Statistical Models

Model Specification	RMSE	Absolute % Error Mean	Absolute % Error Median	Absolute % Error 1st Quartile	Absolute % Error 3rd Quartile	R^2	N
Negative Binomial	434	181%	70%	32%	176%	0.51	868
Poisson	397	198%	75%	34%	191%	0.56	868
Random Forest (Ranger)	336	241%	76%	31%	235%	0.67	868
Random Forest (Caret)	337	262%	79%	32%	242%	0.68	868
XgBoost	365	239%	65%	28%	203%	0.59	868

Figure 3.23 shows the correlation between the estimated AADPT and the observed measures. These results show visually the information presented in Table 3.7.

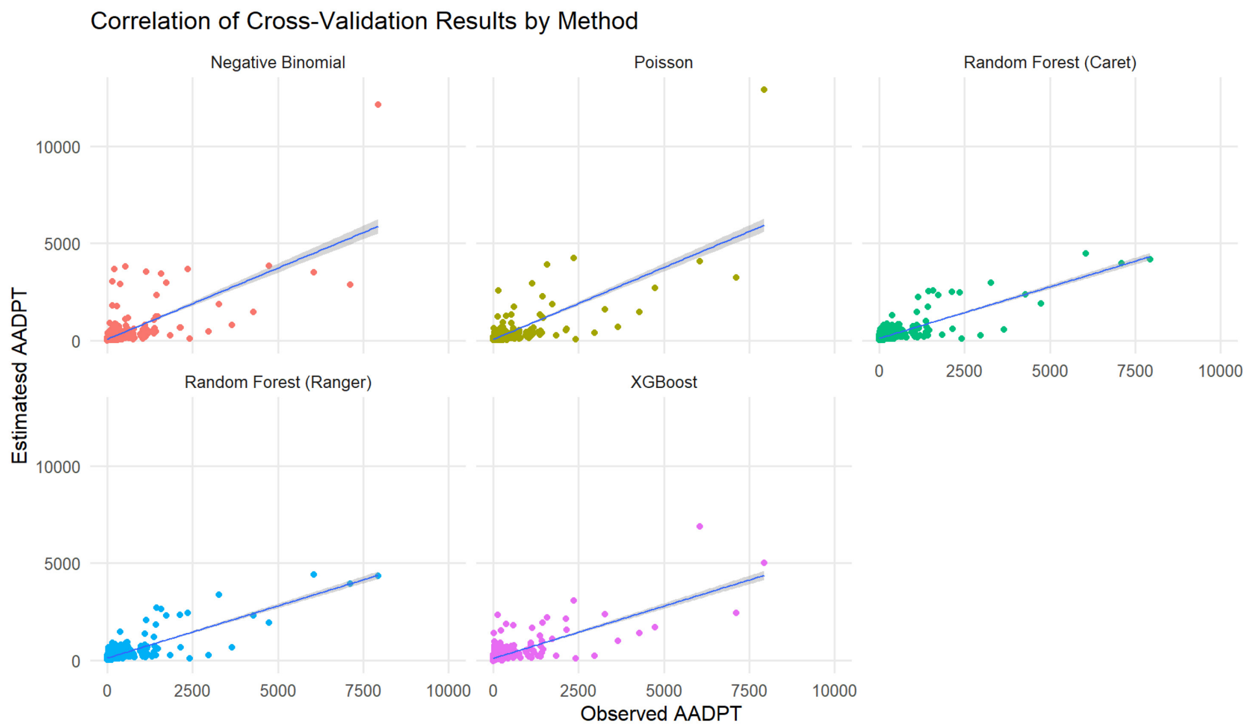


Figure 3.23: Correlation of Cross-Validation Results by Method

To give a sense of how the machine learning models work and which variables are used in each model, **Figure 3.24** presents the top 30 most important variables. **Figure 3.24** shows the name of the variable, the category in which it belongs (by color) and the relative importance for that machine learning model. Significant overlap exists in the top 30 variables selected in both random forest model specifications. The top five variables are identical for both models with many variables present in both models but ranked in a slightly different order. These variables include many measures of walk access to jobs and access to amenities (like parks and schools), travel and transportation measures (e.g., vehicle ownership, proportion of workers commuting by walking), and Strava trips. The top 30 most important variables in the final XgBoost model had similarities to the random forest models, but one notable difference was the number of Smart Location Database variables utilized in this model specification.

The random forest model using the caret package appears to work best given the cross-validation results, and as discussed later, performs best for reasonable distribution of values when applied to the entire state. Because this is the selected model for application, the chart featured in **Figure 3.25** shows all variable importance measure for all variables used in the final model. These variables are shown in a box and whisker plot to give readers a sense of the number of variables included in the model, the types of variables used, and the relative impact of the variables on the model from the distribution of values.

There are a number of different walk access measures, previously explained in detail, that include measures of access to different kinds of jobs and amenities (e.g., parks and schools). Access to these amenities based on driving time were also available to the machine learning algorithm, but were all within short thresholds (e.g.,10 minutes) and may be representing some type of proximity measure that varies from the walk access measures. All of the walk access derived measures, including access to jobs, population and, amenities (schools, parks, and open space) demonstrated higher variable importance measures than driving-based measures.

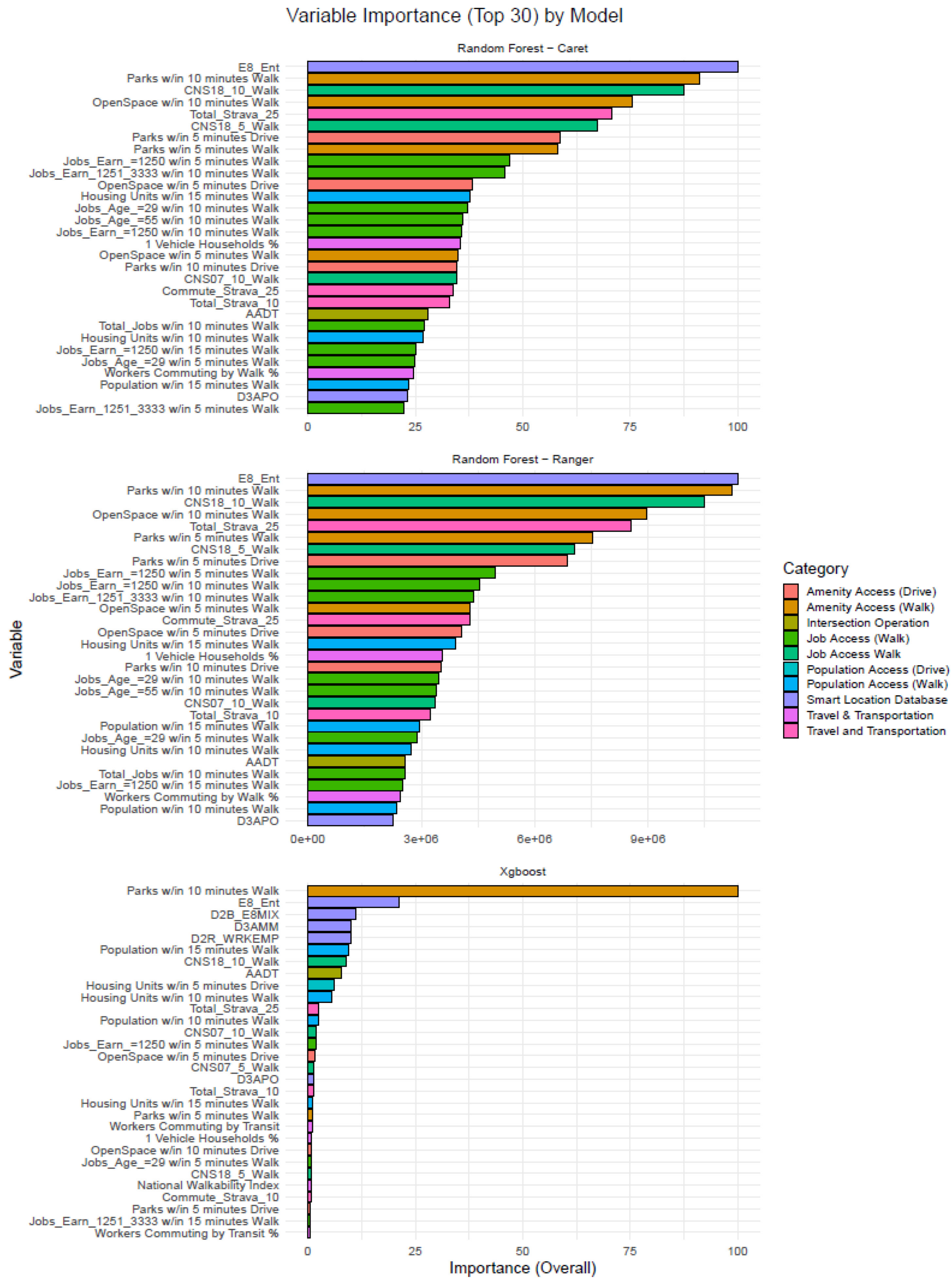


Figure 3.24: Variable Importance (Top 30) by Model

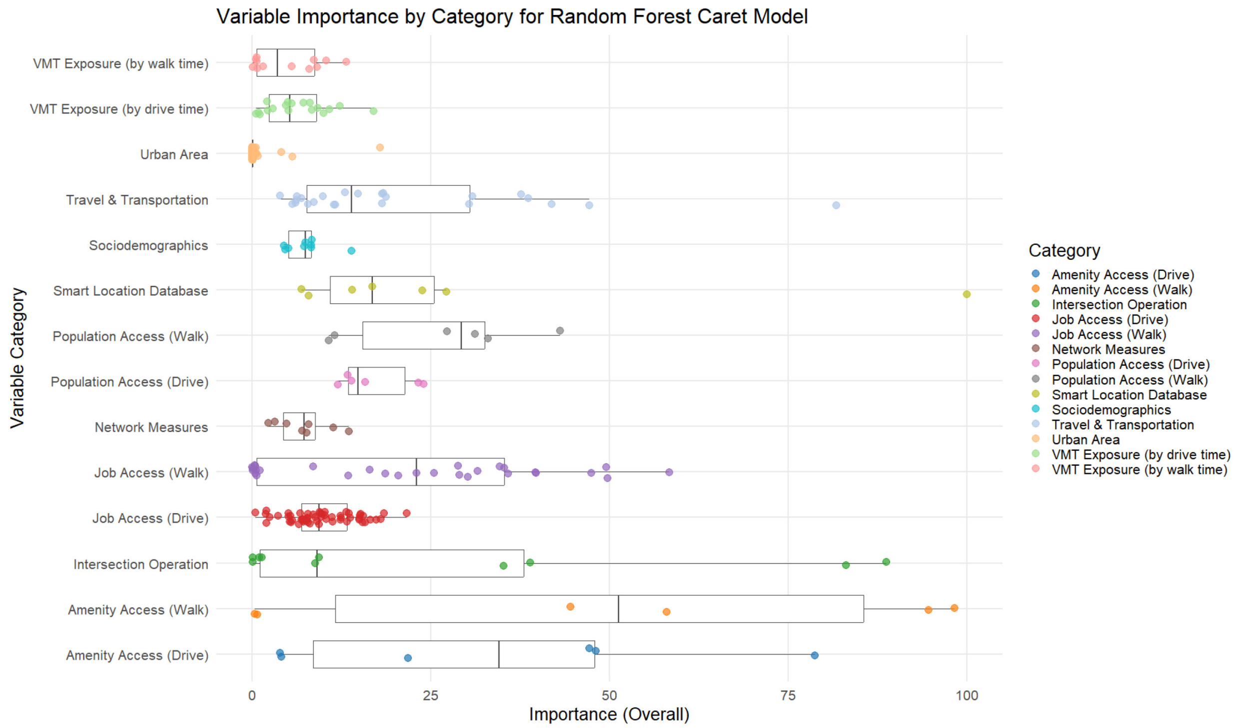


Figure 3.25: Variable Importance by Category for Random Forest Caret Model

3.7.7 Application of Data Fusion Models

The previous subsections describe the development of both statistical and machine learning data fusion models for estimating pedestrian traffic counts. This subsection describes the application of the models and the process to extend the models to estimating AADPT to all intersections in the state. The cross-validation results suggested the machine learning models performed best in terms of RMSE and R^2, though the statistical models performed okay and consideration of their application was initially still warranted. To check for reasonableness, each model was applied to the entire set of intersections and summary statistics calculated. These summary statistics are presented in Table 3.8. Table 3.8 shows the minimum, mean, and maximum values predicted by the different models in addition to percentile values to give readers a sense of how the models predict pedestrian volume. The table highlights the issues with the statistical model where the maximum value predicted is in the hundreds of millions while the other models' maximum values are 7,384 or less, with the random forest specification using the caret package yielding the lowest maximum value of just under 6,000 and the ranger-based random forest model estimating a similar maximum value of 6,021. The XGBoost machine learning model predicted negative values, which is not a realistic quantity. Based on the random forest model producing the most reasonable values when applied and the results from its cross-validation process, this model specification was chosen as the final model to predict AADPT for use in the safety performance functions.

Table 3.8: Summary Statistics of Results from Application of Data Fusion Models

Model	Min	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Mean	Max
Negative Binomial	0.954	44.6	103	245	1610	79,892	649,564,547
Random Forest (caret)	8.01	90.5	169	331	1022	291	5912
Random Forest (ranger)	6.22	109	205	437	1406	380	6021
XGBoost	-61.6	86.3	168	367	2355	447	7384

4.0 PEDESTRIAN SAFETY PERFORMANCE FUNCTION DEVELOPMENT

4.1 METHODOLOGY

Using the pedestrian crash frequency data at urban intersections, a series of pedestrian crash frequency models were estimated. The intent was to arrive at intersection typology-specific and location-specific pedestrian safety performance functions that describe pedestrian crash frequency behavior at such levels. Intersections were disaggregated by the following:

- 4-leg and 3-leg intersections
- Statewide, Portland City Limits, Portland Urban Area, and Portland/Corvallis Urban Areas
    - Local or regional locations were selected based on availability and confidence of particular data, such as pedestrian crossing information and lighting information.
- State (ODOT) System and Non-State System
- Signalized and Unsignalized
- Presence of a marked crossing and no marked crossing

This resulted in a total of eight models for each disaggregation. For example, the 4-leg signalized intersections with marked crossings sample had a statewide model, a Portland City Limits model, a Portland Urban Area model, and a Portland/Corvallis Urban Areas model, and for each one a model was estimated for intersections on the state system and intersections on the non-state system. A flowchart of this workflow is given in Figure 4.1.

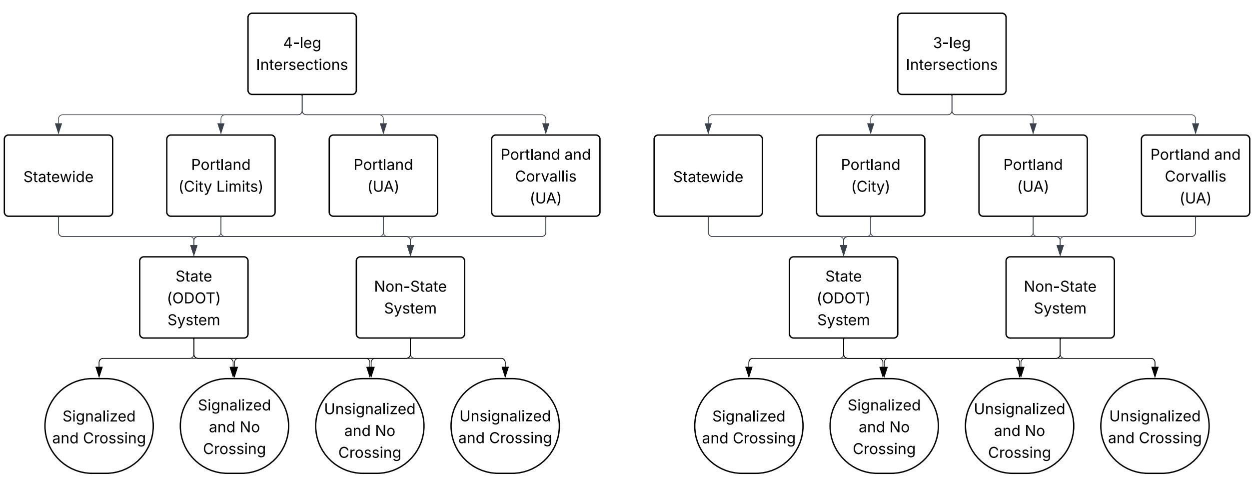


Figure 4.1: Pedestrian Crash Frequency Model Workflow

For this work, all pedestrian crash types and severities were considered. Additionally, the models presented consider exposure only (vehicle volume and pedestrian volume). Although various models were estimated that considered information from the data conflation process, metrics showed that despite some improvements in model fit (per the log-likelihood values), overall, there were not substantial improvements in how the model estimated expected crash frequencies. For this reason, the trade-off between usability and complexity of model specifications was carefully considered, where the overall usability of the models was chosen over complex model specifications. This led to all crash frequency models considering only vehicle volume and pedestrian volume in model specifications. Some examples of these comparisons are given in Appendix A.

To develop the pedestrian crash frequency models, count-data models were used based on the nature of the dependent variable; that is, non-negative integer counts. This work considered two common modeling frameworks for such variables: (1) the Poisson model and (2) the negative binomial model. For each model, a Poisson model was estimated first and a test for overdispersion was conducted. If the overdispersion test suggested the presence of significant overdispersion, a negative binomial model was estimated.

The following subchapters describe the Poisson modeling framework, the overdispersion test, and the negative binomial modeling framework.

4.1.1 Poisson Model

The first model considered was a Poisson model. In a Poisson model, the dependent variable y_i (pedestrian crash frequency) is drawn from a Poisson population with parameter λ_i, where the Poisson model can be formulated as (Greene, 2018):

$$\boldsymbol{P(y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}} \tag{4-1}$$

where $P(y_i)$ is the probability of intersection i having y_i crashes and λ_i is the Poisson parameter for intersection i. The Poisson parameter is also equal to the expected number of crashes on intersection i, $\mathrm{E}[y_i]$. Poisson models are estimated by specifying the Poisson parameter as a function of explanatory variables such that (Greene, 2018):

$$\boldsymbol{\lambda_i = e^{\beta X_i}} \tag{4-2}$$

where $\boldsymbol{X}_i$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of parameters to be estimated. Parameter estimates in the Poisson model are estimated using the following log-likelihood function (Greene, 2018):

$$\boldsymbol{LL(\beta) = \sum_{i=1}^{n}\left[-e^{\beta X_i} + y_i \beta X_i - \ln(y_i!)\right]} \tag{4-3}$$

To interpret the effect of a variable, incidence rate ratios or marginal effects are used. Incidence rate ratios represent the effect of an explanatory variable on the dependent variable as a multiplicative change. Marginal effects are the effect of explanatory variable x_i on the dependent variable due to a one-unit increase in x_i in the units of the dependent variable. In the Poisson model, the Poisson parameter is the dependent variable (see Eq. (4.2)) such that incidence rate ratios and marginal effects are computed as (Greene, 2018):

$$\mathbf{IRR} = \boldsymbol{e^{\beta}} \tag{4-4}$$

$$\mathbf{ME}_{\mathbf{X}_{ik}}^{\boldsymbol{\lambda_i}} = \frac{\partial \boldsymbol{\lambda_i}}{\partial \mathbf{X}_{\boldsymbol{ik}}} = \boldsymbol{\beta_k} \boldsymbol{e}^{(\boldsymbol{\beta}\mathbf{X}_i)} \tag{4-5}$$

For each model estimated, a Poisson model was estimated first. However, Poisson models have a limiting assumption that can cause issues if under- or overdispersion is present. Specifically, the Poisson model operates under the assumption that the mean and variance are equal, $\mathrm{E}[y_i] = \mathrm{Var}[y_i]$. If this equality is not met, the data is underdispersed ($\mathrm{E}[y_i] > \mathrm{Var}[y_i]$) or overdispersed ($\mathrm{E}[y_i] < \mathrm{Var}[y_i]$). If significant under- or overdispersion is present and not accounted for, parameter estimates will no longer be unbiased and standard errors of the estimates will be incorrect, thus leading to inflated t-statistics and scenarios in which the null hypothesis is being rejected when it should not be. To determine if the Poisson assumption holds, a dispersion test must be conducted.

4.1.2 Testing for Dispersion

There are a variety of methods for determining if significant dispersion is present, some of which depend on the software being used (e.g., the significance of the dispersion parameter is given in model specifications).

This work applies a common technique using an ordinary least squares approach proposed by Cameron and Trivedi (1990), which is consistent with the LaGrange multiplier test proposed by Greene (2003). This has since been adopted and versions of it presented by Greene (2018) and Washington et al. (2020).

This method considers a hypothesis test based on the assumption that under the Poisson model $(y_i - \mathrm{E}[y_i])^2 - \mathrm{E}[y_i]$ has a mean of zero, with $\mathrm{E}[y_i]$ being the predicted crash frequency $\hat{\lambda}_i$. The competing hypothesis are then:

$$\boldsymbol{H_o}\text{: } \mathbf{Var}[\boldsymbol{y_i}] = \mathbf{E}[\boldsymbol{y_i}] \tag{4-6}$$

$$\boldsymbol{H_A}\text{: } \mathbf{Var}[\boldsymbol{y_i}] = \mathbf{E}[\boldsymbol{y_i}] + \boldsymbol{\alpha g}(\mathbf{E}[\boldsymbol{y_i}]) \tag{4-7}$$

where $g(\mathrm{E}[y_i])$ is a function of the predicted crash frequency. To conduct the test, a simple linear regression model is estimated such that Z_i is regressed on W_i, where:

$$\boldsymbol{Z_i} = \frac{(\boldsymbol{y_i} - \mathbf{E}[\boldsymbol{y_i}])^2 - \boldsymbol{y_i}}{\mathbf{E}[\boldsymbol{y_i}]\sqrt{2}}$$

(4-8)

$$\boldsymbol{W_i} = \frac{\boldsymbol{g}(\mathbf{E}[\boldsymbol{y_i}])}{\sqrt{2}}$$

(4-9)

Based on the null hypothesis (i.e., equidispersion), the resulting test statistics have chi-square distributions with one degree of freedom. If the test statistics are greater than the chi-square critical value for one degree of freedom, the null hypothesis is rejected in favor of the alternative (i.e., dispersion is present).

Throughout this work, it was determined that some disaggregate models had significant overdispersion. When this occurred, a negative binomial model was estimated.

4.1.3 Negative Binomial Model

The functional form of the negative binomial model remains the same as the Poisson model, with one important addition (Greene, 2018; Washington et al., 2020):

$$\boldsymbol{\lambda_i} = \boldsymbol{e}^{\boldsymbol{\beta X_i + \varepsilon_i}}$$

(4-10)

where ε_i is a Gamma-distributed disturbance term with mean 1 and variance α. The addition of the Gamma-distributed disturbance term, ε_i, allows the variance to differ from the mean, hence overcoming the limiting assumption of the Poisson model. This is accomplished by defining the variance as (Greene, 2018; Washington et al., 2020):

$$\mathrm{Var}[y_i] = \mathrm{E}[y_i]\big[1 + \alpha\mathrm{E}[y_i]\big] = \mathrm{E}[y_i] + \alpha\mathrm{E}[y_i]^2$$

(4-11)

where α is known as the dispersion parameter. Typical ranges and interpretations for α include:

- $\alpha < 0$ indicates underdispersion. This requires specific variants of Poisson and negative binomial models, or in some cases, hybrid or two-part models.
- $0 < \alpha < 1$ indicates moderate overdispersion. In most scenarios, this can be handled by estimating a negative binomial model.
- $\alpha > 1$ indicates substantial overdispersion. In most scenarios, this can be handled by estimating a negative binomial model.

By adding the Gamma-distributed error term, ε_i, and dispersion parameter α, the probability density function is now representation of the negative binomial model (Anastasopoulos and Mannering, 2009; Washington et al., 2020):

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

(4-12)

where $\Gamma(\cdot)$ is a gamma function, which is used as the likelihood function and used to generate the log-likelihood function for parameter estimation.

To interpret variable effects, the same approaches are applied as for the Poisson model.

4.2 SAFETY PERFORMANCE FUNCTION RESULTS

This chapter presents the results from the pedestrian crash frequency models. All model results are presented as summaries, while full model specifications are given in Appendix B (4-leg intersections) and Appendix C (3-leg intersections). The corresponding cumulative residual (CURE) plot for each model is also presented. Following the presentation of model summaries by location, intersection typology, and jurisdiction, a comparison of model estimates across disaggregate models is provided. Lastly, a series of predicted pedestrian crash frequency plots are presented based on model estimates, each of which consider various pedestrian volumes over a range of traffic volumes.

4.3 4-LEG SIGNALIZED INTERSECTIONS

The following subchapters present model results for all 4-leg signalized intersection models. The models are presented in the following two groups: (1) a marked crossing is present and (2) no marked crossing is present.

4.3.1 4-Leg Signalized Intersections with Marked Crossings

The following subchapters will present model specifications based on location and intersection jurisdiction.

4.3.1.1 State System

A summary of model results by location for 4-leg signalized intersections on the state system is given in Table 4-1, while full model specifications are in Appendix B (Table B.1 to Table B.4). Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

Results were consistent across locations, where average annual daily traffic and annual average daily pedestrian traffic are expected to increase pedestrian crash frequency. In the statewide model and the Portland/Corvallis Urban Areas model, the effects for both variables were significant. In the Portland City Limits model, only the effect of annual average daily traffic was significant, and in the Portland Urban Area model, only the effect of annual average daily pedestrian traffic was significant.

The CURE plots for each model are shown in Figure 4.1. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash values). Overall, the four models have good fit over the range of all crash values, but there could be some improvement when fitting small or large crash values.

Table 4.1: Summary of 4-Leg Signalized Intersection with Marked Crossing Models on State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	*	464
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	464
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	*	109
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	NS	109
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	NS	240
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	*	240
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	260
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	260

* Significant with at least 90% confidence
NS = not significant

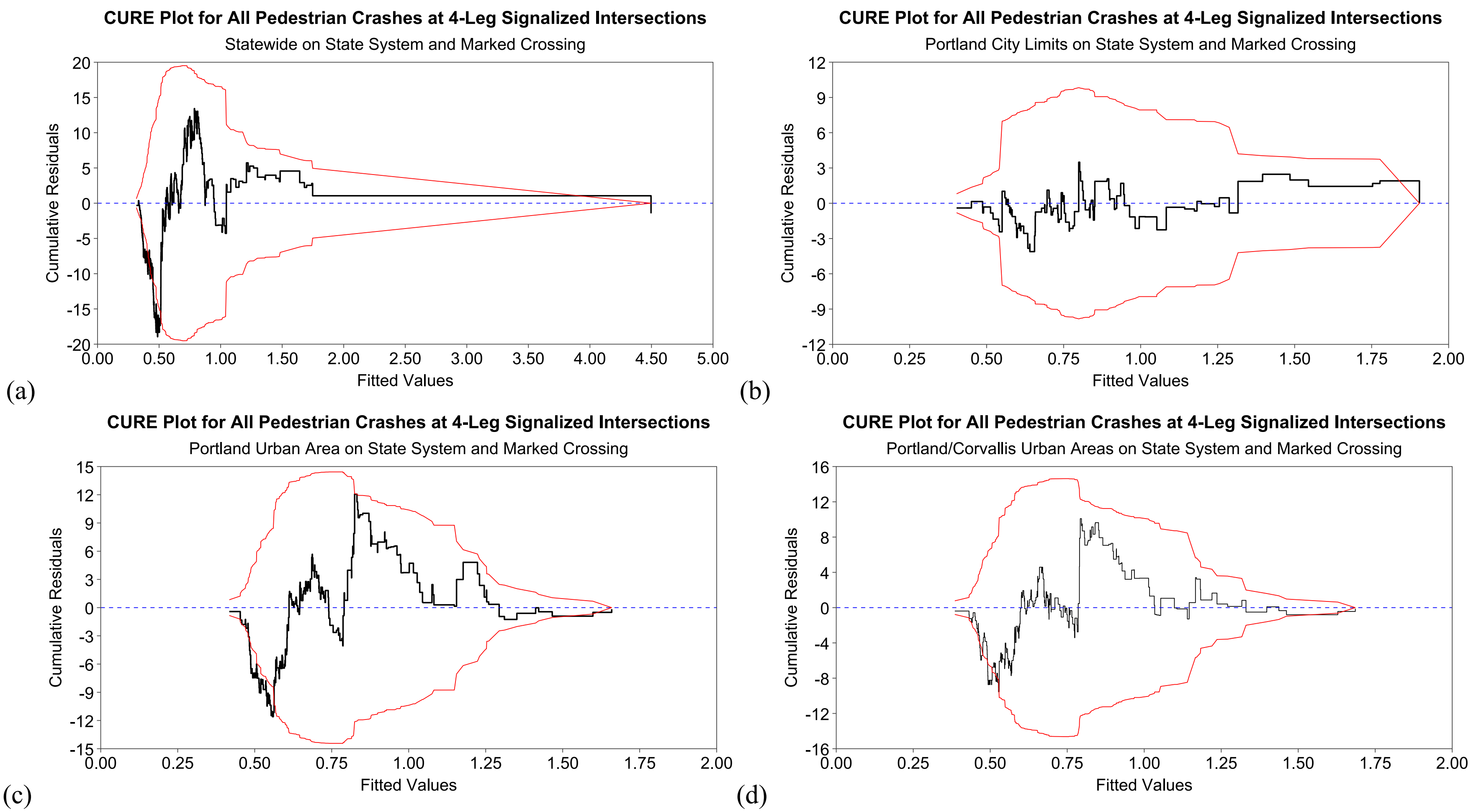


Figure 4.1: CURE Plots for 4-Leg Signalized Intersections and Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on State System

4.3.1.2 Non-State System

A summary of model results by location for 4-leg signalized intersections on the non-state system is given in Table 4-2, while full model specifications are in Appendix B (Table B.5 to Table B.8). Table 4-2Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

Results were the same across all models, where average annual daily traffic and annual average daily pedestrian traffic are expected to increase pedestrian crash frequency. In each model, the effects of the variables were statistically significant.

The CURE plots for each model are shown in Figure 4.2. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash values). These models fit the crash frequency values well at small and midrange values, and do not exceed the fitted bounds but for larger values.

Table 4.2: Summary of 4-Leg Signalized Intersection with Marked Crossing Models on Non-State System

Variable	Geographic Area	Effect	Sig.	***n***
Average Annual Daily Traffic (Thousands)	Statewide	+	*	1,235
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	1,235
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	*	563
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	*	563
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	*	844
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	*	844
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	868
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	868

* Significant with at least 90% confidence
NS = not significant

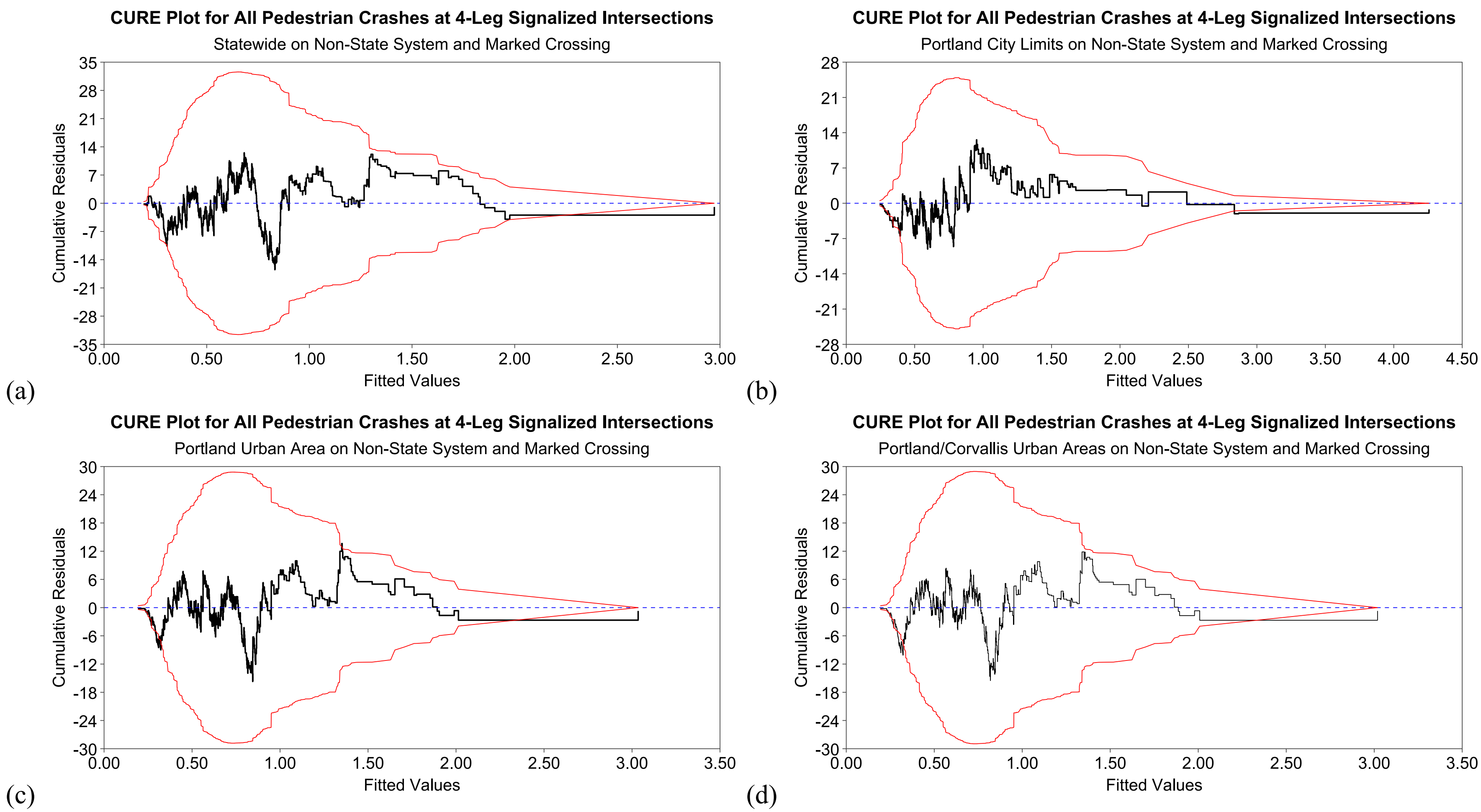


Figure 4.2: CURE Plots for 4-Leg Signalized Intersections and Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on Non-State System

4.3.2 4-Leg Signalized Intersection with No Marked Crossing

The following subchapters will present a model summary based on location and intersection jurisdiction.

4.3.2.1 State System

A summary of model results by location for 4-leg signalized intersections with no marked crossing on the state system is given in Table 4-3, while full model specifications are in Appendix B (Table B.9 to Table B.12). Table 4-3Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

As with previous 4-leg intersection models, results were consistent across locations. In the statewide model, average annual daily traffic and average annual daily pedestrian traffic were both significant contributing factors; this was also true in the Portland/Corvallis Urban Areas model. In the Portland City Limits model, only average annual daily pedestrian traffic was a significant contributing factor, while average annual daily traffic was the only significant contributing factor in the Portland Urban Area model. When significant, average annual daily traffic and average annual daily pedestrian traffic are expected to increase pedestrian crash frequency.

The CURE plots for each model are shown in Figure 4.3. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash values). Referring to Figure 4.3, each model performs well for most crash values, where there is room to improve fitted values for higher crash frequencies.

Table 4.3: Summary of 4-Leg Signalized Intersection with No Marked Crossing Models on State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	*	383
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	383
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	NS	17
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	*	17
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	*	78
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	NS	78
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	87
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	87

* Significant with at least 90% confidence
NS = not significant

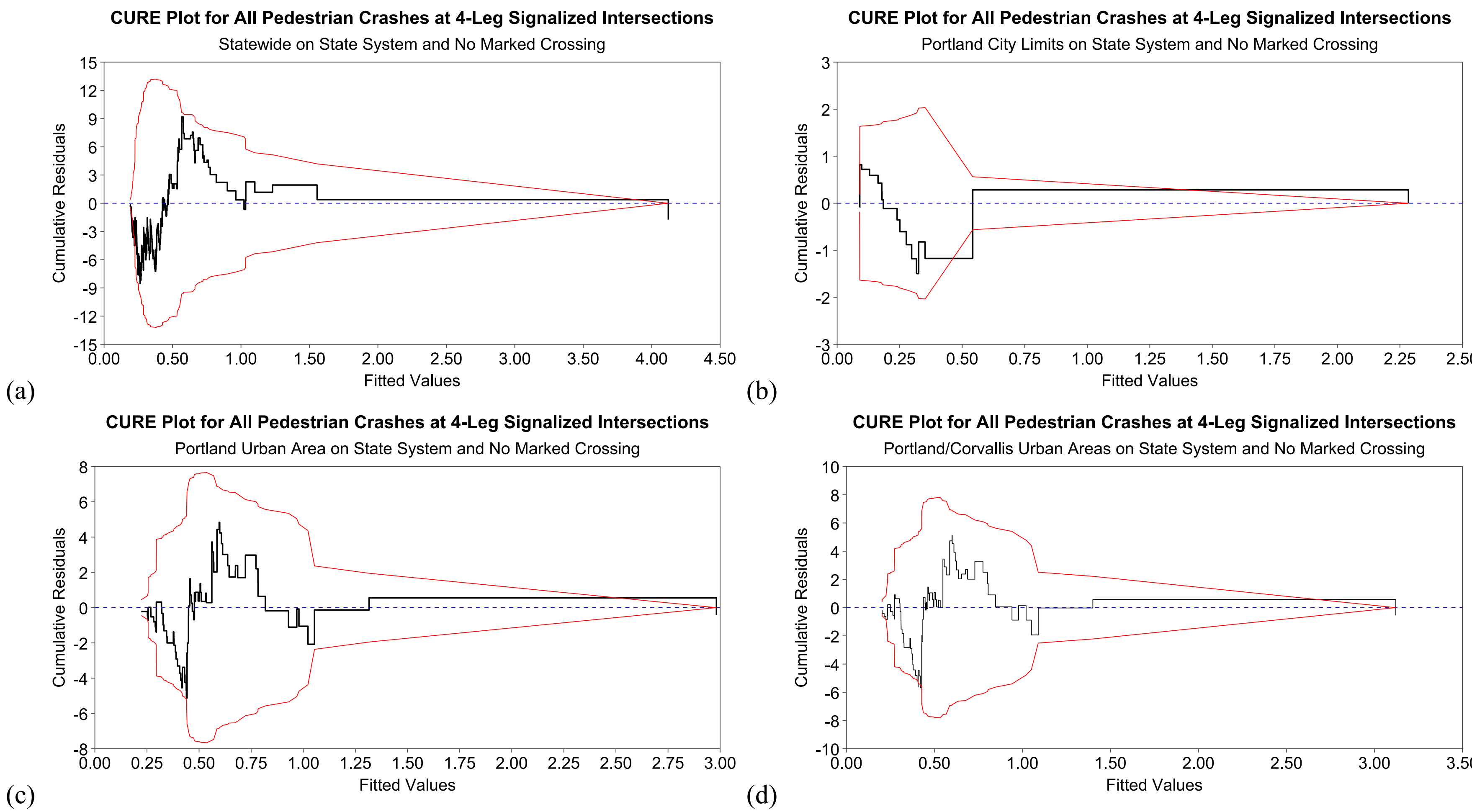


Figure 4.3: CURE Plots for 4-Leg Signalized Intersections and No Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on State System

4.3.2.2 Non-State System

A summary of model results by location for 4-leg signalized intersections with no marked crossing on the non-state system is given in Table 4-4, while full model specifications are in Appendix B (Table B.13 to Table B.16). Table 4-4Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

In these models, the effects of average annual daily traffic and average annual daily pedestrian traffic were significant for each location (statewide, Portland City Limits, Portland Urban Area, and Portland/Corvallis Urban Areas). In all models, vehicle volume and pedestrian volume are expected to increase pedestrian crash frequency.

The CURE plots for each model are shown in Figure 4.4. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash values). Referring to Figure 4.4, each model performs well across the range of crash frequencies.

Table 4.4: Summary of 4-Leg Signalized Intersection with No Marked Crossing Models on Non-State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	*	358
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	358
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	*	56
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	*	56
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	*	129
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	*	129
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	137
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	137

* Significant with at least 90% confidence
NS = not significant

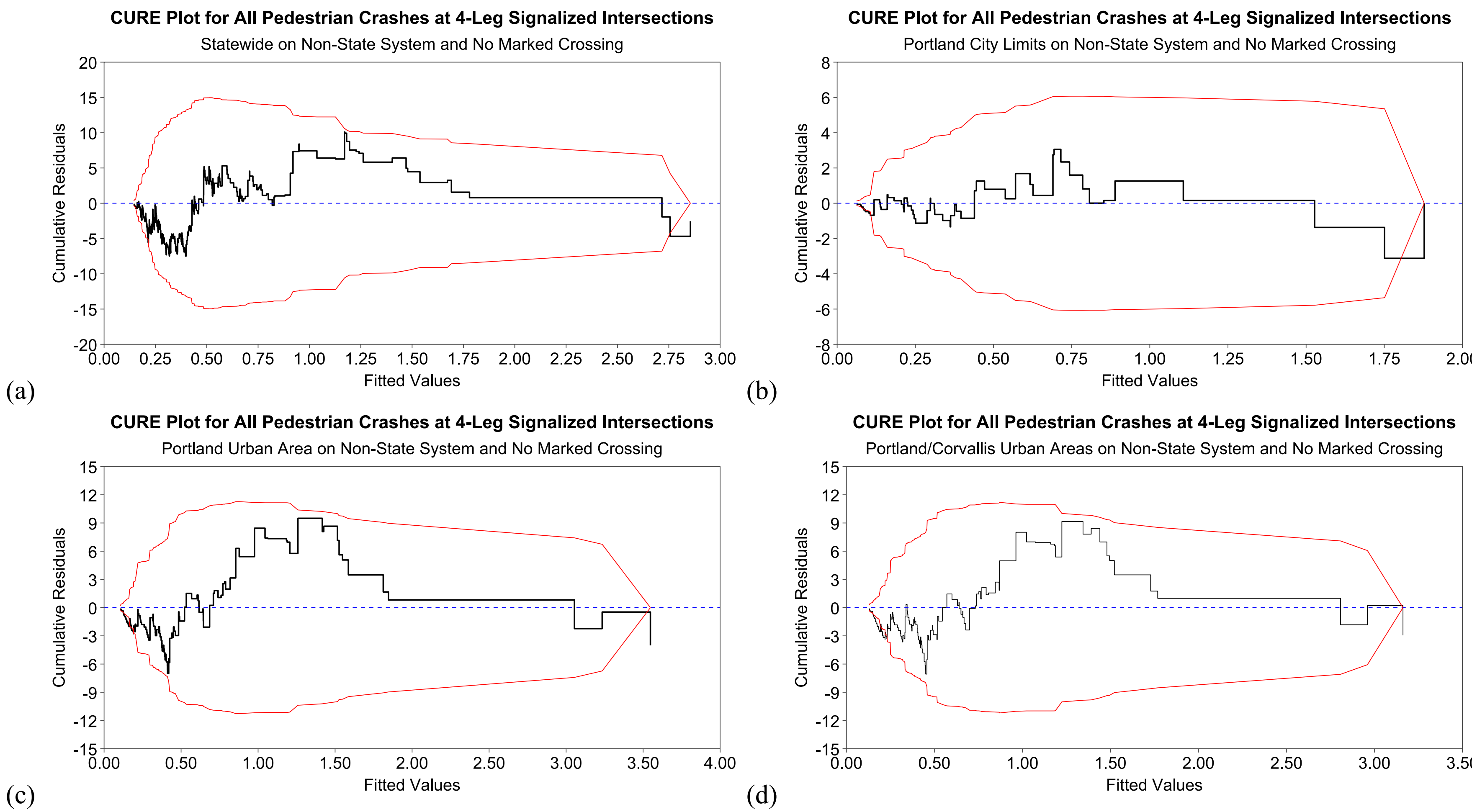


Figure 4.4: CURE Plots for 4-Leg Signalized Intersections and No Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on Non-State System

4.4 4-LEG UNSIGNALIZED INTERSECTIONS

The following subchapters present model results for all 4-leg unsignalized intersection models. The models are presented in the following two groups: (1) a marked crossing is present and (2) no marked crossing is present.

4.4.1 4-Leg Unsignalized Intersections with Marked Crossing

The following subchapters will present model specifications based on location and intersection jurisdiction.

4.4.1.1 State System

A summary of model results by location for 4-leg unsignalized intersections with a marked crossing on the state system is given in Table 4.5, while full model specifications are in Appendix B (Table B.17 to Table B.19). Table 4.5 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

Unlike the 4-leg signalized intersection models, the results for the 4-leg unsignalized intersection models vary. In addition, sample sizes became an issue when considering these disaggregate data models. In the statewide model, while average annual daily traffic and average annual daily pedestrian traffic are expected to increase pedestrian crash frequency, only the effects of average annual daily traffic were significant. In the urban area models (Portland Urban Area and Portland/Corvallis Urban Areas), both vehicle volume and pedestrian volume are expected to increase pedestrian crash frequency. The effect of vehicle volume was significant in the Portland Urban Area model, while the effects of both vehicle volume and pedestrian volume were significant in the Portland/Corvallis Urban Areas model. For both urban area models, the sample size was small (16 intersections and 22 intersections, respectively).

The Portland City Limits model could not be estimated, as there were only two intersections.

The CURE plots for each model are shown in Figure 4.5. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash values). Referring to Figure 4.5, each model performs well for most crash values.

Table 4.5: Summary of 4-Leg Unsignalized Intersection with Marked Crossing Models on State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	*	160
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	NS	160
Average Annual Daily Traffic (Thousands)	Portland City Limits	—	—	2
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	—	—	2
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	*	16
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	NS	16
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	22
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	NS	22

* Significant with at least 90% confidence
NS = not significant

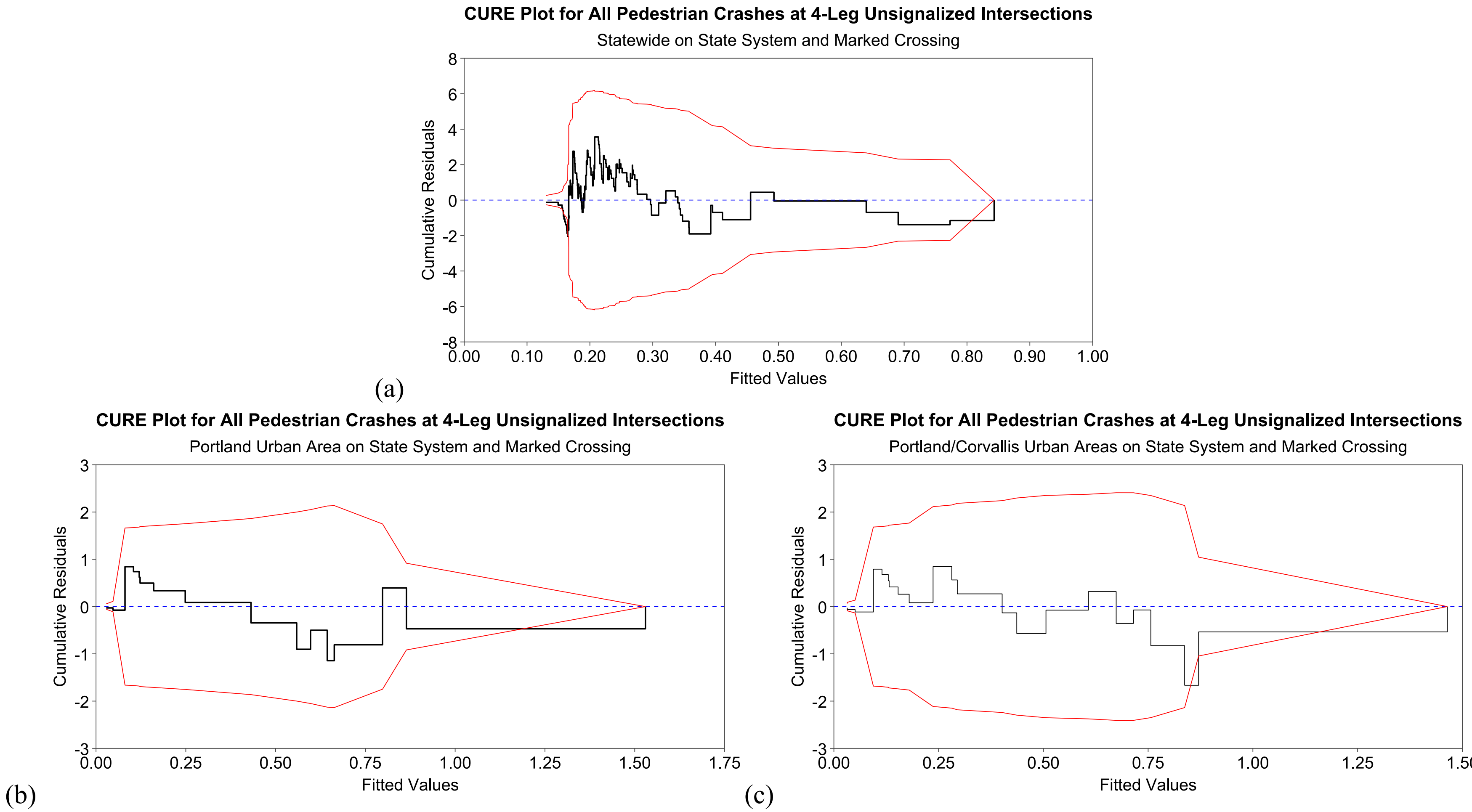


Figure 4.5: CURE Plots for 4-Leg Unsignalized Intersections and Marked Crossing (a) Statewide, (b) Portland Urban Area, and (c) Portland/Corvallis Urban Areas on State System

4.4.1.2 Non-State System

A summary of model results by location for 4-leg unsignalized intersections with a marked crossing on the non-state system is given in Table 4-6, while full model specifications are in Appendix B (Table B.20 to Table B.23). Table 4-6Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

Contrary to the 4-leg unsignalized intersection model with a marked crossing on the state system, the non-state system models had consistent results and did not have sample size issues. In all models, both traffic volume and pedestrian volume had significant effects that increased expected pedestrian crash frequency. In addition, the smallest sample was 340, compared to three samples of less than 25 in the 4-leg unsignalized intersection with a marked crossing on the state system group.

The CURE plots for each model are shown in Figure 4.6. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash values). Referring to Figure 4.6, each model performs well across the range of crash frequency values.

Table 4.6: Summary of 4-Leg Unsignalized Intersection with Marked Crossing Models on Non-State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	*	1,086
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	1,086
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	*	340
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	*	340
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	*	565
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	*	565
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	594
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	594

* Significant with at least 90% confidence
NS = not significant

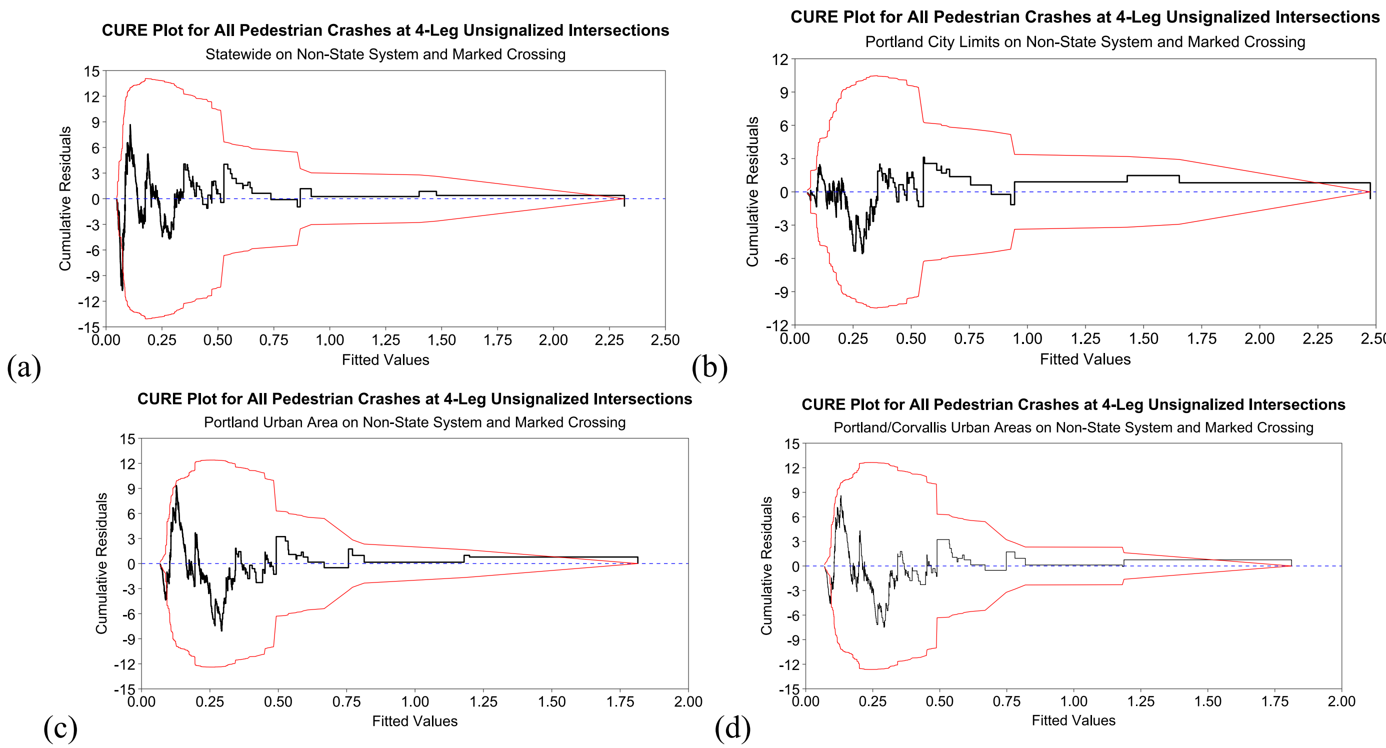


Figure 4.6: CURE Plots for 4-Leg Unsignalized Intersections and Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on Non-State System

4.4.2 4-Leg Unsignalized Intersections with No Marked Crossing

The following subchapters will present model specifications based on location and intersection jurisdiction.

4.4.2.1 State System

A summary of model results by location for 4-leg unsignalized intersections with no marked crossing on the state system is given in Table 4-7, while full model specifications are in Appendix B (Table B.24 to Table B.27). Table 4-7Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

The results for these models show that traffic volume is not a significant contributing factor to pedestrian crash frequency, in addition to having a negative effect (i.e., it decreases the expected number of pedestrian crashes). These results highlight the important consideration of model complexity versus model usability. A common reason for opposite effects stems from omitted variable bias, such that, in this context, additional characteristics may be required to accurately explain pedestrian crash frequency behavior. In addition, as shown in APPENDIX D, these samples have a large number of zero crash frequencies, which can also lead to biased estimates. More advanced modeling approaches may be required to uncover underlying behavior and interactions, but potentially at the cost of usability by transportation agencies.

Pedestrian volume was a significant contributing factor in each model and is expected to increase pedestrian crash frequency.

The CURE plots for each model are shown in Figure 4.7. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash values). Referring to Figure 4.7, the statewide model struggles with fitted lower crash frequencies (i.e., zeros), as does the Portland City Limits model and the Portland/Corvallis Urban Areas model.

Table 4.7: Summary of 4-Leg Unsignalized Intersection with No Marked Crossing Models on State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	NS	1,208
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	1,208
Average Annual Daily Traffic (Thousands)	Portland City Limits	−	NS	127
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	*	127
Average Annual Daily Traffic (Thousands)	Portland Urban Area	−	NS	201
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	*	201
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	−	NS	226
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	226

* Significant with at least 90% confidence
NS = not significant

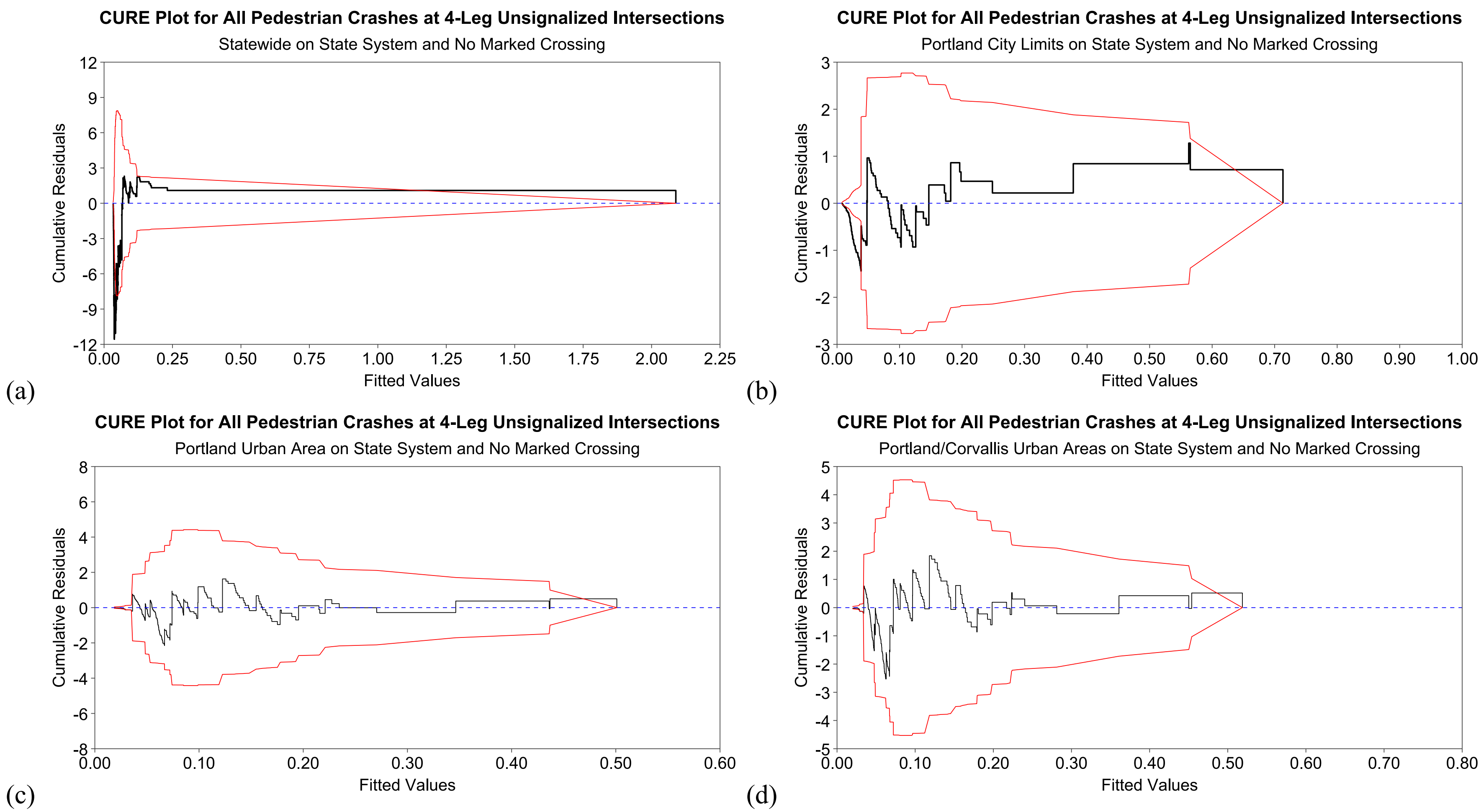


Figure 4.7: CURE Plots for 4-Leg Unsignalized Intersections and No Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on State System

4.4.2.2 Non-State System

A summary of model results by location for 4-leg unsignalized intersections with no marked crossing on the non-state system is given in Table 4-8, while full model specifications are in Appendix B (Table B.28 to Table B.31). Table 4-8Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

Unlike the results for the models on the state system, the results from these models were consistent and had results that were expected. In each model, the effects of vehicle volume and pedestrian volume were significant, and both are expected to increase pedestrian crash frequency.

The CURE plots for each model are shown in Figure 4.8. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash values). Most models perform well over the range of crash frequencies, with the exception being the statewide model. The statewide model struggles with small values and has some struggles with midrange crash frequency values.

Table 4.8: Summary of 4-Leg Unsignalized Intersection with No Marked Crossing Models on Non-State System

Variable	Geographic Area	Effect	Sig.	n
Average Annual Daily Traffic (Thousands)	Statewide	+	*	7,897
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	7,897
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	*	1,634
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	*	1,634
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	*	3,065
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	*	3,065
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	3,219
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	3,219

* Significant with at least 90% confidence
NS = not significant

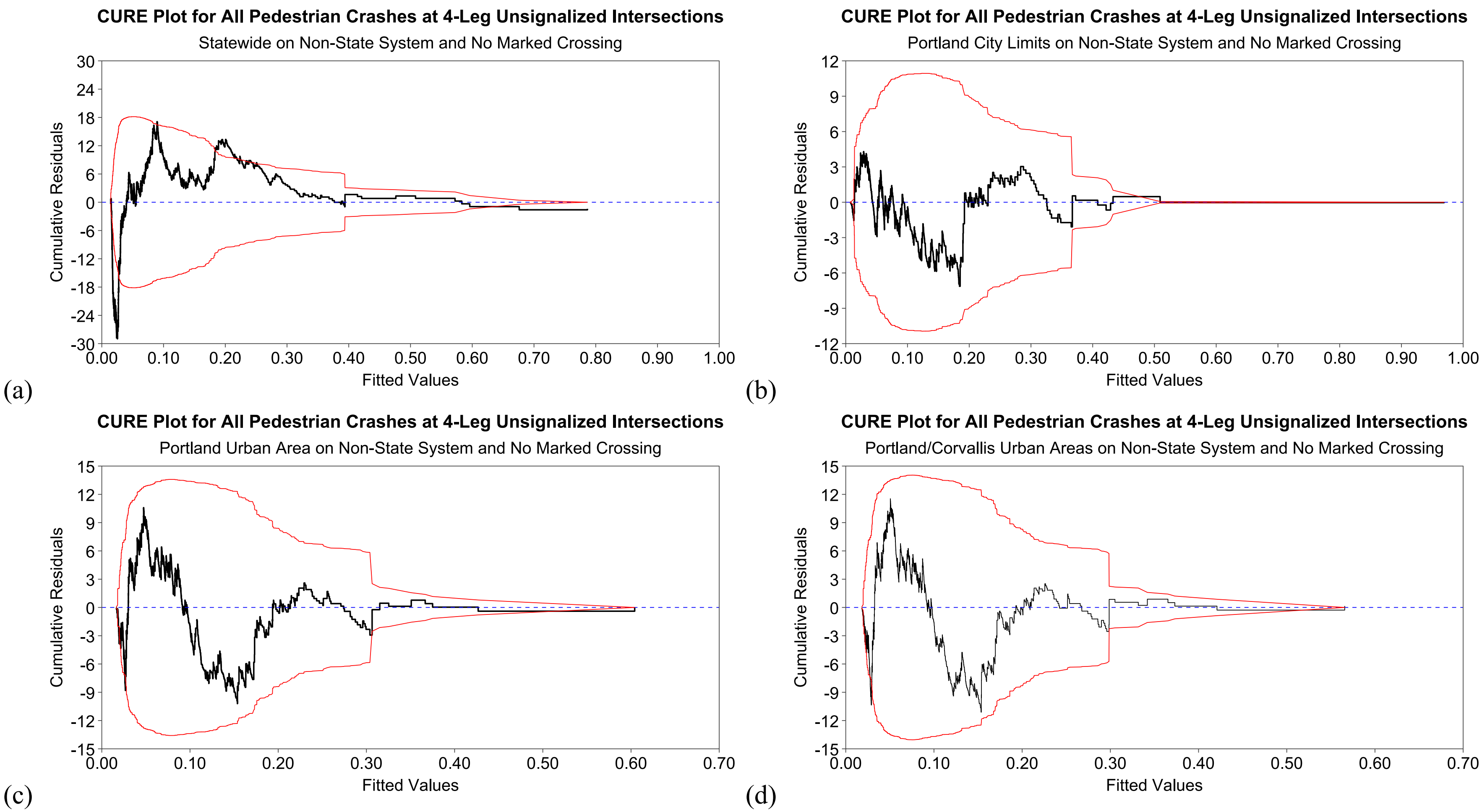


Figure 4.8: CURE Plots for 4-Leg Unsignalized Intersections and No Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on Non-State System

4.5 3-LEG SIGNALIZED INTERSECTIONS

The following subchapters present model results for all 3-leg signalized intersection models. The models are presented in the following two groups: (1) a marked crossing is present and (2) no marked crossing is present.

4.5.1 3-Leg Signalized Intersections with Marked Crossing

The following subchapters will present model summaries based on location and intersection jurisdiction

4.5.1.1 State System

A summary of model results by location for 3-leg signalized intersections with a marked crossing on the state system is given in Table 4-9, while full model specifications are in Appendix C (Table C.1 to Table C.4). Table 4-9Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

Vehicle volume was not a significant contributing factor in each of the models, while pedestrian volume was a significant contributing factor in all models but one (Portland City Limits model, a sample size of 36). The region-specific models have limited sample sizes, ranging from 36 to 112, potentially impacting parameter estimates. Additionally, the coefficient for vehicle volume, although not significant, suggest possible bias (negative coefficient instead of positive coefficient).

The CURE plots for each model are shown in Figure 4.9. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash values).

The statewide model performs well over the range of crash frequencies. The region-specific models struggle with low crash frequency values, which is highlighted most for the Portland Urban Area model and the Portland/Corvallis Urban Areas model. The Portland City Limits model, while the cumulative residuals extend slightly beyond the fitted bounds for smaller crash frequency values, performs relatively well across the model range.

Table 4.9: Summary of 3-Leg Signalized Intersection with Marked Crossing Models on State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	NS	169
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	169
Average Annual Daily Traffic (Thousands)	Portland City Limits	−	NS	36
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	NS	36
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	NS	111
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	*	111
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	NS	112
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	112

* Significant with at least 90% confidence
NS = not significant

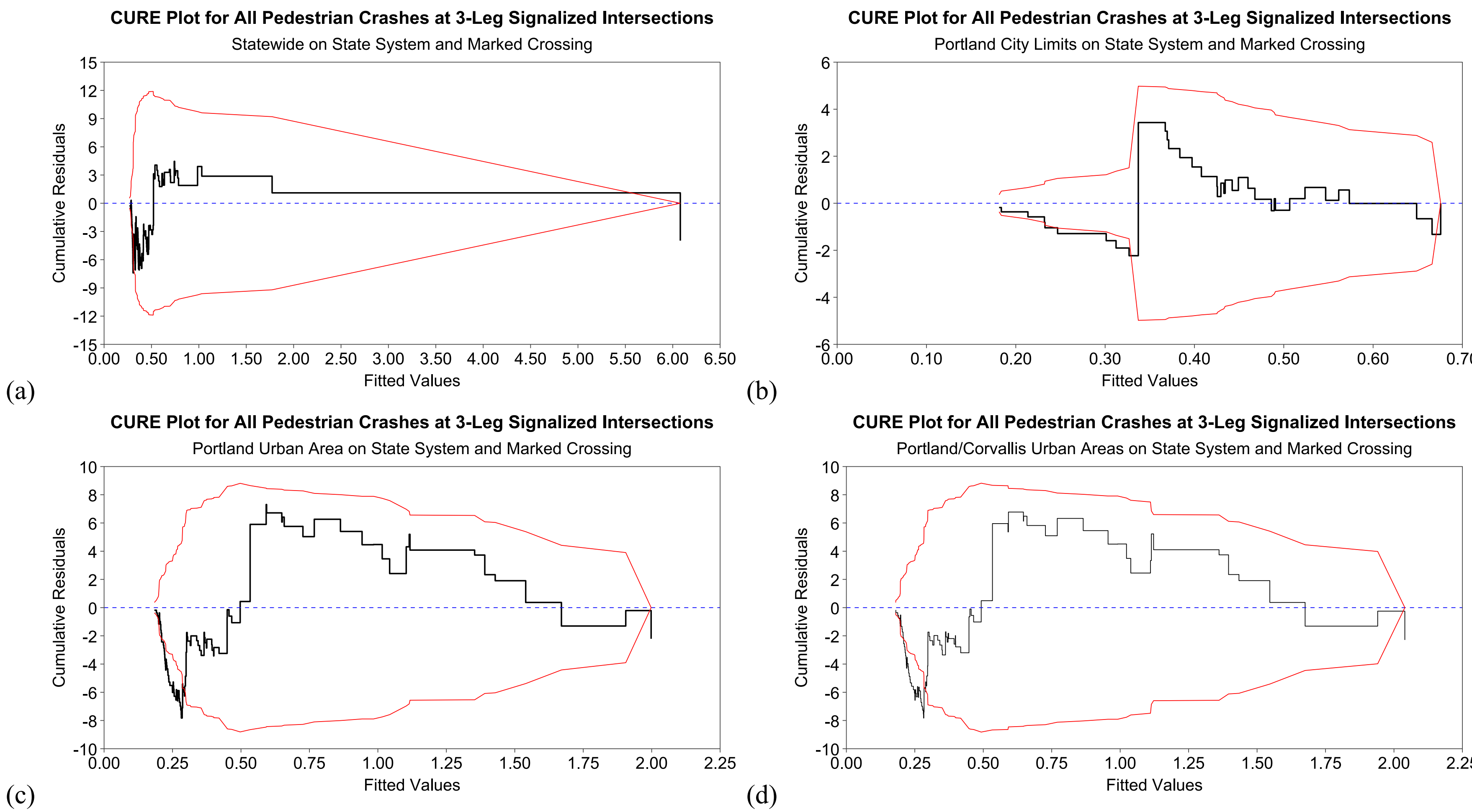


Figure 4.9: CURE Plots for 3-Leg Signalized Intersections and Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on State System

4.5.1.2 Non-State System

A summary of model results by location for 3-leg signalized intersections with a marked crossing on the non-state system is given in Table 4-10, while full model specifications are in Appendix C (Table C.5 to Table C.8). Table 4-10Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

Vehicle volume was a significant contributing factor in each of the models, and in each model resulted in an expected increase in pedestrian crash frequency. Pedestrian volume was not a significant contributing factor in the statewide model, the Portland City Limits model, and the Portland Urban Area model, but did have significant effects in the Portland/Corvallis Urban Areas model. Pedestrian volume, even if effects were not significant, is expected to increase pedestrian crash frequency. The sample sizes for the models ranged from 160 (Portland City Limits model) to 411 (statewide model).

The CURE plots for each model are shown in Figure 4.10. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash frequencies).

The statewide model performs well over the range of crash frequencies. The cumulative residuals line extends slightly beyond the fitted bounds for the region-specific models at lower fitted values, suggesting there may be marginal issues with small crash frequencies. Nonetheless, all models, based on the CURE plots, perform well across the range of crash frequencies.

Table 4.10: Summary of 3-Leg Signalized Intersection with Marked Crossing Models on Non-State System

Variable	Geographic Area	Effect	Sig.	***n***
Average Annual Daily Traffic (Thousands)	Statewide	+	*	411
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	NS	411
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	*	160
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	NS	160
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	*	297
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	NS	297
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	304
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	304

* Significant with at least 90% confidence
NS = not significant

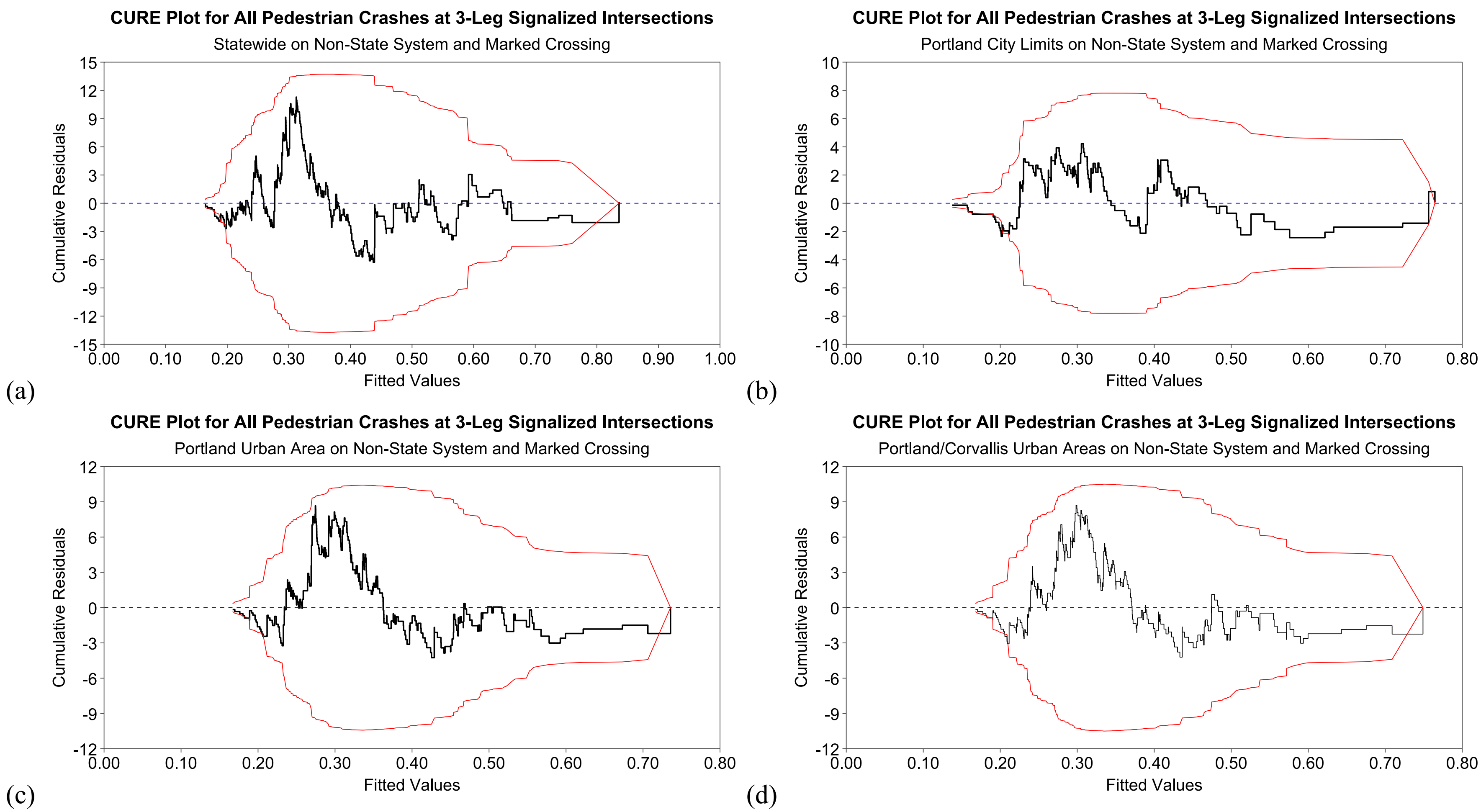


Figure 4.10: CURE Plots for 3-Leg Signalized Intersections and Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on Non-State System

4.5.2 3-Leg Signalized Intersections with No Marked Crossing

The following subchapters will present model specifications based on location and intersection jurisdiction

4.5.2.1 State System

A summary of model results by location for 3-leg signalized intersections with no marked crossing on the state system is given in Table 4-11, while full model specifications are in Appendix C (Table C.9 to Table C.12). Table 4-11Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

This group of models suffered from relatively small sample sizes compared to other disaggregate model groups. While the sample size for the statewide model was 147, sample sizes for the region-specific models ranged from 21 to 48. No effects were determined to be significant for vehicle volume or pedestrian volume, and coefficient signs for vehicle volume variables suggest possible bias (negative effects when positive effects are expected). Inference from these models is limited and more complex modeling approaches and specifications are likely required.

The CURE plots for each model are shown in Figure 4.11. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash frequencies).

Based on Figure 4.11, the statewide model performs the best over the range of crash frequencies. The Portland City Limits model has marginal struggles with smaller crash frequencies, while the Portland Urban Area and Portland/Corvallis Urban Area models struggle moderately with smaller crash frequencies.

Table 4.11: Summary of 3-Leg Signalized Intersection with No Marked Crossing Models on State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	−	NS	147
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	NS	147
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	NS	21
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	NS	21
Average Annual Daily Traffic (Thousands)	Portland Urban Area	−	NS	47
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	NS	47
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	−	NS	48
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	NS	48

NS = not significant

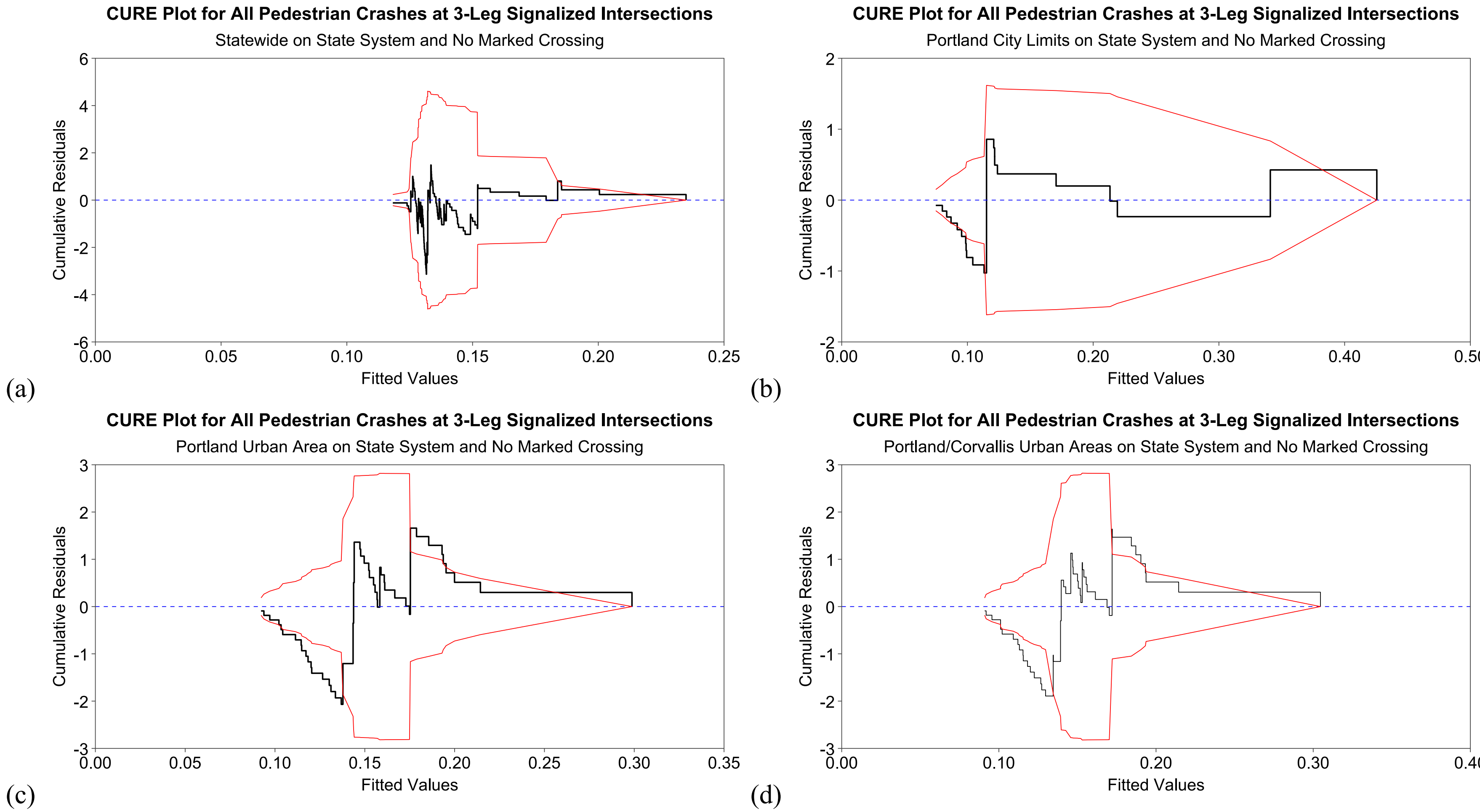


Figure 4.11: CURE Plots for 3-Leg Signalized Intersections and No Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on State System

4.5.2.1 Non-State System

A summary of model results by location for 3-leg signalized intersections with no marked crossing on the non- state system is given in Table 4-12, while full model specifications are in Appendix C (Table C.13 to Table C.16). Table 4-11Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

As was the case with the 3-leg signalized intersections with no marked crossing on the state system group of models, these models suffered from relatively small sample sizes compared to other disaggregate model groups. While the sample size for the statewide model was 197, sample sizes for the region-specific models ranged from 39 to 98. Vehicle volume effects were significant in the statewide model only (increase in expected pedestrian crash frequency) and no pedestrian volume effects were determined to be significant. Despite this, all coefficient signs were expected (positive).

The CURE plots for each model are shown in Figure 4.12. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash frequencies).

Based on Figure 4.11, the statewide model performed well over the range of crash frequencies but there is room for improvement for smaller crash frequencies. The Portland City Limits model also performed well over the range of crash frequencies but there is room for improvement at higher crash frequencies. The urban area models performed well despite some slight deviations beyond the fitted bounds at lower crash frequency values. Overall, each model performed well.

Table 4.12: Summary of 3-Leg Signalized Intersection with No Marked Crossing Models on Non-State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	*	197
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	NS	797
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	NS	39
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	NS	39
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	NS	95
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	NS	95
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	NS	98
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	NS	98

* Significant with at least 90% confidence
NS = not significant

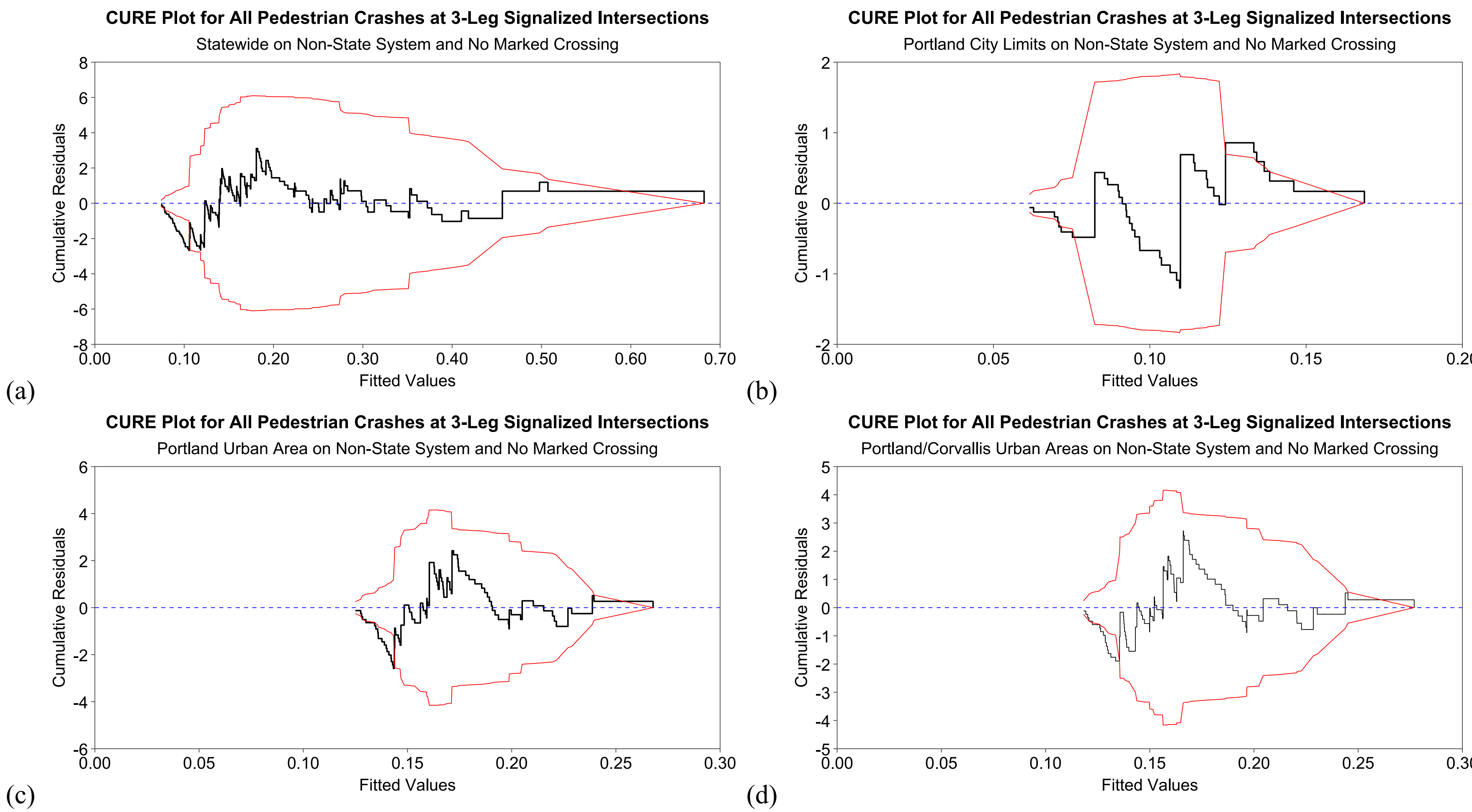


Figure 4.12: CURE Plots for 3-Leg Signalized Intersections and No Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on Non-State System

4.6 3-LEG UNSIGNALIZED INTERSECTIONS

The following subchapters present model results for all 3-leg signalized intersection models. The models are presented in the following two groups: (1) a marked crossing is present and (2) no marked crossing is present.

4.6.1 3-Leg Unsignalized Intersections with Marked Crossing

The following subchapters will present model specifications based on location and intersection jurisdiction.

4.6.1.1 State System

A summary of model results by location for 3-leg unsignalized intersections with a marked crossing on the state system is given in Table 4-13, while full model specifications are in Appendix C (Table C.17 to Table C.20). Table 4-13Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

Similar to other 3-leg intersection groups, sample size was a concern. The largest sample was for the statewide model (151), while sample sizes for the region-specific models ranged from 28 to 43. Neither vehicle volume nor pedestrian volume had significant effects in any model, and the coefficient for pedestrian volume in the Portland City Limits model was negative.

The CURE plots for each model are shown in Figure 4.13. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash frequencies).

Despite no significant effects of vehicle volume and pedestrian volume, the models performed well. Of the models, the urban area models performed best, with the cumulative residuals extending beyond the fitted bounds at high crash frequencies. The statewide model and Portland City Limits model had issues at low, midrange, and high crash frequencies; albeit, the cumulative residuals only marginally extended beyond the fitted bounds.

Table 4.13: Summary of 3-Leg Unsignalized Intersection with Marked Crossing Models on State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	NS	151
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	NS	151
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	NS	28
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	−	NS	28
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	NS	42
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	NS	42
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	NS	43
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	NS	43

NS = not significant

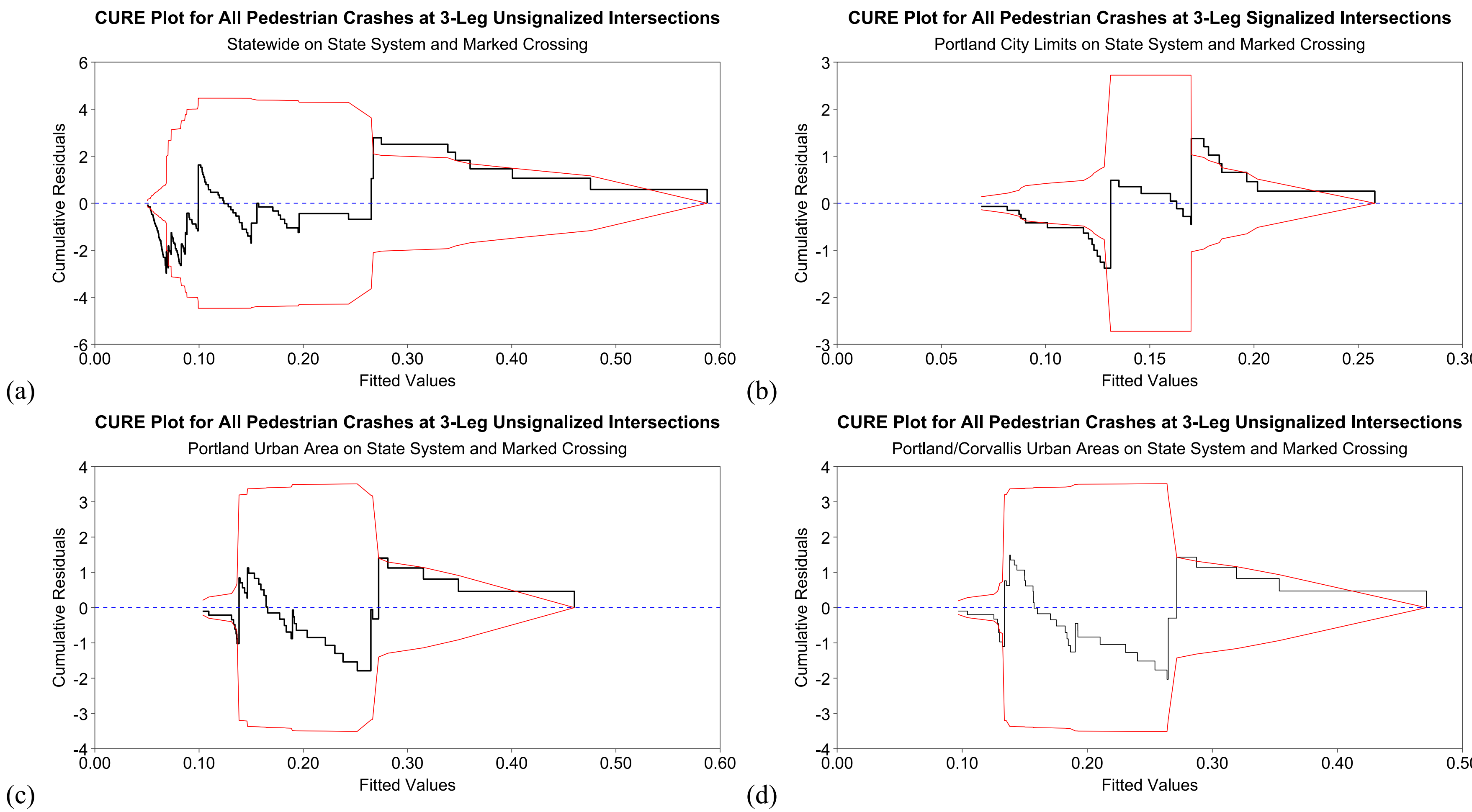


Figure 4.13: CURE Plots for 3-Leg Unsignalized Intersections and Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on State System

4.6.1.2 Non-State System

A summary of model results by location for 3-leg unsignalized intersections with a marked crossing on the non-state system is given in Table 4-14, while full model specifications are in Appendix C (Table C.21 to Table C.24). Table 4-14Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

Unlike previous 3-leg intersection models, sample size was not much of a concern in this model group, as sample sizes ranged from 254 to 1,244. Vehicle volume had significant effects in all models except the Portland City Limits model, and in each model, it is expected to increase pedestrian crash frequency. Pedestrian volume had significant effects in all models and is expected to increase pedestrian crash frequency.

The CURE plots for each model are shown in Figure 4.14. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash frequencies).

Figure 4.14 shows that each of the models performed well across the range of crash frequencies. The statewide model has room for some improvement when considering smaller crash frequencies, while the other three models have cumulative residuals that oscillate about zero and stay within the fitted bounds over all crash frequencies.

Table 4.14: Summary of 3-Leg Unsignalized Intersection with Marked Crossing Models on Non-State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	*	1,244
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	1,244
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	NS	254
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	*	254
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	*	619
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	*	619
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	640
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	640

* Significant with at least 90% confidence
NS = not significant

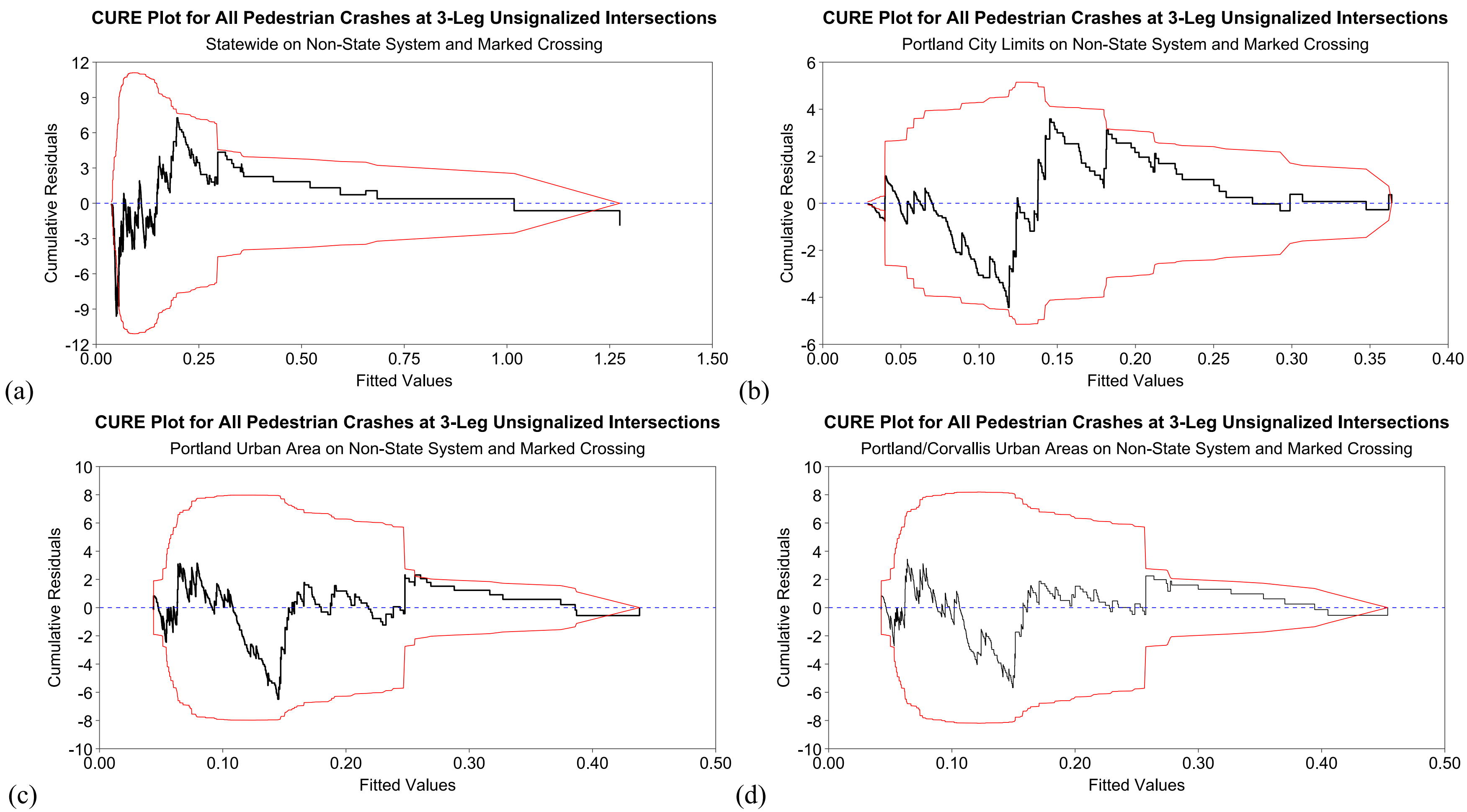


Figure 4.14: CURE Plots for 3-Leg Unsignalized Intersections and Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on Non-State System

4.6.2 3-Leg Unsignalized Intersections with No Marked Crossing

The following subchapters will present model specifications based on location and intersection jurisdiction.

4.6.2.1 State System

A summary of model results by location for 3-leg unsignalized intersections with no marked crossing on the state system is given in Table 4-15, while full model specifications are in Appendix C (Table C.25 to Table C.28). Table 4-15Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

Despite sample size not being a concern within this group of disaggregate models, results varied. Both vehicle volume and pedestrian volume had significant effects in the statewide model (increase in expected pedestrian crash frequency), while both vehicle volume and pedestrian volume did not have significant effects in the Portland City Limits model (the coefficient was also negative for vehicle volume). In the two urban area models, vehicle volume effects were not significant and pedestrian volume effects were significant.

The CURE plots for each model are shown in Figure 4.15. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash frequencies).

Figure 4.15 shows that each of the models have room for improvement, the first case such thus far. The statewide model does not perform well at lower crash frequencies, which may be linked to a high number of zeros. The statewide model was the only model in which the effects from both variables were significant. The Portland City Limits model can also be improved across all ranges of crash frequencies (the effects were not significant in this model). The urban area models have room for improvement at all levels of crash frequencies, with notable deviations from the fitted bounds at low and midrange crash frequency values.

Table 4.15: Summary of 3-Leg Unsignalized Intersection with No Marked Crossing Models on State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	*	3,993
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	3,993
Average Annual Daily Traffic (Thousands)	Portland City Limits	−	NS	561
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	NS	561
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	NS	1,144
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	*	1,144
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	NS	1,224
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	1,224

* Significant with at least 90% confidence
NS = not significant

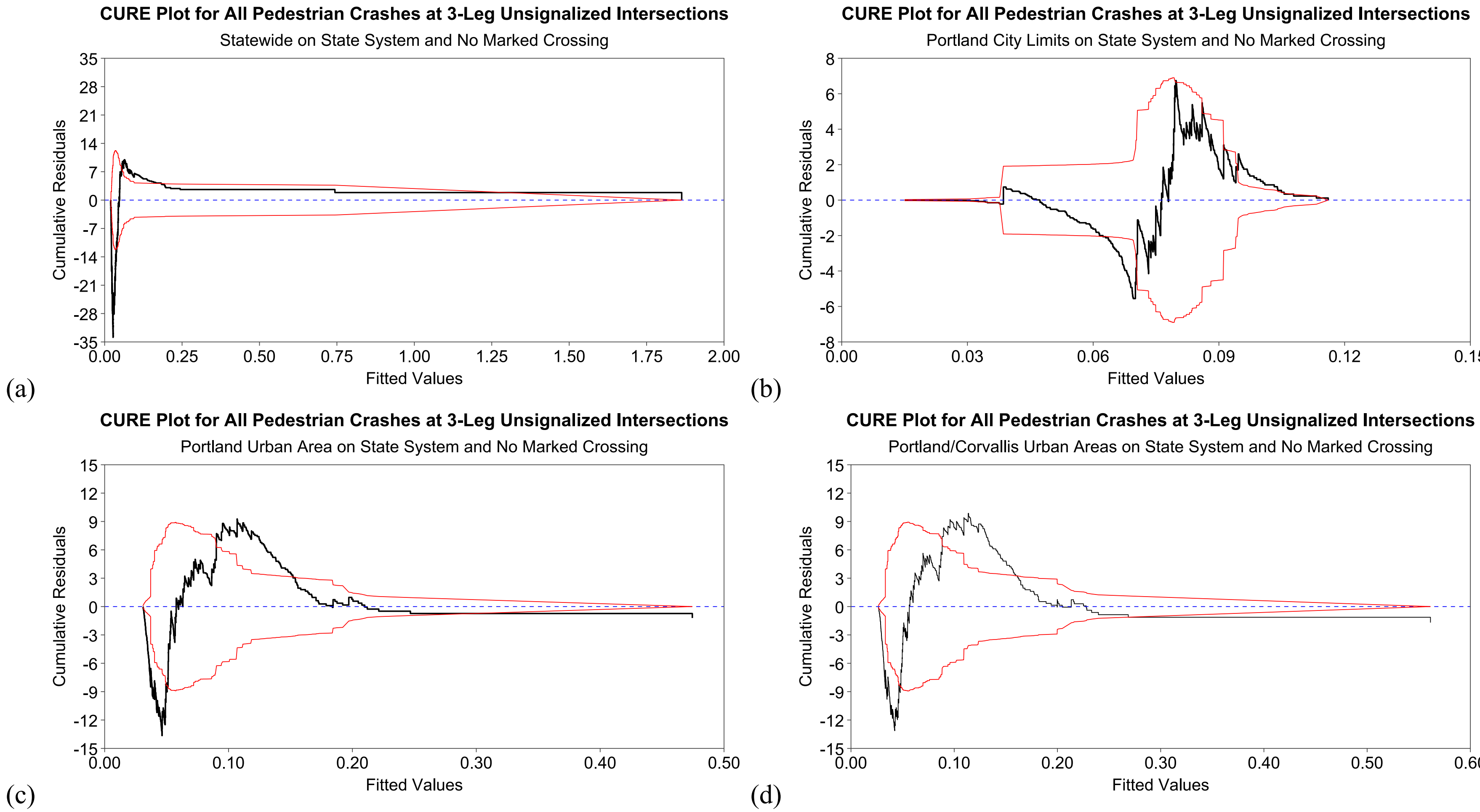


Figure 4.15: CURE Plots for 3-Leg Unsignalized Intersections and No Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on State System

4.6.2.2 Non-State System

A summary of model results by location for 3-leg unsignalized intersections with no marked crossing on the non-state system is given in Table 4-16, while full model specifications are in Appendix C (Table C.29 to Table C.32). Table 4-16Table 4-1 shows the effect of the variable (increase or decrease in pedestrian crash frequency), if the effects are statistically significant, and how many observations were used to estimate the model.

This group of models had the greatest sample sizes. The effects of vehicle volume were significant in each model, and the effects of pedestrian volume were significant in each model. Per log-likelihood values, these models also consistently had the best model fit in terms of improvement over the null model (model with only a constant estimated).

The CURE plots for each model are shown in Figure 4.16. These plots can be used to visually assess how well the crash frequency model fits. If the line representing the cumulative residuals (the black line) stays within the fitted bounds (the red lines) and oscillates about zero, the crash frequency model has good fit over the range of the model (i.e., all crash frequencies).

Despite some of the best fit models per log-likelihood values, Figure 4.16 shows the struggles the models have with fitting crash frequencies over different ranges. The statewide model can be improved across all crash frequency values, as the cumulative residuals extend beyond the fitted bounds at low, midrange, and high values. This pattern is also observed in the urban area models. The best performing model according to Figure 4.16 is the Portland City Limits model, which does well with the exception of some midrange crash frequency values.

Table 4.16: Summary of 3-Leg Unsignalized Intersection with No Marked Crossing Models on Non-State System

Variable	Geographic Area	Effect	Sig.	*n*
Average Annual Daily Traffic (Thousands)	Statewide	+	*	21,134
Average Annual Daily Pedestrian Traffic (Thousands)	Statewide	+	*	21,134
Average Annual Daily Traffic (Thousands)	Portland City Limits	+	*	2,918
Average Annual Daily Pedestrian Traffic (Thousands)	Portland City Limits	+	*	2,918
Average Annual Daily Traffic (Thousands)	Portland Urban Area	+	*	8,341
Average Annual Daily Pedestrian Traffic (Thousands)	Portland Urban Area	+	*	8,341
Average Annual Daily Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	8,668
Average Annual Daily Pedestrian Traffic (Thousands)	Portland/Corvallis Urban Areas	+	*	8,668

* Significant with at least 90% confidence

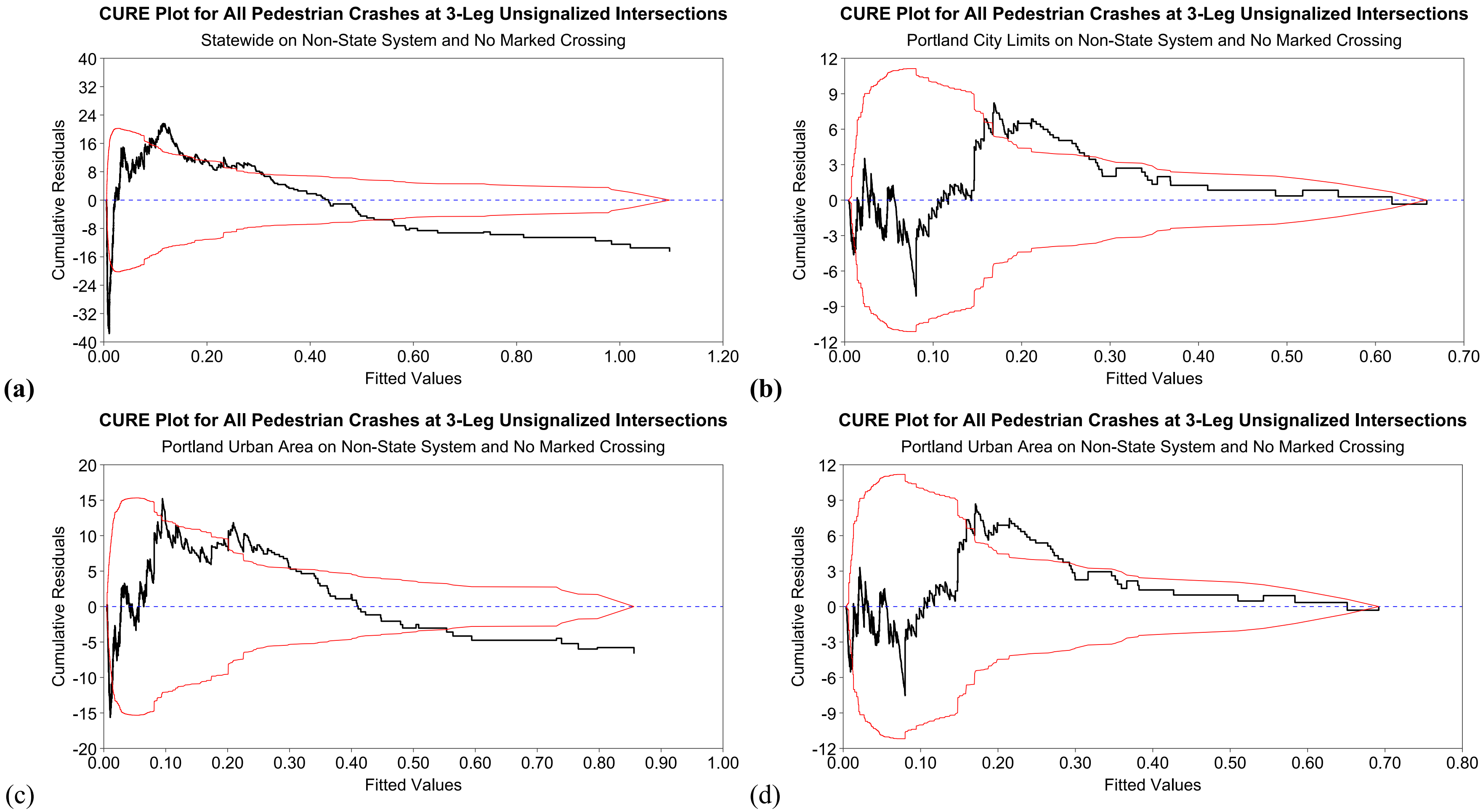


Figure 4.16: CURE Plots for 3-Leg Unsignalized Intersections and No Marked Crossing (a) Statewide, (b) Portland City Limits, (c) Portland Urban Area, and (d) Portland/Corvallis Urban Areas on Non-State System

4.7 COMPARISON BETWEEN 4-LEG INTERSECTIONS

Figure 4.17 and Figure 4.18 show the comparisons of model coefficients for average annual daily traffic and average annual daily pedestrian traffic, respectively, across all 4-leg intersection models.

In most models, the coefficient for average annual daily traffic was positive and the 95% confidence interval did not include zero. This is generally associated with vehicle volume being a significant contributing factor with at least 95% confidence. This also indicates that among the majority of models, vehicle volume is expected to increase pedestrian crash frequency. These results are generally expected.

On the other hand, there were three models in which the coefficients for average annual daily traffic were negative, contrary to expectations and previous research. Despite there being no significance, the expectation is that the coefficients will still be positive, indicating an increase in the expected number of pedestrian crashes. When this is observed, from a modeling framework perspective, it can stem from two potential sources: (1) omitted variable bias and (2) the model is unable to account for the large number of zeros, should a large number of zeros be present. Regardless of the reason, this scenario can impact the fitted values of the model when generating crash prediction plots. It is worth nothing, however, that the three intersection typologies with these estimates were unsignalized intersections with no marked crossing. These results, despite potential methodological limitations, may suggest that pedestrians may be dissuaded from crossing these intersection types as vehicle volume increases. These results warrant further investigation into these intersection typologies.

Average annual daily pedestrian traffic had a positive coefficient in all models and for most models, the 95% confidence interval did not include zero. In all but four models, the effect of pedestrian volume was significant. These results are generally expected.

Being that the models contain only vehicle and pedestrian volumes, omitted variable bias may be present. If there are variables not estimated with the model that impact the dependent variable (pedestrian crash frequency) and have some correlation with other explanatory variables (vehicle or pedestrian volume), the effects of these variables can be picked up by the variables included in the model. For all models, if this is present, the variables picking up these effects would be vehicle volume or pedestrian volume. With this in mind, it is important to reiterate the reason for estimating models with only vehicle and pedestrian volume. As additional variables are included in model specifications, model complexity increases. In the context of this work, the data process to obtain additional variables to be used in the model may not always be achievable by transportation agencies. This sparks the consideration of usability versus complexity, where usability was chosen. To address this, additional variables can be included in model specifications, such as variables related to the built-environment, land use, socioeconomics, demographics, roadway design, and others.

Addressing a preponderance of zeros will add even more complexity, as more advanced and sophisticated models are required. Often times, these more advanced and sophisticated models also require specific software that may not be readily available. In addition, some of the most common models that address this zero inflation should be carefully considered in a transportation safety context. The manner in which many of these models define zeros does not fully align with, in the case of this work, zero crash intersections. The concept is that no intersection is 100% safe, and because transportation safety includes human participants and human error, there is always a chance a crash can occur. While statistical fit may improve, defending its practical application may prove difficult. In the end, the decision to choose a model that accounts for a large number of zeros should be supported by a suite of models, tests for zero effects, and practical reasoning for assuming zeros belong in their own count state and are true zeros.

To estimate the expected number of crashes, each model was used to generate the crash prediction plots shown in Figure 4.19 to Figure 4.22. All crash prediction plots show the importance of vehicle volume in estimating pedestrian crash frequency (as vehicle volume increases, expected pedestrian crash frequency increases). In most cases, the prediction trends remain consistent until large values of vehicle volume, which then results in a substantial increase in the expected number of crashes. The plots are also reliant on the ranges of values used to estimate the model; that is, some plots show steep increases in expected pedestrian crash frequency once vehicle volume exceeds 75,000, which was the maximum observed vehicle volume in some models.

Also of note is the behavior of estimated crash frequency if the sample size is small (small number of intersections) (see Figure 4.21). To estimate reliable models, variation is needed, and variation comes with more observations. The final note is on the importance of vehicle volume and how potentially biased estimates can impact crash estimations (see Figure 4.22). For all state system models in Figure 4.22, note that pedestrian volume is a significant contributing factor. However, due to the large influence vehicle volume has on crash estimations, the insignificance of its effects and its coefficient being less than zero results in plots that show pedestrian crash frequency decreasing as vehicle volume increases. This further illustrates the importance of the usability versus complexity tradeoff and ties in with the previous discussion.

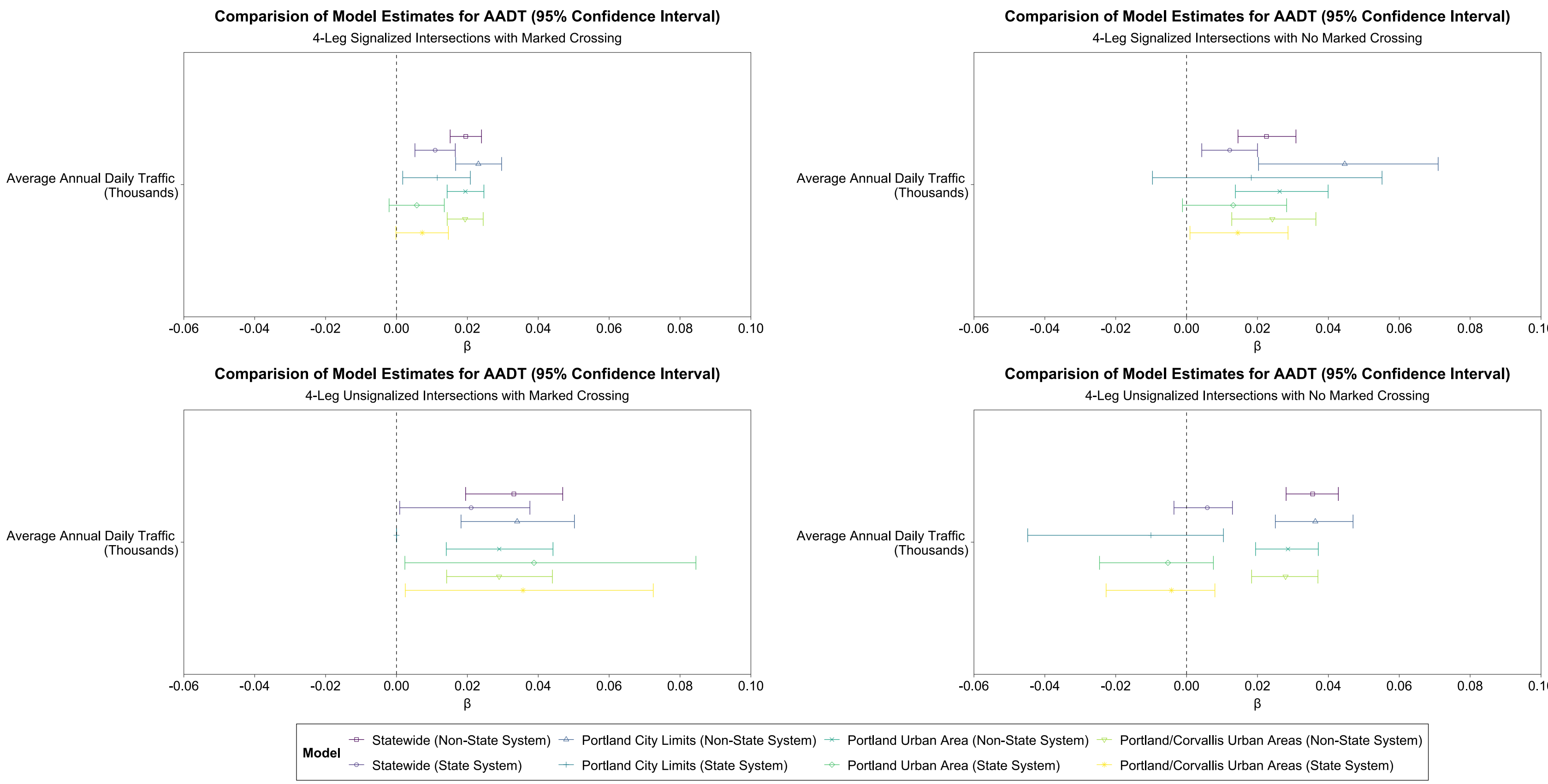


Figure 4.17: Comparison of Model Estimates for AADT Between 4-Leg Intersection Models

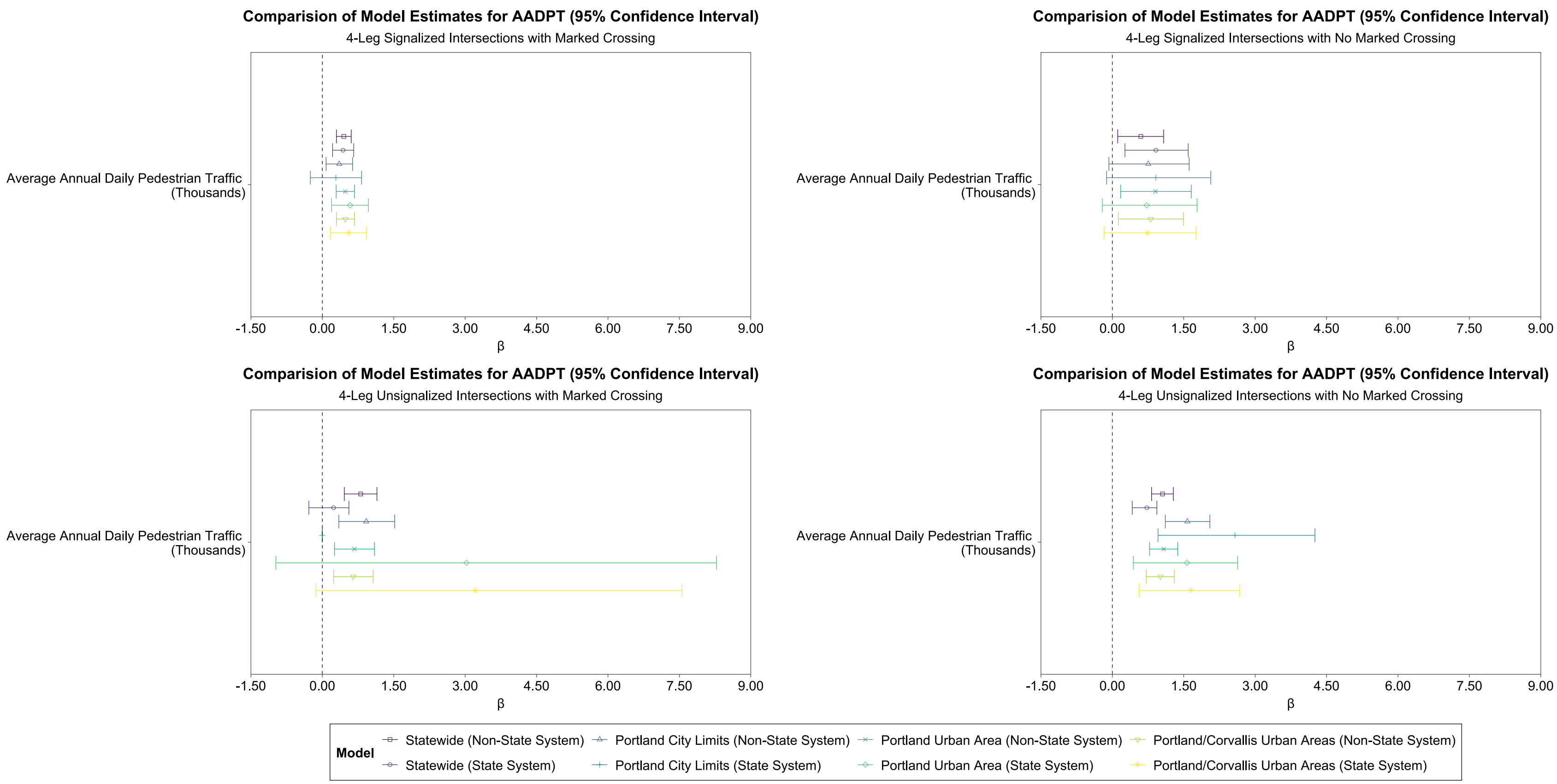


Figure 4.18: Comparison of Model Estimates for AADPT Between 4-Leg Intersection Models

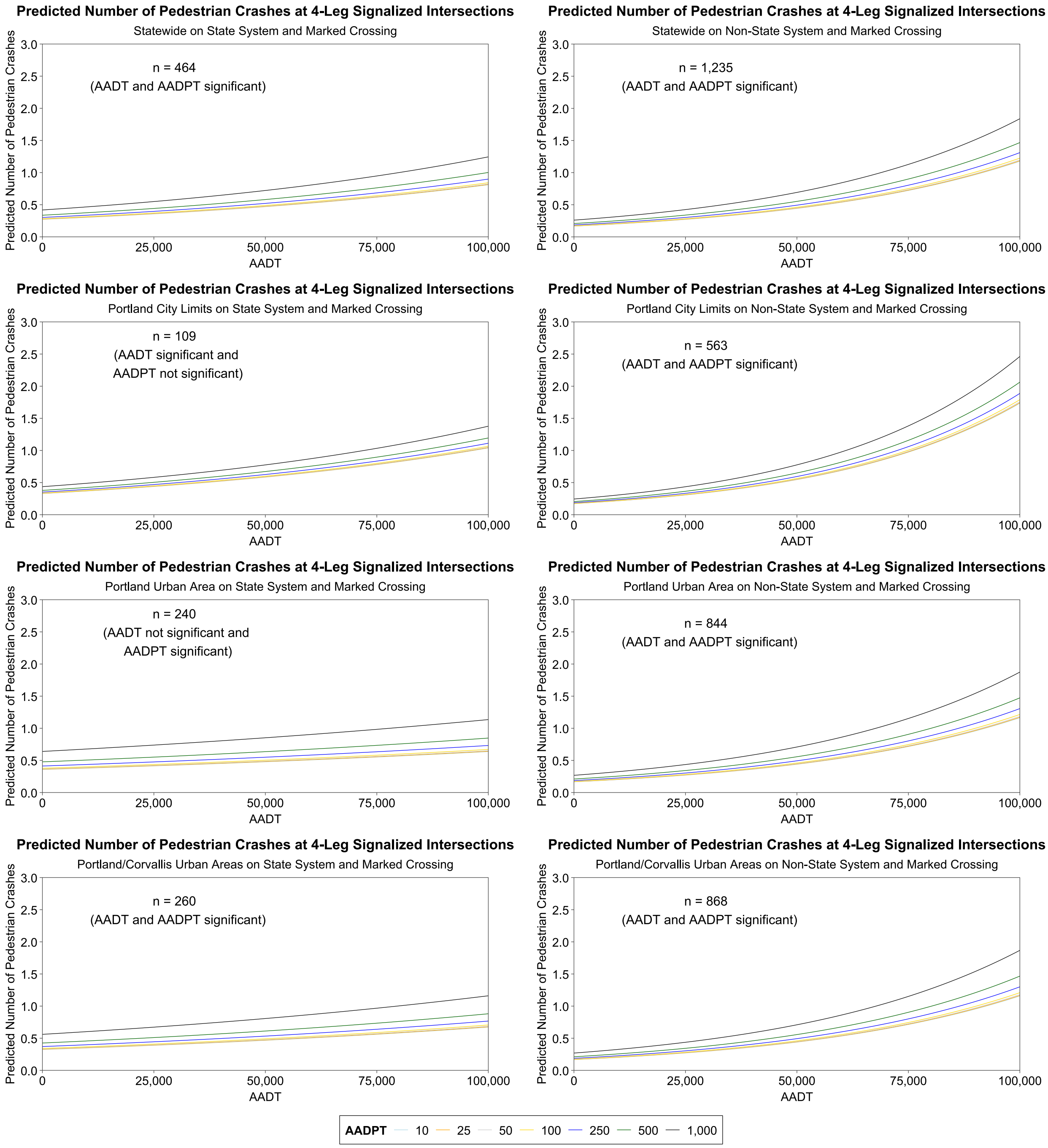


Figure 4.19: Predicted Number of Pedestrian Crashes at 4-Leg Signalized Intersections with Marked Crossing

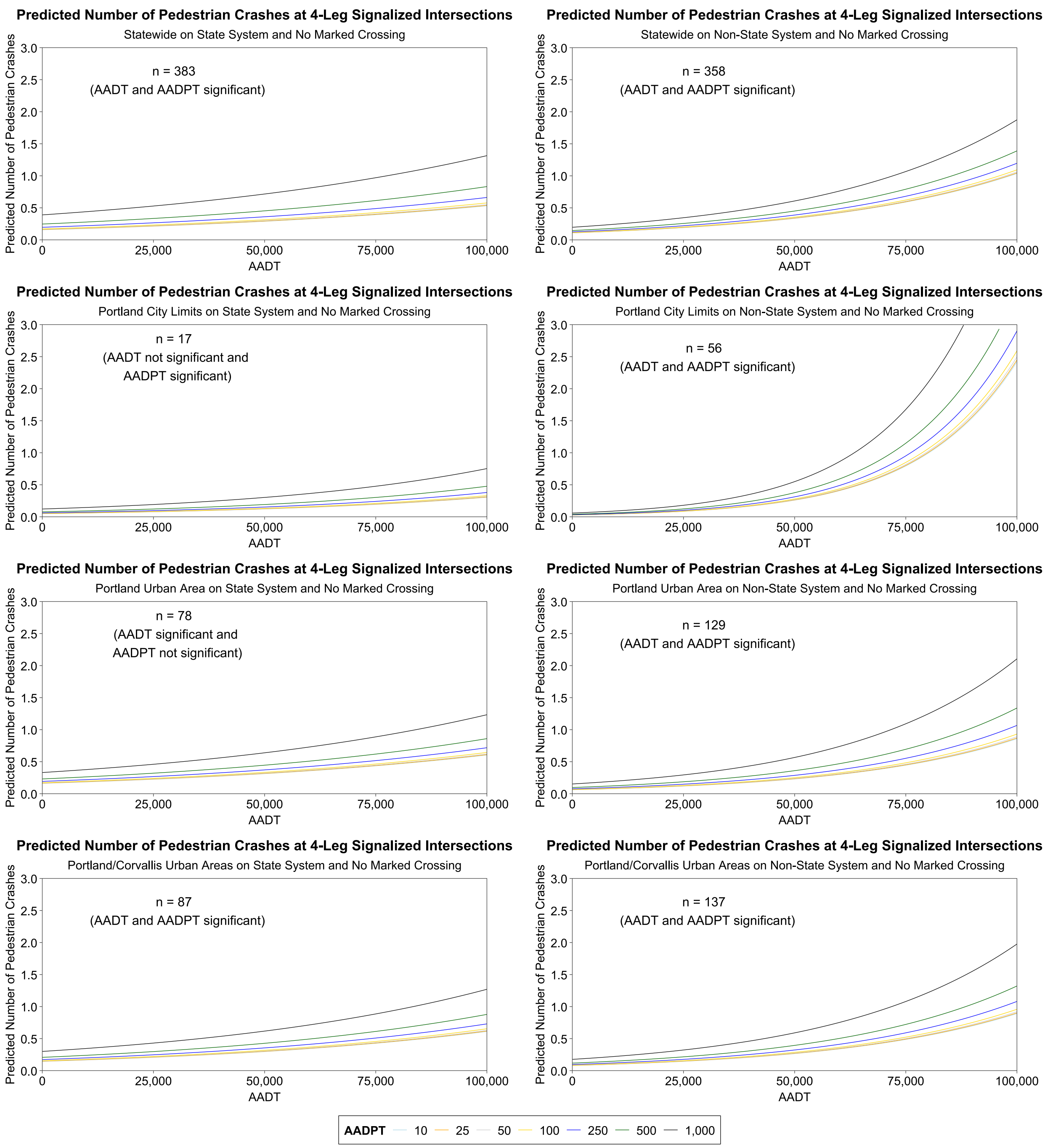


Figure 4.20: Predicted Number of Pedestrian Crashes at 4-Leg Signalized Intersections with No Marked Crossing

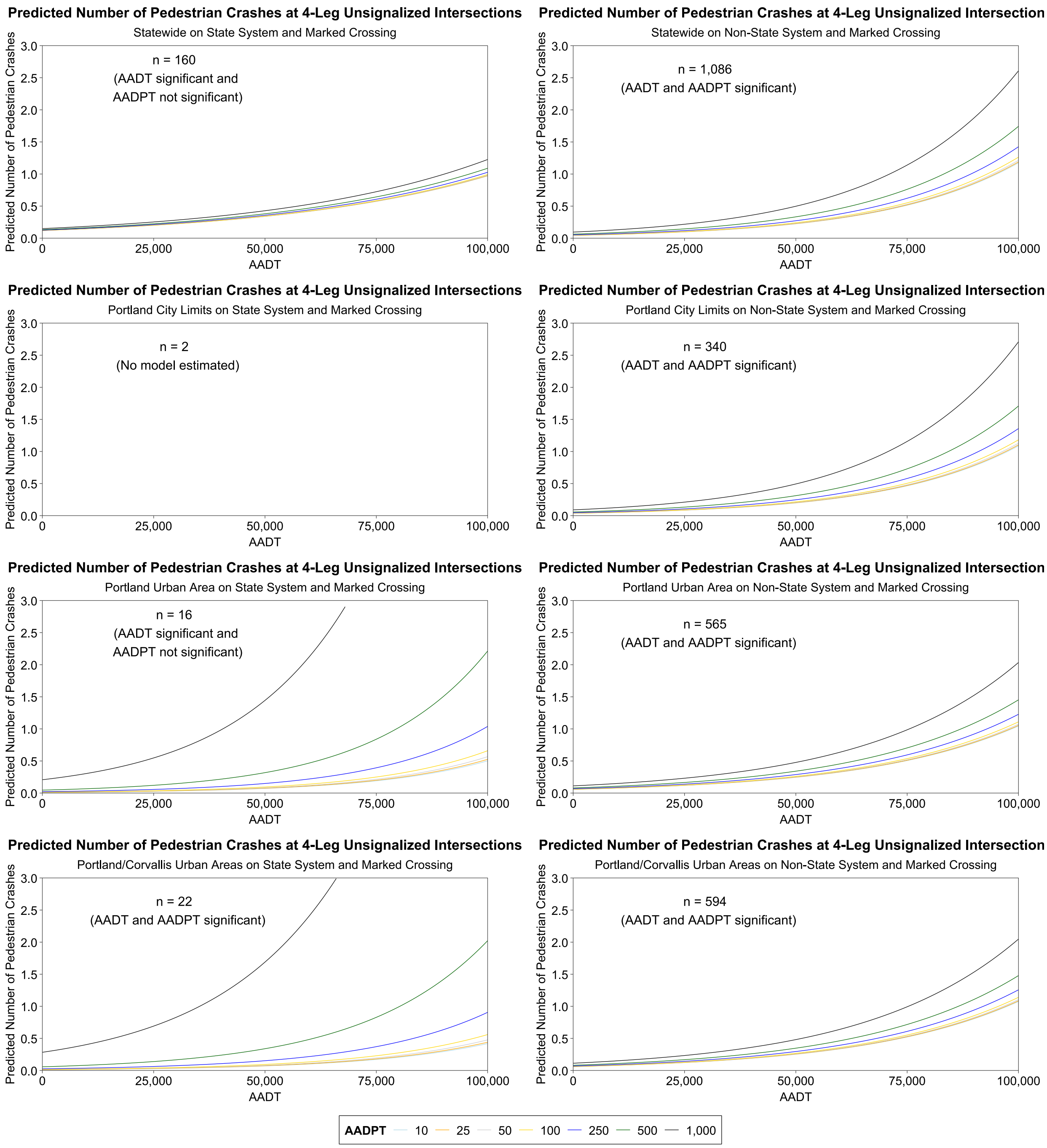


Figure 4.21: Predicted Number of Pedestrian Crashes at 4-Leg Unsignalized Intersections with Marked Crossing

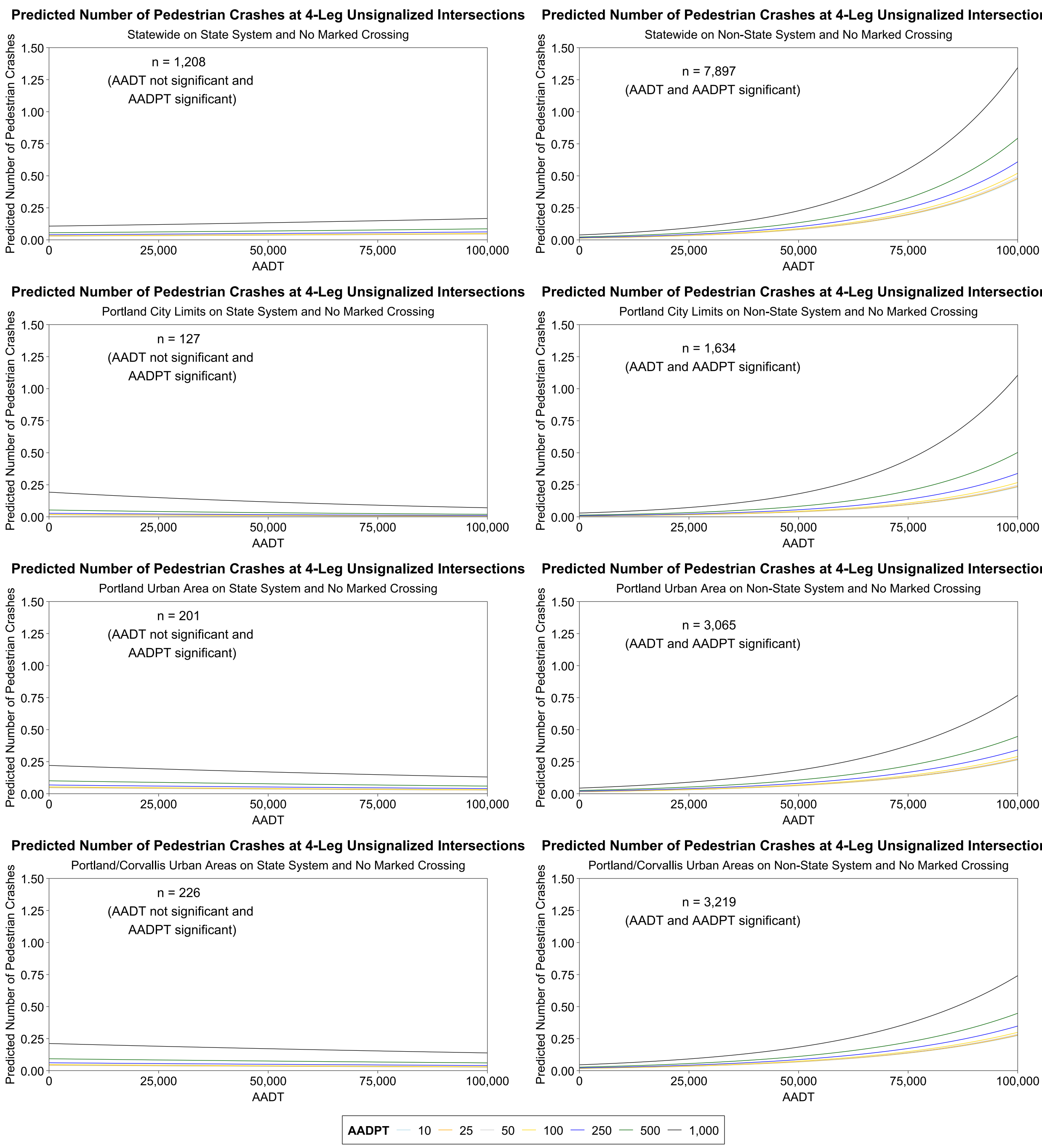


Figure 4.22: Predicted Number of Pedestrian Crashes at 4-Leg Unsignalized Intersections with No Marked Crossing

4.8 COMPARISON BETWEEN 3-LEG INTERSECTIONS

Figure 4.23 and Figure 4.24 show the comparisons of model coefficients for average annual daily traffic and average annual daily pedestrian traffic, respectively, across all 3-leg intersection models.

Similar to the 4-leg intersection models, in most 3-leg models, the coefficient for average annual daily traffic was positive and the 95% confidence interval did not include zero. This is generally associated with vehicle volume being a significant contributing factor with at least 95% confidence. This also indicates that among the majority of models, vehicle volume is expected to increase pedestrian crash frequency. These results are generally expected.

On the other hand, there were five models (two more than the 4-leg intersection models) in which the coefficients for average annual daily traffic were negative, contrary to expectations and previous research. Three of the five models were for the Portland/Corvallis Urban Areas on the non-state system. Despite there being no significance, the expectation is that the coefficients will still be positive, indicating an increase in the expected number of pedestrian crashes. When this is observed, it can stem from two potential sources: (1) omitted variable bias and (2) the model is unable to account for the large number of zeros, should a large number of zeros be present. Regardless of the reason, this scenario can impact the fitted values of the model when generating crash prediction plots. It should be noted that the number of zeros in the observed crash distributions was notably larger for 3-leg intersections for most locations, jurisdictions, and marked crossing considerations (see Appendix D).

Unlike the 4-leg intersection models, although average annual daily pedestrian traffic had a positive coefficient in all models but one, the significance of the effect of pedestrian volume varied. For example, pedestrian volume effects were not significant in the following models:

- Portland City Limits model (signalized, state system, and marked crossing)
- Statewide, Portland City Limits, and Portland Urban Area (signalized, non-state system, and marked crossing)
- All models (signalized, state system, and no marked crossing)
- All models (signalized, non-state system, and no marked crossing)
- All models (unsignalized, state system, and marked crossing)
- Portland City Limits (unsignalized, state system, and no marked crossing)

Overall, the 3-leg intersections had fewer crashes and more intersections with zero crashes. Additionally, with the exception of the unsignalized groups (non-state system with marked crossing, state system with no marked crossing, and non-state system with no marked crossing), the sample sizes were generally small relative to the 4-leg models. This was not exclusive to pedestrian volume, as vehicle volume effects were also insignificant across more models. These results were unexpected but could be addressed through some of the methods discussed in the proceeding paragraphs.

Akin to the 4-leg intersection models, the 3-leg intersection models contain only vehicle and pedestrian volumes, leading to the potential for omitted variable bias. If there are variables not included in the model that impact the dependent variable (pedestrian crash frequency) and have some correlation with other explanatory variables (vehicle or pedestrian volume), the effects of these variables can be picked up by the variables that were included in the model. For all models, if this is present, the variables picking up these effects would be vehicle volume or pedestrian volume. With this in mind, it is important to once more reiterate the reason for estimating models with only vehicle and pedestrian volume. As additional variables are included in model specifications, model complexity increases, and usability of the model can become more complex. In the context of this work, the data process to obtain additional variables to be used in the model may not always be achievable by transportation agencies. This leads to the consideration of usability versus complexity, where usability was chosen. To address this, additional variables can be included in model specifications, such as variables related to the built-environment, land use, socioeconomics, demographics, roadway design, and others.

To address a preponderance of zeros adds even more complexity, as more advanced and sophisticated models are required. To estimate such models, specific software is often required but it may not be readily available to transportation safety engineers or planners. Further, the more common models that address this zero inflation, and therefore the most accessible, should be carefully considered in a transportation safety context. The manner in which most of these models define zeros does not fully align with, in the case of this work, zero crash intersections. To restate, the concept is that no intersection is 100% safe, and because transportation safety includes human participants and human error, there is always a chance a crash can occur. While statistical fit may improve by applying one of these models, defending its practical application in this context may prove difficult. In the end, the decision to choose a model that accounts for a large number of zeros should be supported by a suite of models, tests for zero effects, and practical reasoning for assuming zeros belong in their own count state.

To estimate the expected number of crashes, each model was used to generate the crash prediction plots shown in Figure 4.25 to Figure 4.28. As was the case with the 4-leg models, all 3-leg crash prediction plots show the importance of vehicle volume in estimating pedestrian crash frequency. For plots that show typical crash estimation trends, the predictions remain consistent until large values of vehicle volume, which then results in a steep increase in the expected number of pedestrian crashes. The 3-leg crash prediction plots rely on the ranges of values used to estimate the model; therefore, some of the prediction plots increase steepling after vehicle volumes of 70,000 (the max used for estimation).

Of note with the 3-leg crash prediction plots are the various plots that do not follow typical behavior. One example is the collection of state system plots in Figure 4.25. In this example, the plot suggests that pedestrians crashes decrease with increased vehicle volume within Portland City Limits on the state system and with a marked crossing, a non-significant but still counterintuitive finding. These plots highlight two things: (1) the importance of vehicle volume and (2) the relationship to pedestrian crash frequency if the coefficient for average annual daily traffic is negative. For most state system models, this behavior was observed. These results suggest a need for more complex and sophisticated models.

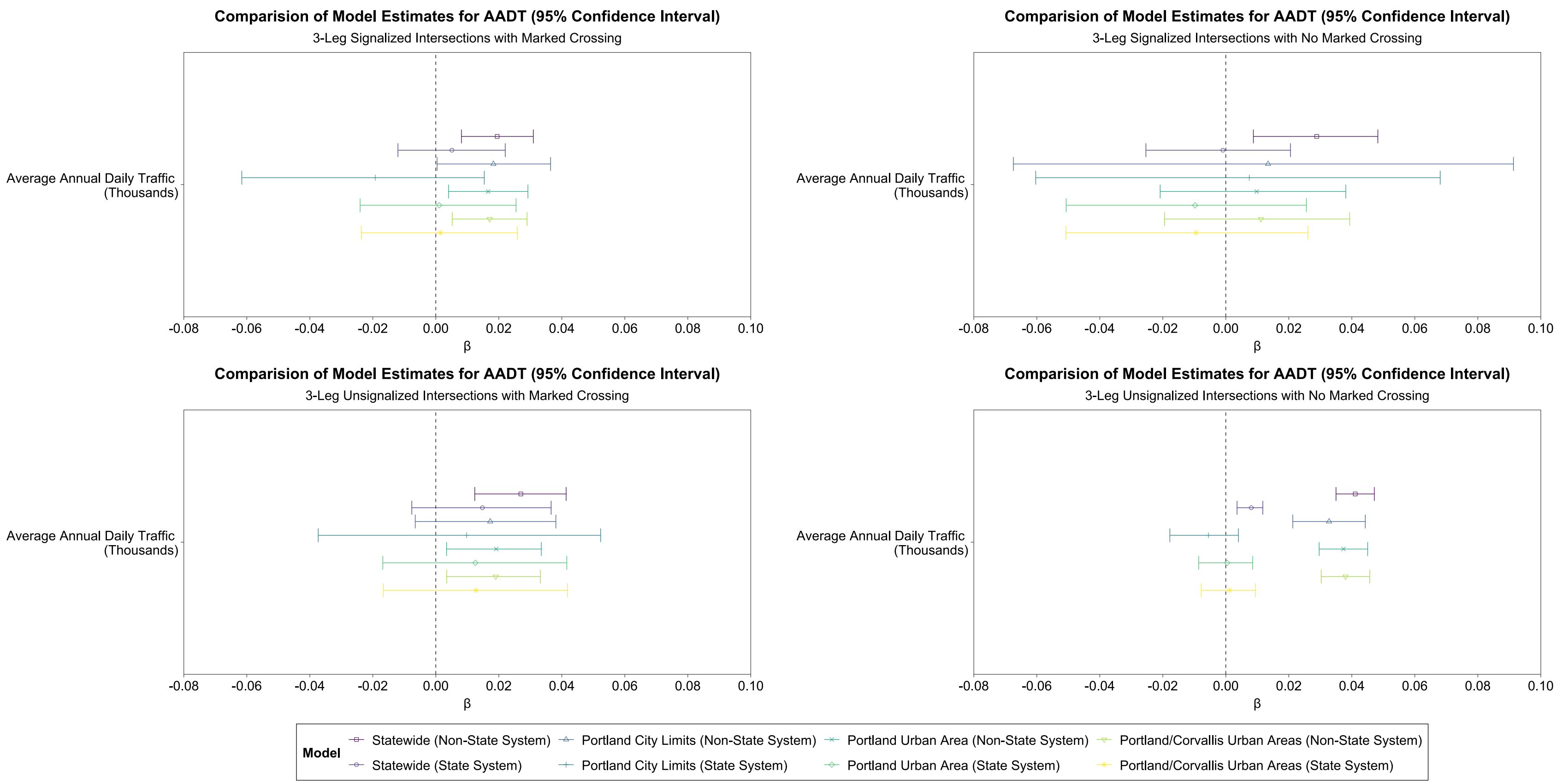


Figure 4.23: Comparison of Model Estimates for AADT Between 3-Leg Intersection Models

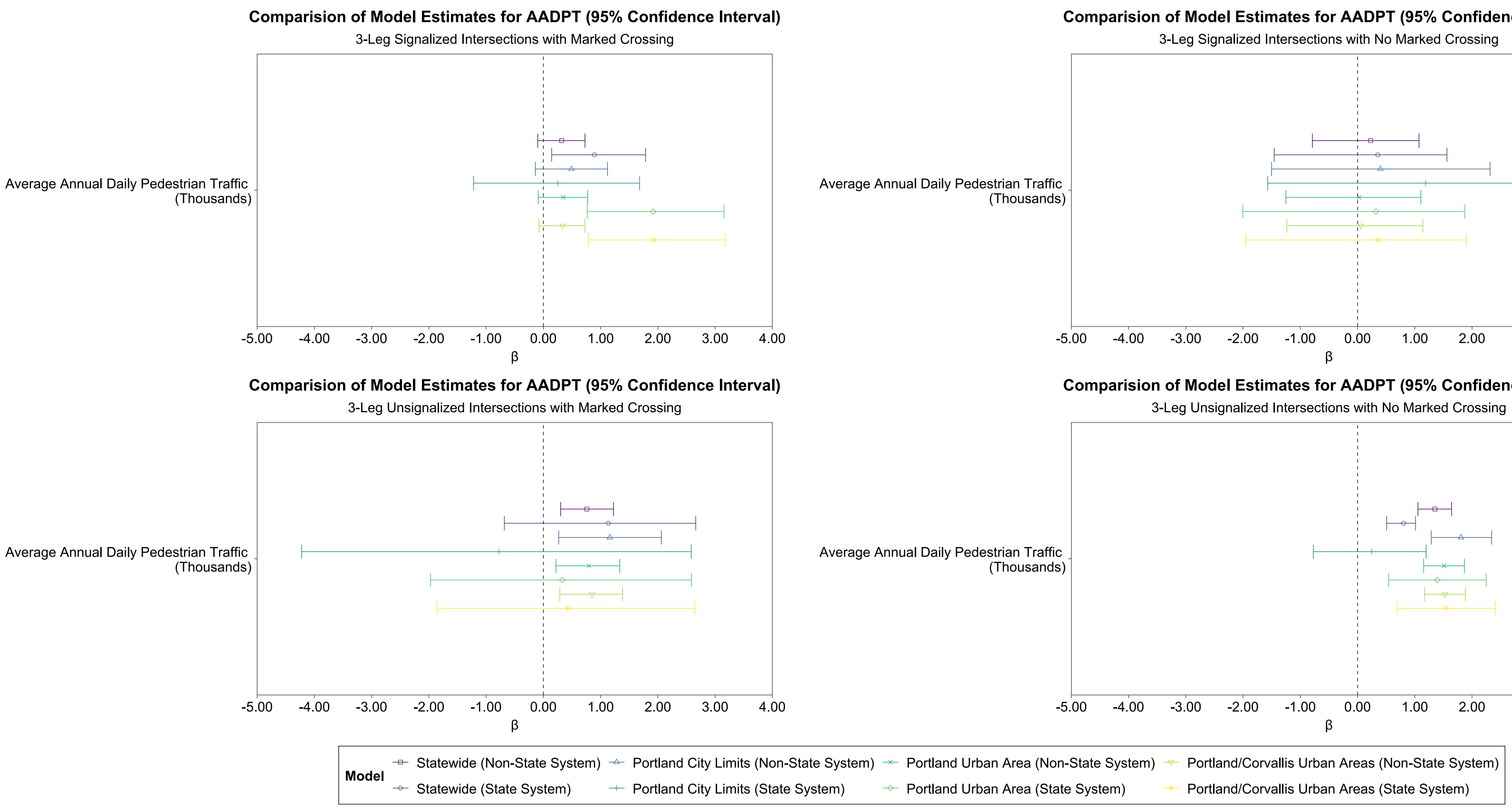


Figure 4.24: Comparison of Model Estimates for AADPT Between 3-Leg Intersection Models

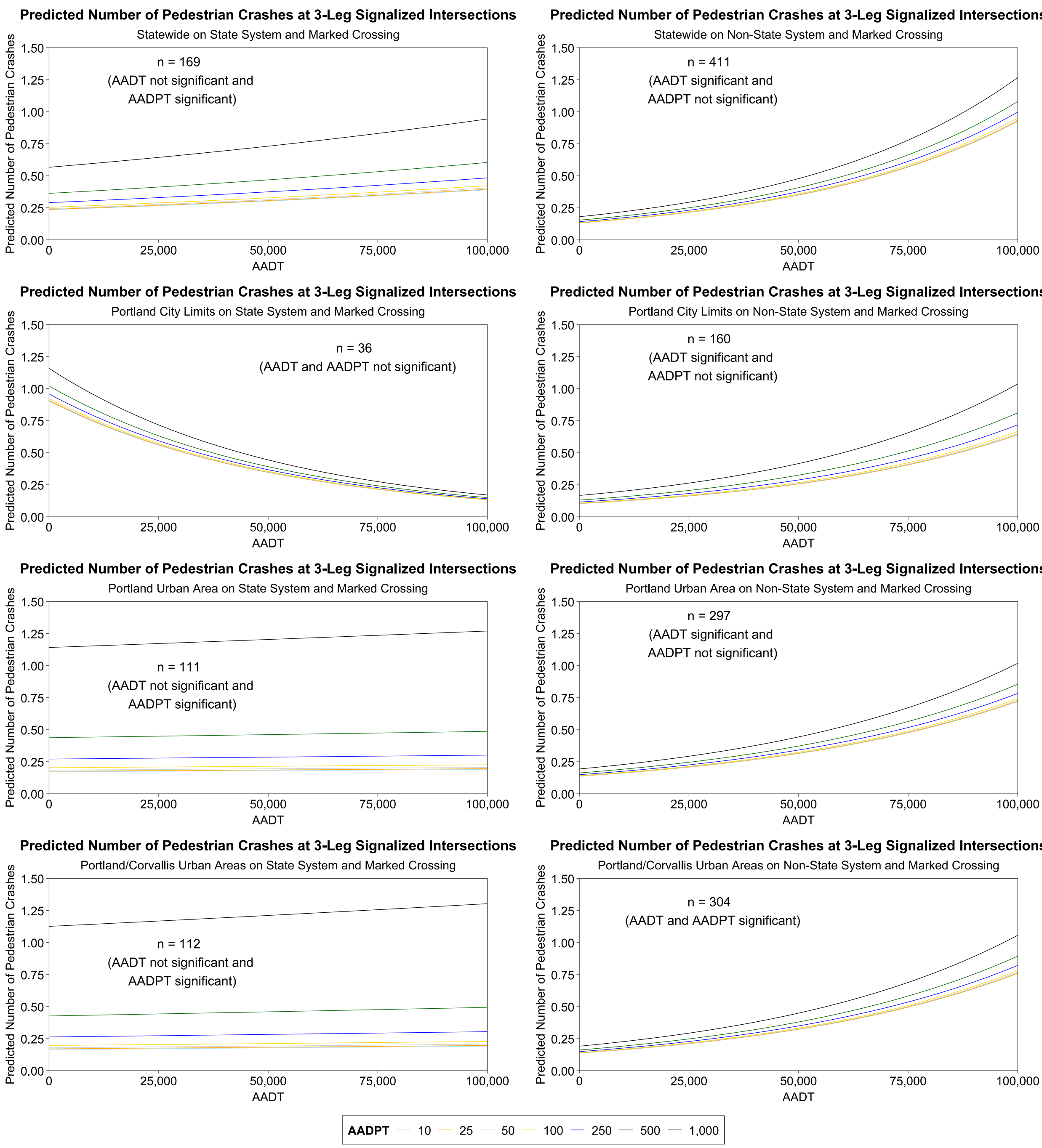


Figure 4.25: Predicted Number of Pedestrian Crashes at 3-Leg Signalized Intersections with Marked Crossing

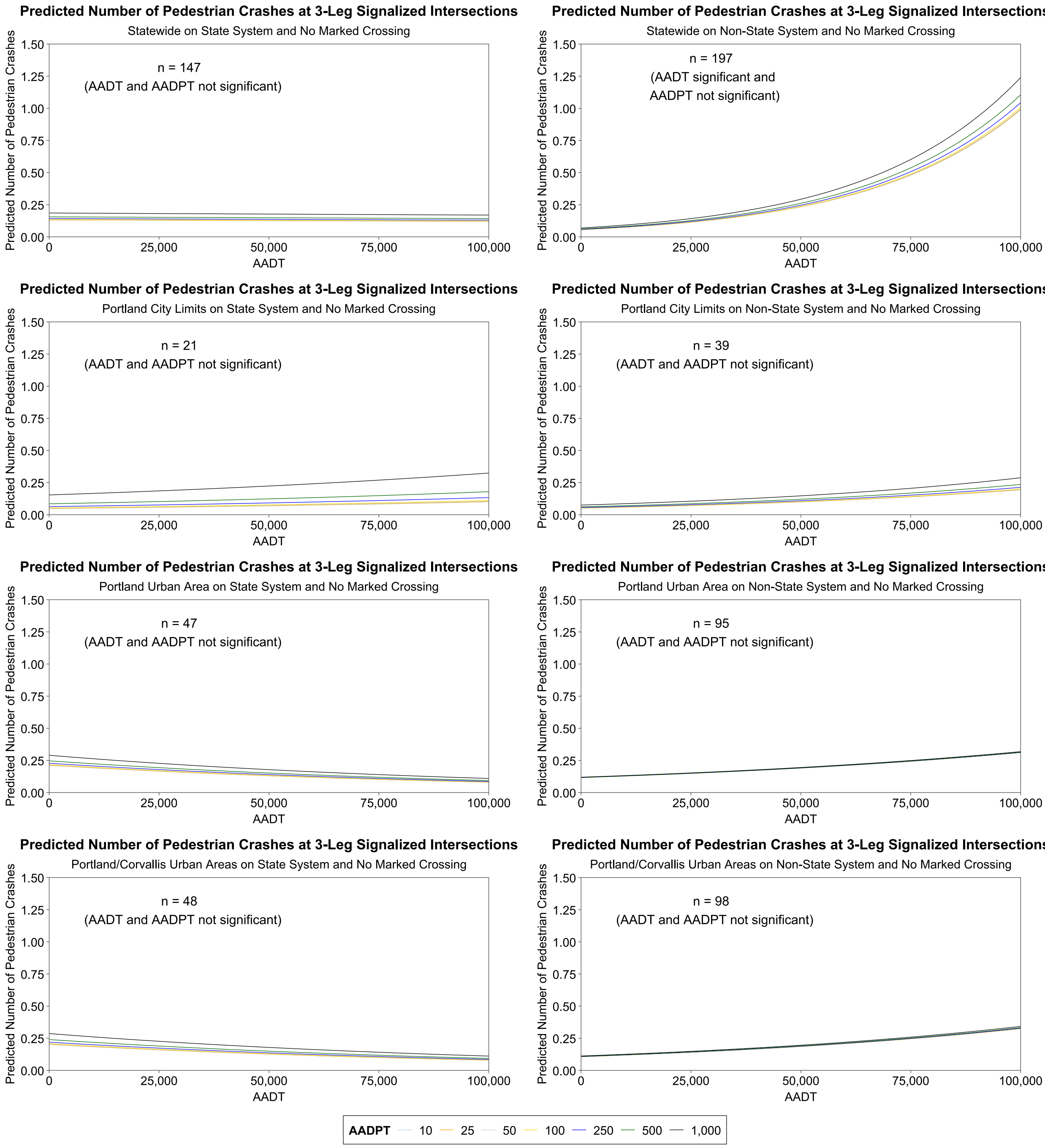


Figure 4.26: Predicted Number of Pedestrian Crashes at 3-Leg Signalized Intersections with No Marked Crossing

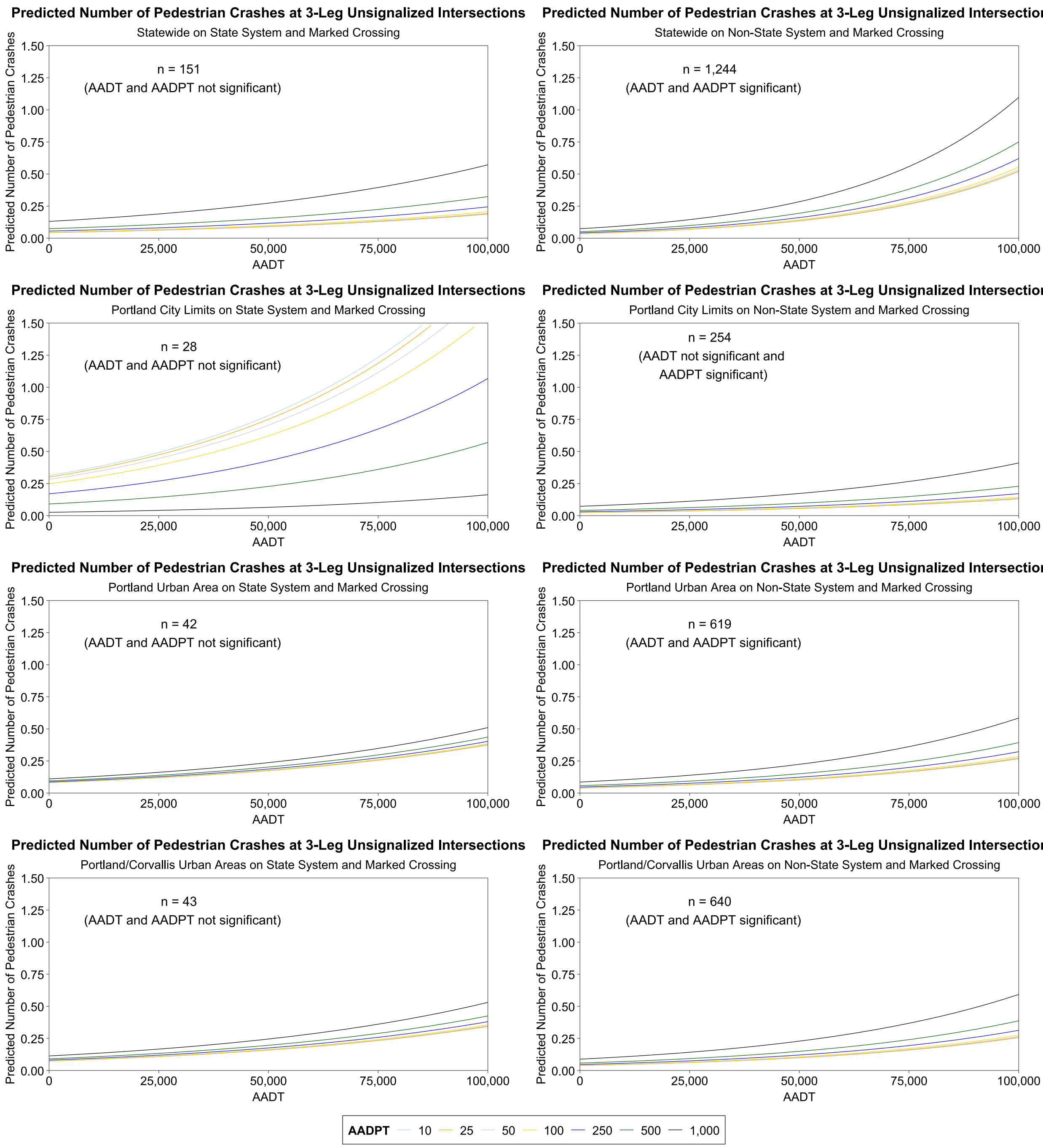


Figure 4.27: Predicted Number of Pedestrian Crashes at 3-Leg Unsignalized Intersections with Marked Crossing

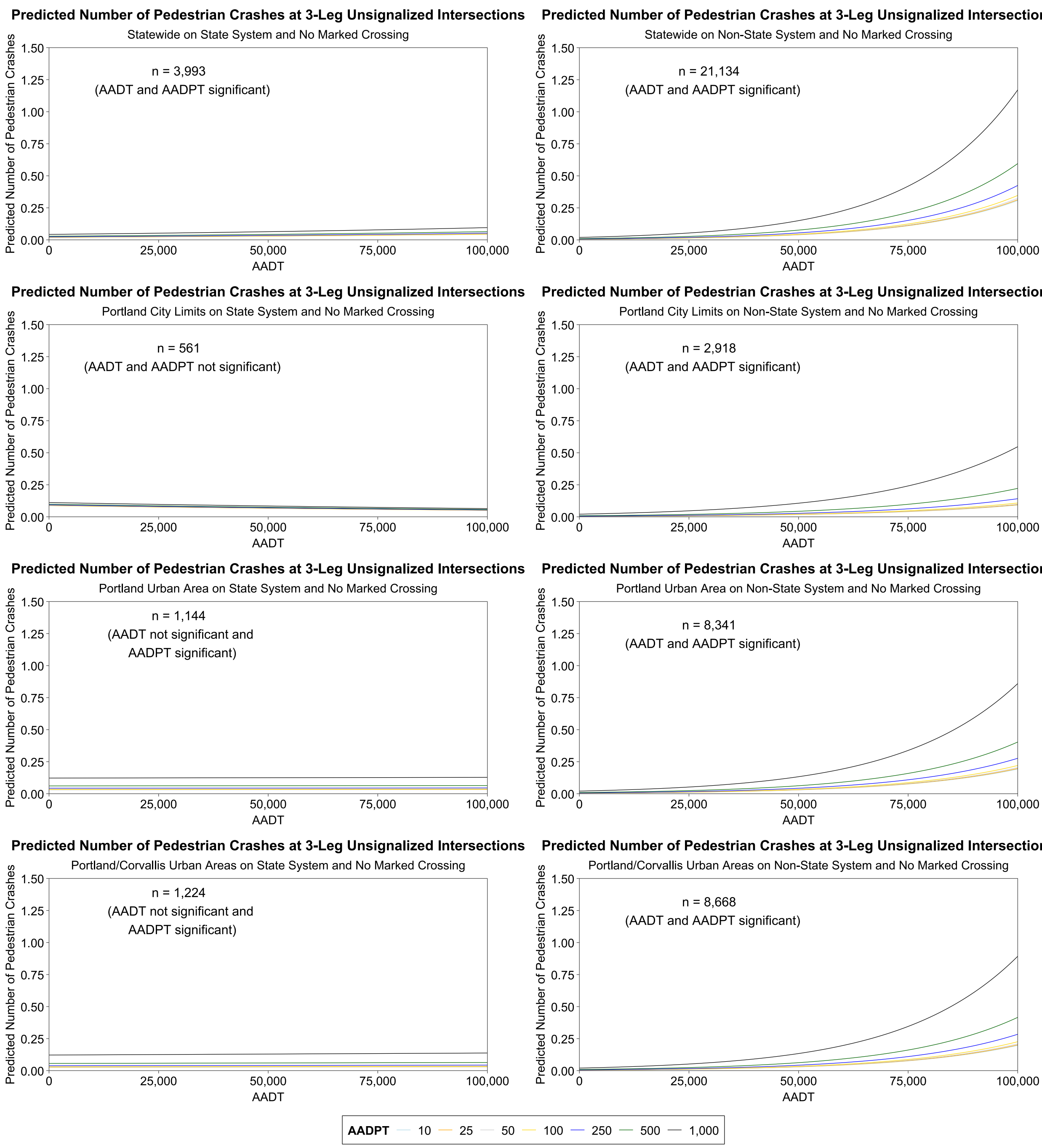


Figure 4.28: Predicted Number of Pedestrian Crashes at 3-Leg Unsignalized Intersections with No Marked Crossing

4.9 SUMMARY

Utilizing a vast database for pedestrian crashes in Oregon, a series of crash frequency models were estimated to develop safety performance functions for urban intersections. A disaggregate data approach was applied, such that models were estimated for different intersection typologies (4-leg and 3-leg, signalized and unsignalized), locations (statewide, Portland City Limits, Portland Urban Area, and Portland/Corvallis Urban Areas), and jurisdictions (state system and non-state system).

After careful consideration, each model included only vehicle exposure and pedestrian exposure. This was done in an attempt to keep the models accessible and increase usability.

Results from the 4-leg intersection models were generally consistent. Vehicle volume estimates were often significant and positive, indicating an increase in pedestrian crash frequency. This was also true for pedestrian volume estimates. There were cases in which the estimates for vehicle volume were negative (and insignificant), which likely stemmed from omitted variables or a large number of zeros. Omitted variables were a byproduct of the selection to include only vehicle and pedestrian exposure variables, while initial tests showed no significant zero-related effects.

The crash estimation plots illustrated the importance of vehicle volume in estimating pedestrian crash frequency and performed well over the range of values used for model estimation. Any issues observed with crash estimation were linked to sample size and/or limitations related to estimating only vehicle and pedestrian volumes.

Results from the 3-leg intersection models were consistent, but for different reasons. While the majority of estimates for vehicle and pedestrian volumes were positive, indicating increases in pedestrian crash frequency, the significance of their effects varied substantially. Nearly all models had estimates that were not significant. In general, the 3-leg intersections had fewer crashes and more intersections with zero crashes, which likely led to some of these results. This led to some issues with the crash estimations for disaggregate groups, in which model estimates were insignificant. These issues were observed widely, but most prevalent for the model groups with no marked crossing.

Despite some of these challenges, results do indicate that Oregon-specific pedestrian safety performance functions can be developed and used. In nearly every disaggregation, the statewide models performed well, resulting in a model that can be used for pedestrian safety analysis in Oregon. Many of the 4-leg intersection models that were location-specific can also be used, and for any location where the model did not perform well, the statewide model can be applied. This work provides Oregon with a state-specific pedestrian crash frequency model representing Oregon crash behavior.

4.9.1 Limitations

While results are generally promising, there are some distinct limitations that can be addressed in future work.

The decision to include only vehicle and pedestrian volume data can be modified. When making the decision between usability and model complexity, there are inherent limitations by selecting usability. This was observed in some of the models, where omitted variable bias may be present. To address this, the presented models can be extended to include a variety of additional built-environment, land use, socioeconomic, demographic, roadway design, and other information. The data exists, and some of these results (in the form of model fit measures, not specifications) are given in Appendix A. Overall model fit will improve while increasing model complexity. This approach can be accompanied by a workshop or tool to help agency personnel understand and use more comprehensive models and data.

Another limitation also stems from the decision to prioritize usability over model complexity. While preliminary results suggest no significant zero-related effects, the use of a more advanced and sophisticated model to account for the presence of zeros may improve model fit and crash estimation. The tradeoff for these models is starker than just adding additional variables, as the modeling framework itself must be modified and, often times, specific software is required. Additionally, the interpretation of model estimates becomes more complex, which can lead to potential adoption and application issues. It is recommended to explore these modeling frameworks, and where appropriate, apply them to generate novel, innovative pedestrian safety performance functions. To assist with agency adoption, the application of such models can be illustrated through agency sponsored workshops and tutorial sessions led by university partners.

An additional limitation worth noting is the exclusion of safety performance functions for serious crashes only (fatal and incapacitating). Models for these types of crashes were considered but required a complex modeling framework to account for the presence of significant underdispersion. Underdispersion is much less common in transportation safety analysis, as are the count-data models required to account for this data limitation. Due to traditional safety performance function methods (e.g., Poisson and negative binomial) not being appropriate for significant underdispersion, alternate count-data modeling frameworks are required. While in-depth testing and evaluation of these methods was beyond the scope of the current study, this offers a unique opportunity to apply these alternate methodologies to develop a safety performance function framework for data that is underdispersed. In the end, such an endeavor could lead to breakthroughs in active transportation safety analysis. In addition, Oregon could maintain and update more complex pedestrian safety performance functions by utilizing tools developed through research.

Regarding data limitations, this work relied on estimates for both vehicle volume, and even more so for pedestrian volume. While demand models and metrics indicate good fit and reliable volume estimates, there is still error introduced. Some of the results that were unexpected or counterintuitive may also stem from the error introduced by estimating vehicle and pedestrian exposure. For 4-leg models, unexpected behavior was observed with small samples sizes and intersections with no marked crossing. Error will be more noticeable in small sample sizes, while no marked crossing locations rely almost entirely on estimated volumes. This trend was also true for 3-leg models. This creates an opportunity to explicitly focus on volume estimation and compare a variety of methods, then test each of those methods in a safety performance context.

Lastly, no types of heterogeneity were accounted for in this analysis. Future work can consider both latent and spatial heterogeneity. Applications addressing these in the transportation safety literature always results in improved model fit, which in turn improves crash estimation and network screening.

5.0 REFERENCES

AASHTO. (2010). *Highway safety manual*. https://trid.trb.org/View/1150867

AASHTO. (2014). *Highway safety manual, first edition, 2014 supplement* (Pub. Code HSM-1S). https://trid.trb.org/View/1328763

Broach, J., Kothuri, S., Miah, M. M., McNeil, N., Hyun, K., Mattingly, S., Nordback, K., & Proulx, F. (2023). Evaluating the potential of crowdsourced data to estimate network-wide bicycle volumes. *Transportation Research Record, 2678*(3), 573–589. https://doi.org/10.1177/03611981231182388

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Brüde, U., & Larsson, J. (1993). Models for predicting accidents at junctions where pedestrians and cyclists are involved: How well do they fit? *Accident Analysis & Prevention, 25*(5), 499–509. https://doi.org/10.1016/0001-4575(93)90001-D

Clifton, K. J., Burnier, C. V., & Akar, G. (2009). Severity of injury resulting from pedestrian–vehicle crashes: What can we learn from examining the built environment? *Transportation Research Part D: Transport and Environment, 14*(6), 425–436. https://doi.org/10.1016/j.trd.2009.01.001

Csárdi, G., Nepusz, T., Traag, V., Horvát, S., Zanini, F., Noom, D., & Müller, K. (2025). *igraph: Network analysis and visualization in R* (Version 2.1.4) [R package]. https://doi.org/10.5281/zenodo.7682609

Dai, D., & Jaworski, D. (2016). Influence of built environment on pedestrian crashes: A network-based GIS analysis. *Applied Geography, 73*, 53–61. https://doi.org/10.1016/j.apgeog.2016.06.005

Dolatsara, H. A. (2014). *Development of safety performance functions for non-motorized traffic safety* (Master's thesis). Western Michigan University.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Federal Highway Administration. (2016). *Utah's and Kentucky's innovative use of GIS-based tools: Strategies for using GIS to advance highway safety* (FHWA-SA-16-028). https://safety.fhwa.dot.gov/rsdp/downloads/fhwasa16028.pdf

Federal Highway Administration. (2021, February 19). *FHWA's crash tree diagram tool: Developing and using crash trees* [Video]. https://www.youtube.com/watch?v=QNUZ5b-J3Dw

Federal Highway Administration. (2022). *Traffic monitoring guide* (2022 ed.). U.S. Department of Transportation. https://www.fhwa.dot.gov/policyinformation/tmguide/2022_TMG_Final_Report.pdf

Federal Highway Administration. (n.d.). *The systemic approach to safety*. Retrieved August 5, 2022, from https://safety.fhwa.dot.gov/systemic/

FHWA Roadway Safety Data Program. (n.d.). *Oregon Safety Priority Index System (SPIS)*. Retrieved August 5, 2022, from https://safety.fhwa.dot.gov/rsdp/toolbox-content.aspx?toolid=87

Gates, T. J., Savolainen, P. T., Stapleton, S., Kirsch, T., & Miraskar, S. (2016). *Development of safety performance functions and other decision support tools to assess pedestrian and bicycle safety* (TRCLC 14-6). Wayne State University. https://trid.trb.org/View/1435143

Guerra, E., Dong, X., & Kondo, M. (2019). Do denser neighborhoods have safer streets? Population density and traffic safety in the Philadelphia region. *Journal of Planning Education and Research*. https://doi.org/10.1177/0739456X19845043

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182. https://doi.org/10.1162/153244303322753616

Harwood, D. W., Bauer, K. M., Richard, K. R., Gilmore, D. K., Graham, J. L., Potts, I. B., … Hauer, E. (2008). *Pedestrian safety prediction methodology* (NCHRP Web-Only Document 129). Transportation Research Board. https://trid.trb.org/view/864886

Hu, Y., Zhang, Y., & Shelton, K. S. (2018). Where are the dangerous intersections for pedestrians and cyclists? *Transportation Research Part C: Emerging Technologies, 95*, 431–441. https://doi.org/10.1016/j.trc.2018.07.030

Kaiser, S. K., Klein, N., & Kaack, L. H. (2025). From counting stations to city-wide estimates: Data-driven bicycle volume extrapolation. *Environmental Data Science, 4*, e13. https://doi.org/10.1017/eds.2025.5

Kim, D. (2019). The transportation safety of elderly pedestrians. *Accident Analysis & Prevention, 131*, 268–274. https://doi.org/10.1016/j.aap.2019.07.009

Kothuri, S., Hyun, K., Mattingly, S., & McNeil, N. (2024). *Exploring the use of crowdsourced data sources for pedestrian count estimations* (NITC-RR-1489). Transportation Research and Education Center.

Kothuri, S., Singleton, P., Vahedi Saheli, M., Yates, E., & Broach, J. (2024). *Active transportation counts from existing on-street signal and detection infrastructure* (FHWA-OR-RD-24-03). Oregon Department of Transportation.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Lee, K., & Sener, I. N. (2020). Strava Metro data for bicycle monitoring: A literature review. *Transport Reviews, 41*(1), 27–47. https://doi.org/10.1080/01441647.2020.1798558

Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R News, 2*(3), 18–22. https://CRAN.R-project.org/doc/Rnews/

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 from generalized linear mixed-effects models revisited. *Journal of the Royal Society Interface, 14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Nosal, T., Miranda-Moreno, L. F., & Krstulic, Z. (2014). Incorporating weather: Comparative analysis of AADB estimation methods. *Transportation Research Record, 2468*, 100–110.

Oregon Department of Transportation. (2021). *Nonmotorized traffic monitoring and crash analysis* (FHWA-OR-RD-21-08). https://www.oregon.gov/ODOT/TD/TP_RES/

Roll, J. (2021). Daily traffic count imputation for bicycle and pedestrian traffic. *Transportation Research Record, 2675*(11), 1428–1440. https://doi.org/10.1177/03611981211027161

Roll, J. F. (2023, January 10). *Evaluating third-party traffic volume data* [Conference presentation]. Transportation Research Board Annual Meeting, Washington, DC.

Singleton, P. A., & Runa, F. (2021). Pedestrian traffic signal data accurately estimates pedestrian crossing volumes. *Transportation Research Record, 2675*(6), 429–440. https://doi.org/10.1177/0361198121994126

Tao, T., Lindsey, G., Stern, R., & Levin, M. (2024). The use of crowdsourced mobile data in estimating pedestrian and bicycle traffic. *Journal of Transport and Land Use, 17*(1), 41–65. https://doi.org/10.5198/jtlu.2024.2315

U.S. Environmental Protection Agency. (2021). *Smart Location Database* (Version 3.0). https://catalog.data.gov/dataset/smart-location-database8

Wright, M., & Ziegler, A. (2017). ranger: A fast implementation of random forests. *Journal of Statistical Software, 77*(1), 1–17. https://doi.org/10.18637/jss.v077.i01

APPENDIX A

Table A.1: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Statewide, on State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.613	0.835	48.018	144.382	-453.42
Exposure-Only Model	0.683	0.906	49.373	144.983	-480.19

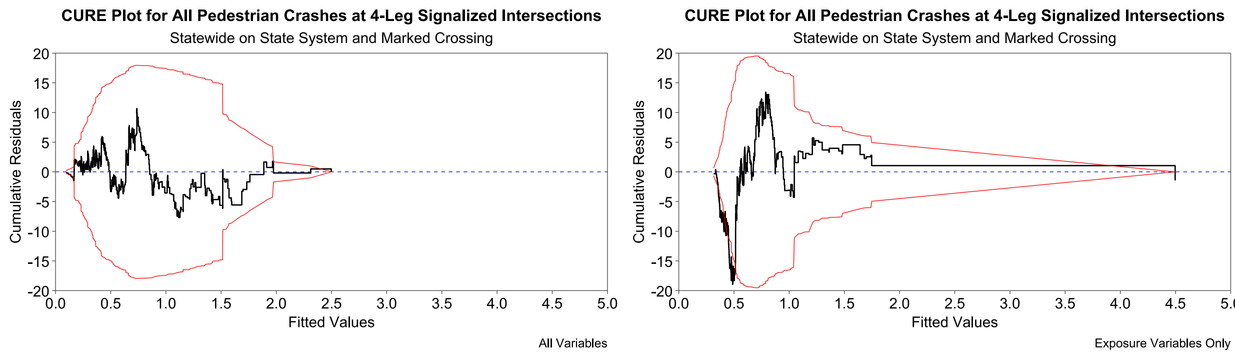


Figure A.1: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Statewide, on State System, and Marked Crossing

Table A.2: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Statewide, on Non-State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.651	0.900	51.837	151.656	-453.42
Exposure-Only Model	0.680	0.927	50.986	151.670	-480.19

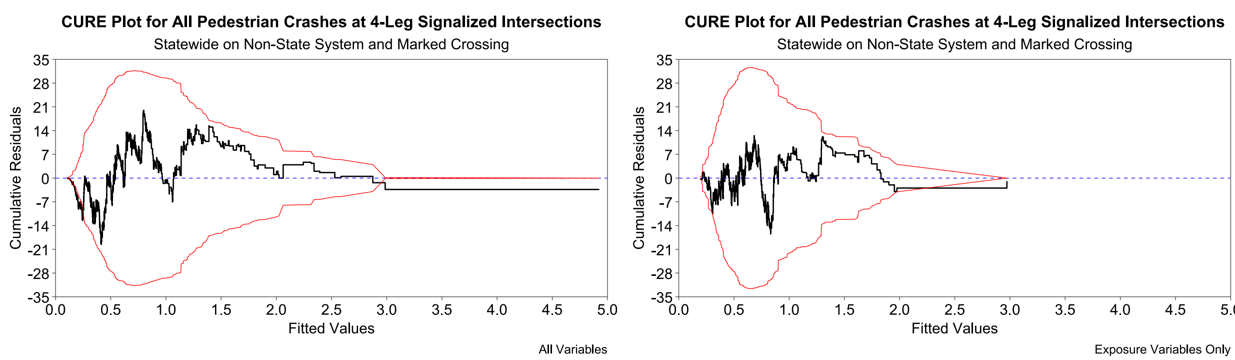


Figure A.2: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Statewide, on Non-State System, and Marked Crossing

Table A.3: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland City Limits, on State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.598	0.849	38.739	117.526	-115.03
Exposure-Only Model	0.718	0.953	37.517	118.439	-128.64

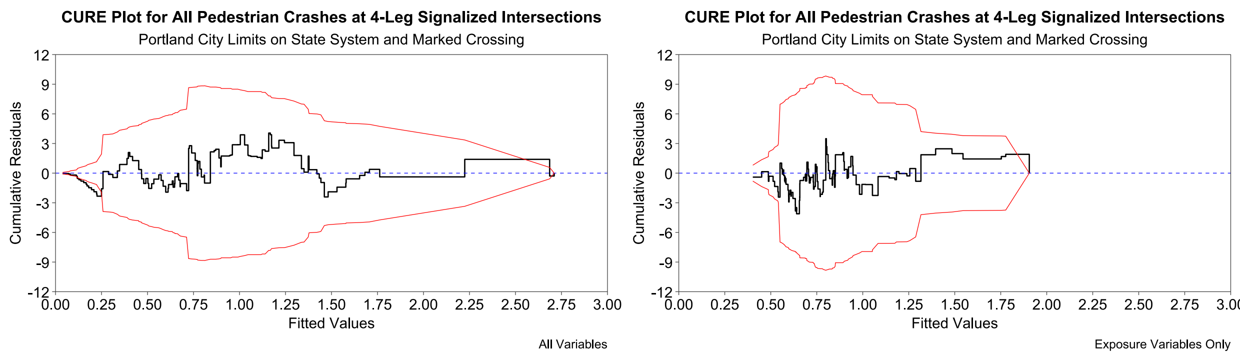


Figure A.3: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland City Limits, on State System, and Marked Crossing

Table A.4: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland City Limits, on Non-State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.724	1.006	50.101	141.616	-115.03
Exposure-Only Model	0.774	1.050	47.077	141.149	-128.64

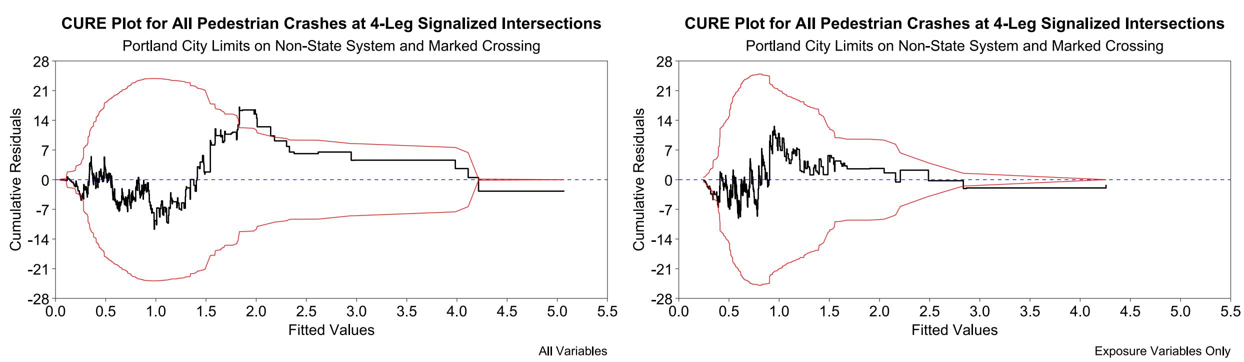


Figure A.4: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland City Limits, on Non-State System, and Marked Crossing

Table A.5: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland Urban Area, on State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.633	0.822	46.662	134.066	-244.21
Exposure-Only Model	0.725	0.931	42.731	132.922	-270.12

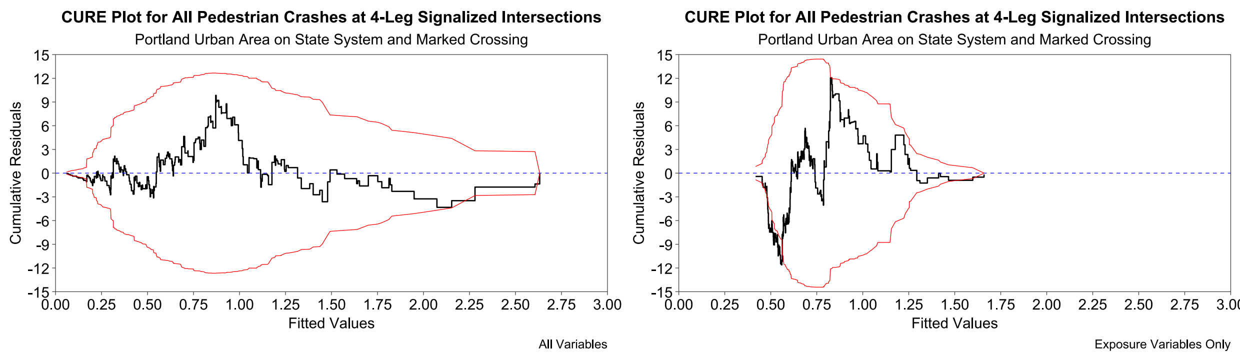


Figure A.5: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland Urban Area, on State System, and Marked Crossing

Table A.6: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland Urban Area, on Non-State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.693	0.956	50.662	146.636	-244.21
Exposure-Only Model	0.730	0.992	48.677	146.491	-270.12

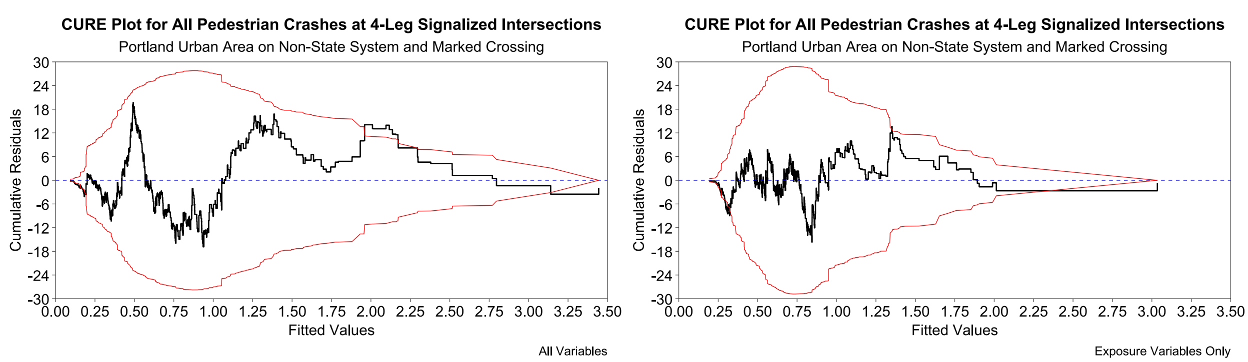


Figure A.6: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland Urban Area, on Non-State System, and Marked Crossing

Table A.7: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland/Corvallis Urban Areas, on State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.594	0.778	43.619	121.141	-110.26
Exposure-Only Model	0.683	0.902	40.161	123.731	-144.58

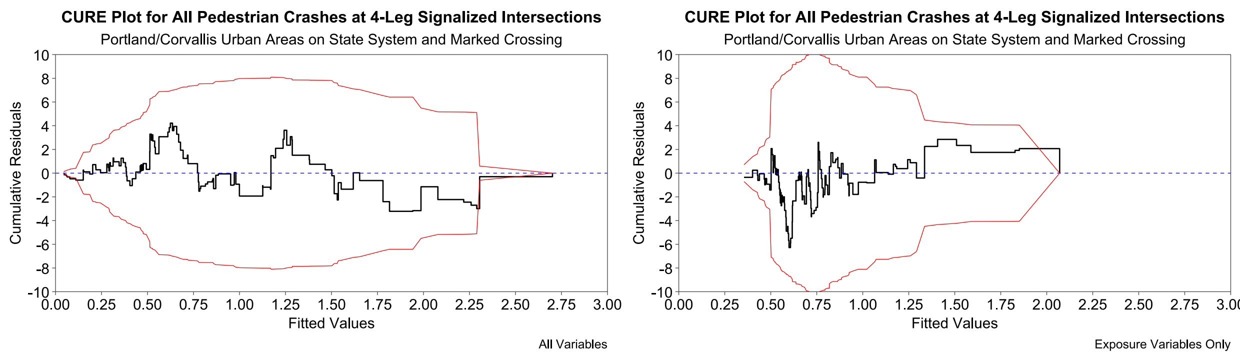


Figure A.7: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland/Corvallis Urban Areas, on State System, and Marked Crossing

Table A.8: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland/Corvallis Urban Areas, on Non-State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.731	1.012	49.384	141.929	-110.26
Exposure-Only Model	0.764	1.035	47.481	141.805	-144.58

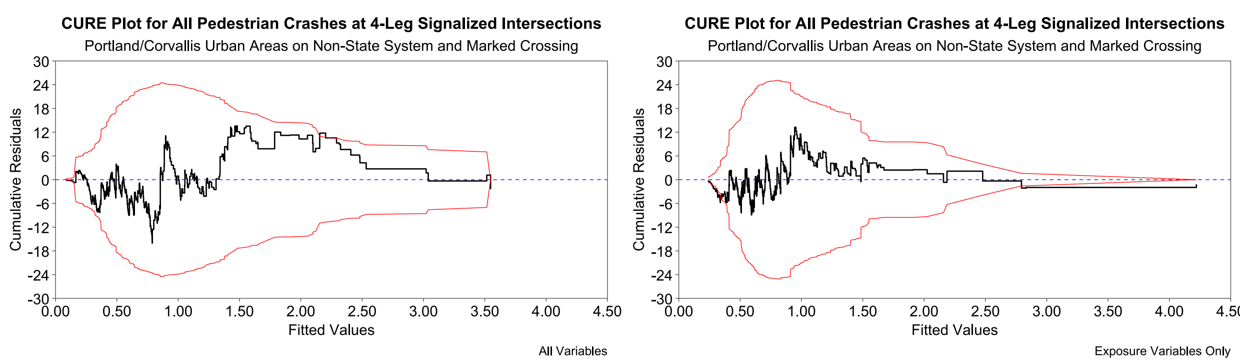


Figure A.8: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland/Corvallis Urban Areas, on Non-State System, and Marked Crossing

Table A.9: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Statewide, on State System, and No Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.434	0.627	61.602	172.293	-262.80
Exposure-Only Model	0.504	0.674	68.627	175.752	-288.45

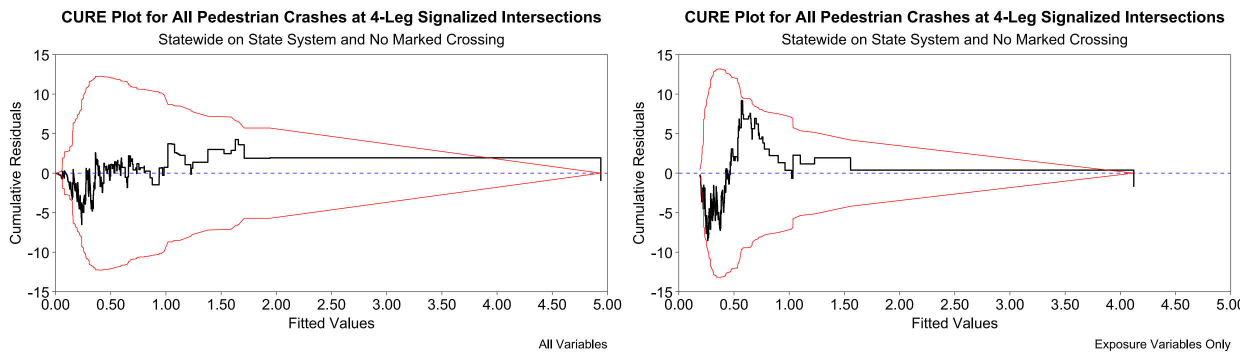


Figure A.9: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Statewide, on State System, and No Marked Crossing

Table A.10: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Statewide, on Non-State System, and No Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.446	0.671	56.477	168.127	-255.92
Exposure-Only Model	0.550	0.791	62.604	171.305	-287.83

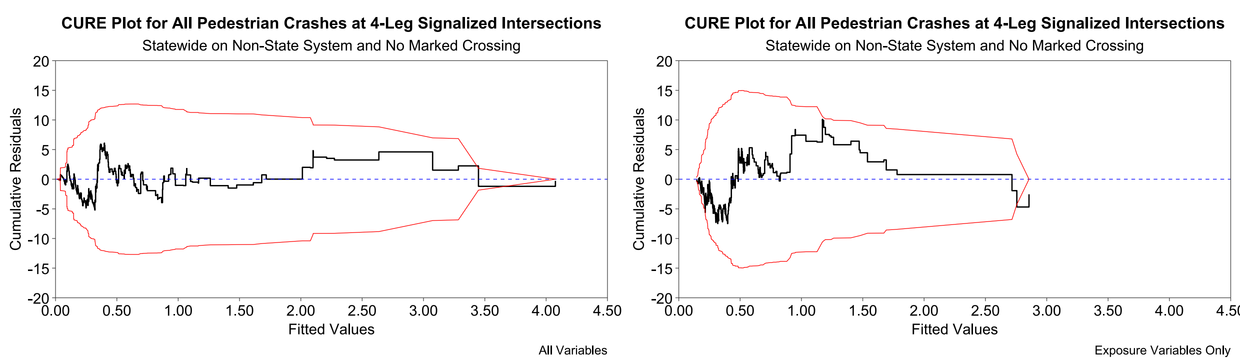


Figure A.10: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Statewide, on Non-State System, and No Marked Crossing

Table A.11: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland City Limits, on Non-State System, and No Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.265	0.455	42.734	163.299	-29.48
Exposure-Only Model	0.562	0.811	64.242	173.142	-44.92

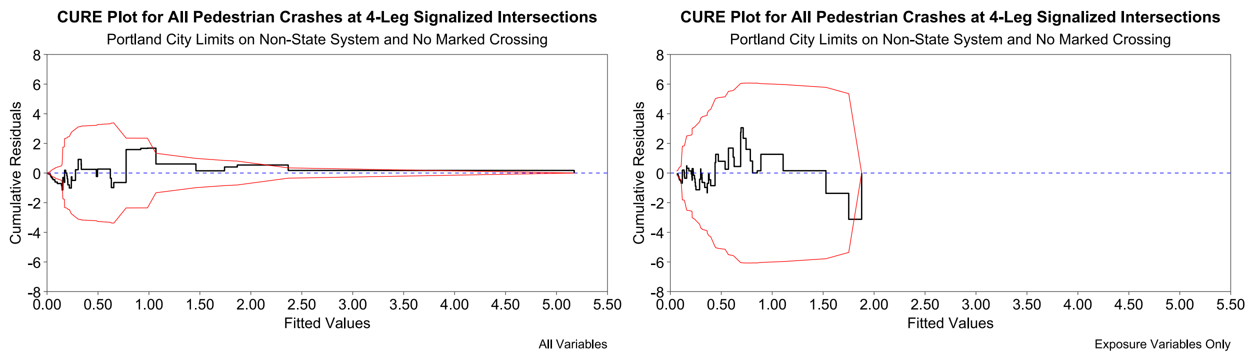


Figure A.11: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland City Limits, on Non-State System, and No Marked Crossing

Table A.12: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland Urban Area, on State System, and No Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.446	0.705	43.261	156.338	-58.83
Exposure-Only Model	0.681	0.878	59.927	164.291	-74.79

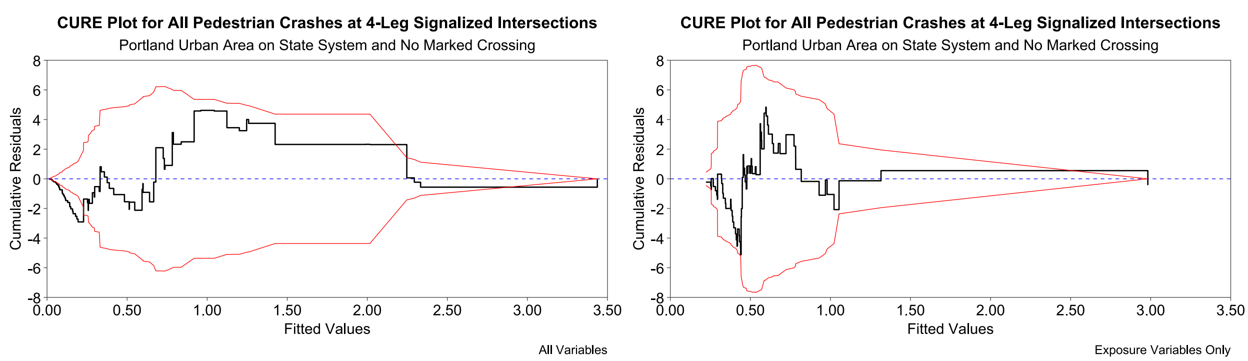


Figure A. 12: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland Urban Area, on State System, and No Marked Crossing

Table A.13: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland Urban Area, on Non-State System, and No Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.548	0.832	59.181	165.636	-103.52
Exposure-Only Model	0.665	0.992	59.43	167.043	-114.17

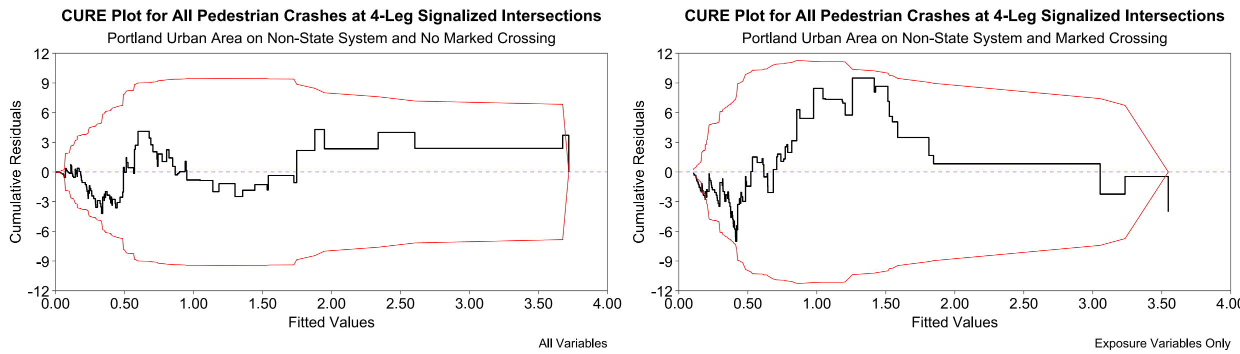


Figure A.13: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland Urban Area, on Non-State System, and No Marked Crossing

Table A.14: Comparison Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland/Corvallis Urban Areas, on Non-State System, and No Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.366	0.589	53.261	159.671	-40.246
Exposure-Only Model	0.572	0.779	61.583	166.701	-53.374

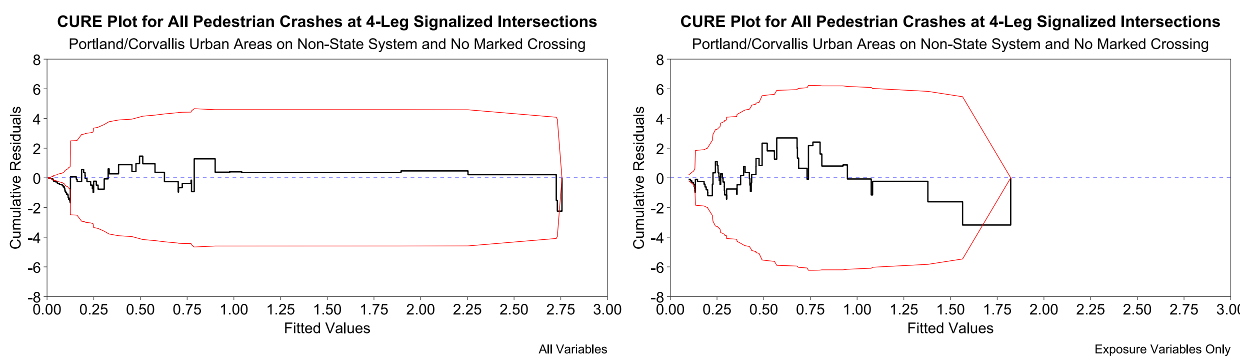


Figure A.14: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Signalized Intersections, Portland/Corvallis Urban Areas, on Non-State System, and No Marked Crossing

Table A.15: Comparison Between Full Model and Exposure-Only Model at 4-Leg Unsignalized Intersections, Statewide, on State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.283	0.421	64.291	180.998	-77.77
Exposure-Only Model	0.360	0.490	77.213	186.088	-92.70

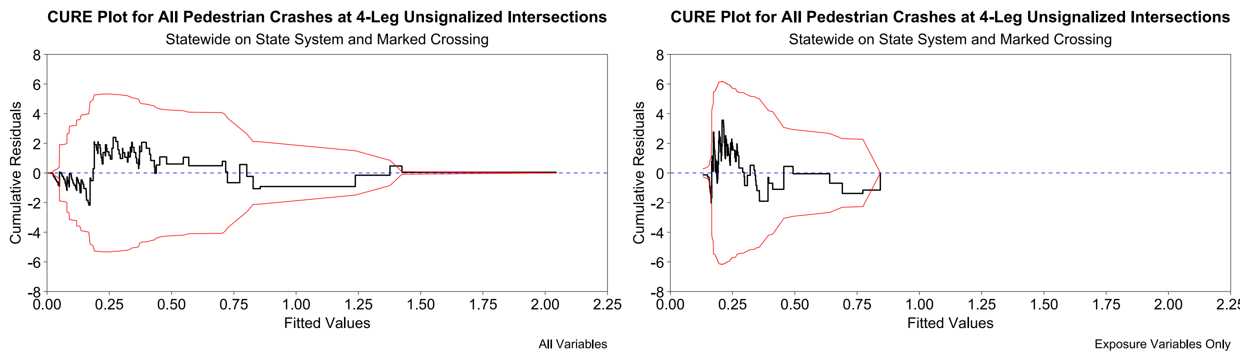


Figure A.15: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Unsignalized Intersections, Statewide, on State System, and Marked Crossing

Table A.16: Comparison Between Full Model and Exposure-Only Model at 4-Leg Unsignalized Intersections, Statewide, on Non-State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.215	0.402	77.872	192.507	-406.71
Exposure-Only Model	0.231	0.427	82.498	193.451	-426.92

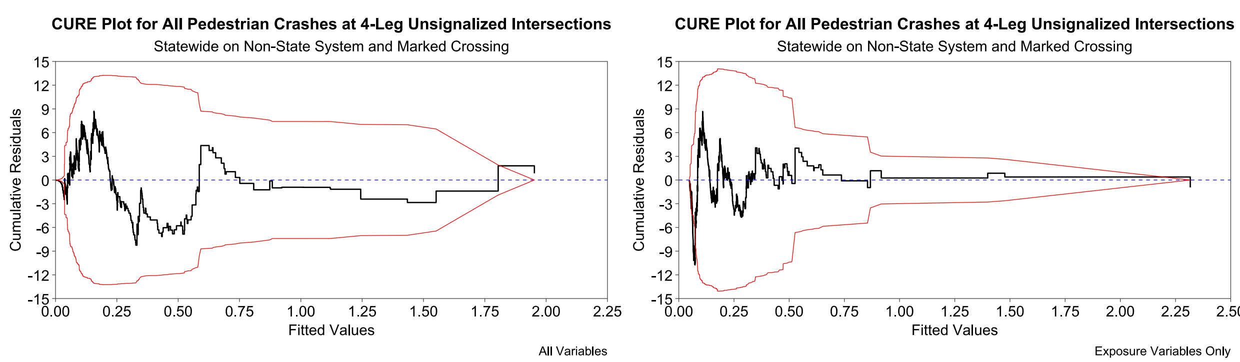


Figure A. 16: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 4-Leg Unsignalized Intersections, Statewide, on Non-State System, and Marked Crossing

Table A.17: Comparison Between Full Model and Exposure-Only Model at 3-Leg Signalized Intersections, Statewide, on State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.426	0.658	55.702	167.959	-115.04
Exposure-Only Model	0.616	0.922	75.456	173.480	-141.97

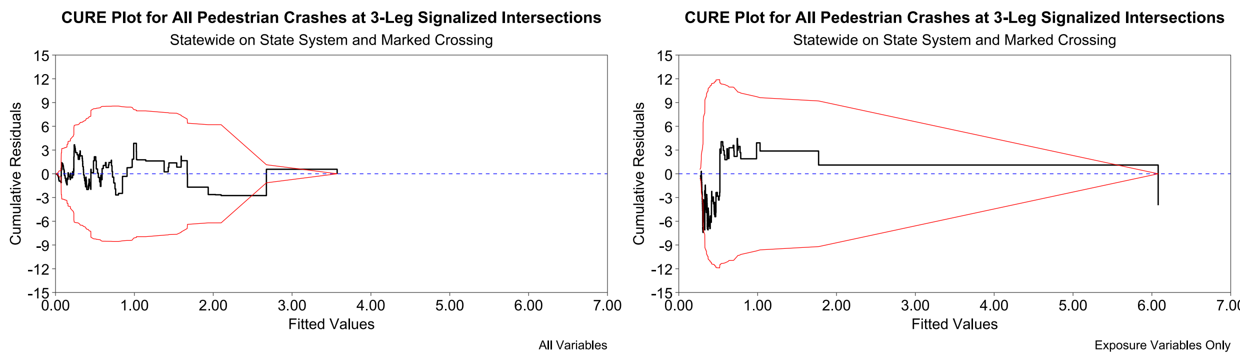


Figure A.17: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 3-Leg Signalized Intersections, Statewide, on State System, and Marked Crossing

Table A.18: Comparison Between Full Model and Exposure-Only Model at 3-Leg Signalized Intersections, Statewide, on Non-State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.438	0.593	63.893	174.426	-285.25
Exposure-Only Model	0.495	0.677	67.708	175.667	-308.83

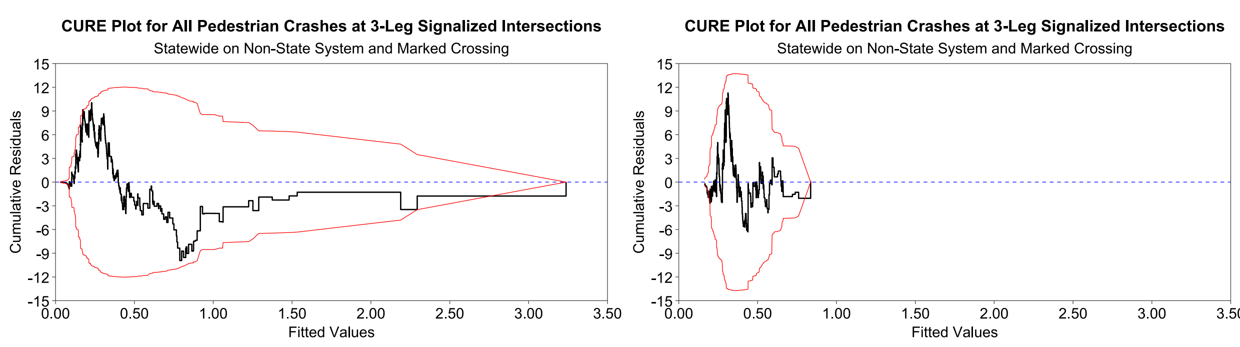


Figure A.18: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 3-Leg Signalized Intersections, Statewide, on Non-State System, and Marked Crossing

Table A.19: Comparison Between Full Model and Exposure-Only Model at 3-Leg Signalized Intersections, Portland Urban Area, on State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.435	0.752	58.047	168.664	-76.43
Exposure-Only Model	0.582	0.883	64.51	173.71	-89.78

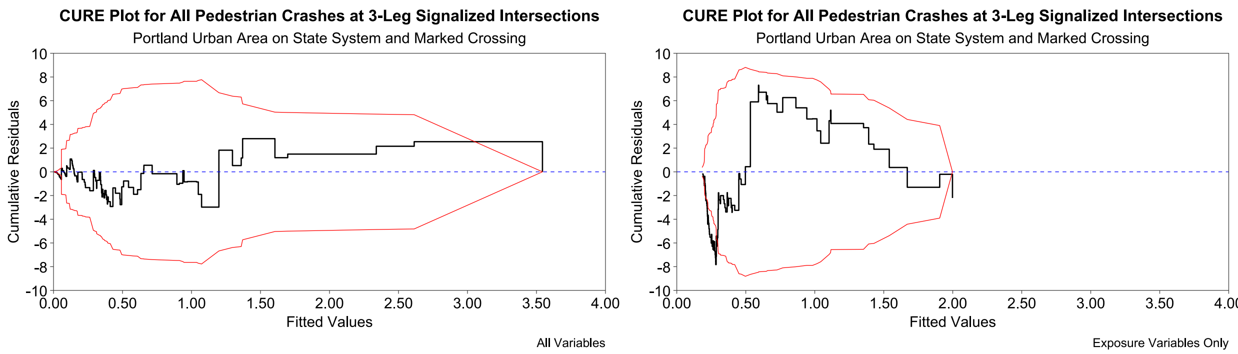


Figure A. 19: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 3-Leg Signalized Intersections, Portland Urban Area, on State System, and Marked Crossing

Table A.20: Comparison Between Full Model and Exposure-Only Model at 3-Leg Signalized Intersections, Portland Urban Area, on Non-State System, and Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.460	0.587	66.748	174.414	-211.49
Exposure-Only Model	0.480	0.605	68.542	175.237	-219.77

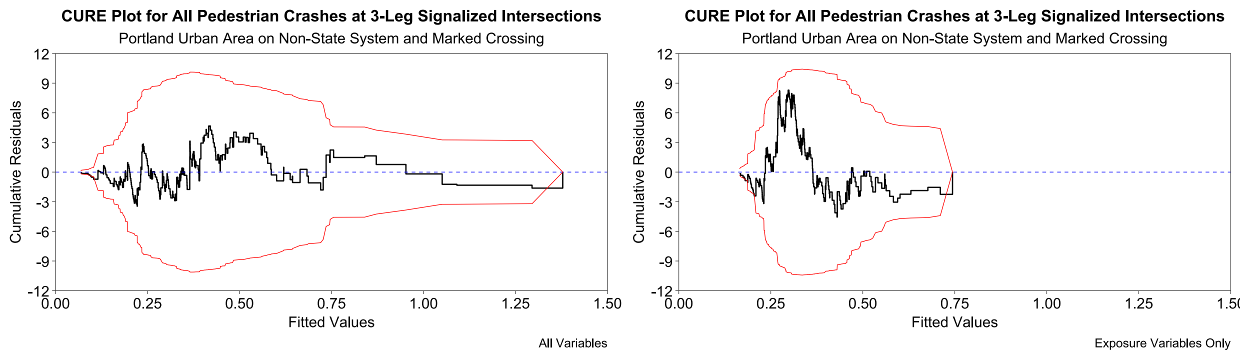


Figure A. 20: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 3-Leg Signalized Intersections, Portland Urban Area, on Non-State System, and Marked Crossing

Table A.21: Comparison Between Full Model and Exposure-Only Model at 3-Leg Signalized Intersections, Statewide, on State System, and No Marked Crossing

Model	MAE	RMSE	MAPE	sMAPE	LL(β)
Full Model	0.174	0.325	65.850	188.883	-47.55
Exposure-Only Model	0.239	0.380	87.111	194.428	-61.18

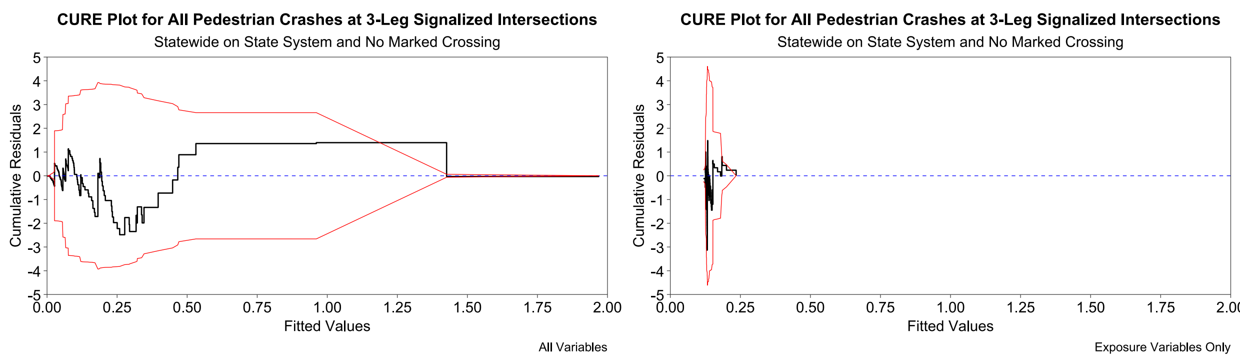


Figure A.21: Comparison of CURE Plots Between Full Model and Exposure-Only Model at 3-Leg Signalized Intersections, Statewide, on State System, and No Marked Crossing

APPENDIX B

Table B.1: 4-Leg Signalized Intersections, Statewide, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.307	0.178	-7.327	≤0.001		
Average Annual Daily Traffic (Thousands)	0.011	0.003	3.908	≤0.001	0.007	1.011
Average Annual Daily Pedestrian Traffic (Thousands)	0.435	0.100	4.337	≤0.001	0.269	1.545
Model Summary						
Number of Observations	464					
Log-likelihood at Zero	-495.511					
Log-likelihood at Convergence	-480.193					
McFadden Pseudo R-squared	0.031					
Dispersion Parameter	0.376					
Standard Error of Dispersion Parameter	0.138					

Table B.2: 4-Leg Signalized Intersections, Portland City Limits, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.115	0.325	-3.429	≤ 0.001		
Average Annual Daily Traffic (Thousands)	0.012	0.005	2.360	0.018	0.009	1.012
Average Annual Daily Pedestrian Traffic (Thousands)	0.285	0.274	1.041	0.298	0.233	1.330
Model Summary						
Number of Observations	109					
Log-likelihood at Zero	-133.264					
Log-likelihood at Convergence	-128.635					
McFadden Pseudo R-squared	0.035					

Table B.3: 4-Leg Signalized Intersections, Portland Urban Area, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.031	0.267	-3.858	≤0.001		
Average Annual Daily Traffic (Thousands)	0.006	0.004	1.481	0.139	0.004	1.006
Average Annual Daily Pedestrian Traffic (Thousands)	0.584	0.192	3.041	0.002	0.419	1.793
Model Summary						
Number of Observations	240					
Log-likelihood at Zero	-276.223					
Log-likelihood at Convergence	-270.120					
McFadden Pseudo R-squared	0.022					
Dispersion Parameter	0.269					
Standard Error of Dispersion Parameter	0.163					

Table B.4: 4-Leg Signalized Intersections, Portland/Corvallis Urban Areas, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.268	0.301	-4.211	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.013	0.005	2.871	0.004	0.010	1.013
Average Annual Daily Pedestrian Traffic (Thousands)	0.287	0.262	1.097	0.273	0.216	1.332
Model Summary						
Number of Observations	129					

Log-likelihood at Zero	-150.878					
Log-likelihood at Convergence	-144.584					
McFadden Pseudo R-squared	0.042					

Table B.5: 4-Leg Signalized Intersections, Statewide, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.797	0.144	-12.443	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.020	0.002	8.816	≤ 0.000	0.012	1.020
Average Annual Daily Pedestrian Traffic (Thousands)	0.451	0.079	5.692	≤ 0.000	0.265	1.570
Model Summary						
Number of Observations	1,235					
Log-likelihood at Zero	-1295.446					
Log-likelihood at Convergence	-1247.082					
McFadden Pseudo R-squared	0.037					
Dispersion Parameter	0.629					
Standard Error of Dispersion Parameter	0.109					

Table B.6: 4-Leg Signalized Intersections, Portland City Limits, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.764	0.274	-6.447	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.023	0.003	7.113	≤ 0.000	0.017	1.023
Average Annual Daily Pedestrian Traffic (Thousands)	0.355	0.142	2.497	0.013	0.259	1.426
Model Summary						
Number of Observations	563					
Log-likelihood at Zero	-662.609					
Log-likelihood at Convergence	-638.058					
McFadden Pseudo R-squared	0.037					
Dispersion Parameter	0.600					
Standard Error of Dispersion Parameter	0.138					

Table B.7: 4-Leg Signalized Intersections, Portland Urban Area, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.798	0.188	-9.569	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.019	0.003	7.465	≤ 0.000	0.013	1.019
Average Annual Daily Pedestrian Traffic (Thousands)	0.482	0.098	4.934	≤ 0.000	0.315	1.619
Model Summary						
Number of Observations	844					
Log-likelihood at Zero	-939.907					
Log-likelihood at Convergence	-905.521					
McFadden Pseudo R-squared	0.037					
Dispersion Parameter	0.642					
Standard Error of Dispersion Parameter	0.124					

Table B.8: 4-Leg Signalized Intersections, Portland/Corvallis Urban Areas, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.772	0.256	-6.921	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.023	0.003	7.339	≤ 0.000	0.017	1.023
Average Annual Daily Pedestrian Traffic (Thousands)	0.364	0.135	2.702	0.007	0.261	1.439
Model Summary						
Number of Observations	587					
Log-likelihood at Zero	-684.239					
Log-likelihood at Convergence	-657.983					
McFadden Pseudo R-squared	0.038					
Dispersion Parameter	0.578					
Standard Error of Dispersion Parameter	0.134					

Table B.9: 4-Leg Signalized Intersections, Statewide, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.859	0.224	-8.316	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.012	0.004	3.132	0.002	0.004	1.012

Average Annual Daily Pedestrian Traffic (Thousands)	0.916	0.279	3.278	≤ 0.001	0.329	2.499
Model Summary						
Number of Observations	383					
Log-likelihood at Zero	-298.815					
Log-likelihood at Convergence	-288.451					
McFadden Pseudo R-squared	0.035					
Dispersion Parameter	0.719					
Standard Error of Dispersion Parameter	0.317					

Table B.10: 4-Leg Signalized Intersections, Portland City Limits, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.021	1.117	-2.705	0.007		
Average Annual Daily Traffic (Thousands)	0.018	0.016	1.163	0.245	0.006	1.018
Average Annual Daily Pedestrian Traffic (Thousands)	0.913	0.532	1.717	0.086	0.322	2.492
Model Summary						
Number of Observations	17					
Log-likelihood at Zero	-13.635					
Log-likelihood at Convergence	-10.484					
McFadden Pseudo R-squared	0.231					

Table B.11: 4-Leg Signalized Intersections, Portland Urban Area, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.828	0.575	-3.181	≤ 0.001		
Average Annual Daily Traffic (Thousands)	0.013	0.007	1.837	0.066	0.007	1.013
Average Annual Daily Pedestrian Traffic (Thousands)	0.721	0.445	1.62	0.105	0.392	2.056
Model Summary						
Number of Observations	78					
Log-likelihood at Zero	-77.5					
Log-likelihood at Convergence	-74.786					
McFadden Pseudo R-squared	0.035					
Dispersion Parameter	0.993					

Standard Error of Dispersion Parameter	0.633					

Table B.12: 4-Leg Signalized Intersections, Portland/Corvallis Urban Areas, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.744	0.870	-3.154	0.002		
Average Annual Daily Traffic (Thousands)	0.018	0.014	1.319	0.187	0.006	1.018
Average Annual Daily Pedestrian Traffic (Thousands)	0.768	0.476	1.612	0.107	0.236	2.155
Model Summary						
Number of Observations	26					
Log-likelihood at Zero	-18.816					
Log-likelihood at Convergence	-15.408					
McFadden Pseudo R-squared	0.181					

Table B.13: 4-Leg Signalized Intersections, Statewide, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.224	0.244	-9.127	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.023	0.004	5.695	≤ 0.000	0.010	1.023
Average Annual Daily Pedestrian Traffic (Thousands)	0.599	0.235	2.544	0.011	0.254	1.820
Model Summary						
Number of Observations	358					
Log-likelihood at Zero	-307.272					
Log-likelihood at Convergence	-287.826					
McFadden Pseudo R-squared	0.063					
Dispersion Parameter	0.890					
Standard Error of Dispersion Parameter	0.303					

Table B.14: 4-Leg Signalized Intersections, Portland City Limits, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.585	0.825	-4.348	≤ 0.000		

Average Annual Daily Traffic (Thousands)	0.045	0.013	3.470	≤ 0.001	0.019	1.046
Average Annual Daily Pedestrian Traffic (Thousands)	0.755	0.427	1.769	0.077	0.324	2.128
Model Summary						
Number of Observations	56					
Log-likelihood at Zero	-53.419					
Log-likelihood at Convergence	-44.917					
McFadden Pseudo R-squared	0.159					

Table B.15: 4-Leg Signalized Intersections, Portland Urban Area, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.789	0.496	-5.626	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.026	0.006	4.342	≤ 0.000	0.014	1.026
Average Annual Daily Pedestrian Traffic (Thousands)	0.907	0.362	2.502	0.012	0.499	2.477
Model Summary						
Number of Observations	129					
Log-likelihood at Zero	-124.773					
Log-likelihood at Convergence	-114.167					
McFadden Pseudo R-squared	0.085					
Dispersion Parameter	1.206					
Standard Error of Dispersion Parameter	0.503					

Table B.16: 4-Leg Signalized Intersections, Portland/Corvallis Urban Areas, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.024	0.695	-4.350	≤ 0.001		
Average Annual Daily Traffic (Thousands)	0.041	0.012	3.455	≤ 0.001	0.018	1.042
Average Annual Daily Pedestrian Traffic (Thousands)	0.510	0.377	1.351	0.177	0.271	1.665
Model Summary	Value					
Number of Observations	64					
Log-likelihood at Zero	-61.064					
Log-likelihood at Convergence	-53.374					
McFadden Pseudo R-squared	0.126					

Table B.17: 4-Leg Unsignalized Intersections, Statewide, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.139	0.340	-6.300	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.021	0.009	2.272	0.023	0.005	1.021
Average Annual Daily Pedestrian Traffic (Thousands)	0.236	0.204	1.155	0.248	0.055	1.266
Model Summary						
Number of Observations	160					
Log-likelihood at Zero	-95.796					
Log-likelihood at Convergence	-92.696					
McFadden Pseudo R-squared	0.032					

Table B.18: 4-Leg Unsignalized Intersections, Portland City Limits, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-4.599	2.278	-2.019	0.043		
Average Annual Daily Traffic (Thousands)	0.039	0.020	1.931	0.053	0.017	1.040
Average Annual Daily Pedestrian Traffic (Thousands)	3.027	2.217	1.365	0.172	1.324	20.635
Model Summary	Value					
Number of Observations	16					
Log-likelihood at Zero	-14.597					
Log-likelihood at Convergence	-11.43					
McFadden Pseudo R-squared	0.217					

Table B.19: 4-Leg Unsignalized Intersections, Portland/Corvallis Urban Areas, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	0.298	2.567	0.116	0.908		
Average Annual Daily Traffic (Thousands)	-0.118	0.137	-0.861	0.389	-0.044	0.889
Average Annual Daily Pedestrian Traffic (Thousands)	3.566	3.943	0.905	0.366	1.337	35.375
Model Summary						
Number of Observations	8					

Log-likelihood at Zero	-5.942					
Log-likelihood at Convergence	-4.711					
McFadden Pseudo R-squared	0.207					

Table B.20: 4-Leg Unsignalized Intersections, Statewide, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.159	0.189	-16.710	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.033	0.006	5.119	≤ 0.000	0.005	1.034
Average Annual Daily Pedestrian Traffic (Thousands)	0.806	0.171	4.715	≤ 0.000	0.113	2.239
Model Summary						
Number of Observations	1,086					
Log-likelihood at Zero	-459.099					
Log-likelihood at Convergence	-426.922					
McFadden Pseudo R-squared	0.070					
Dispersion Parameter	1.220					
Standard Error of Dispersion Parameter	0.454					

Table B.21: 4-Leg Unsignalized Intersections, Portland City Limits, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.330	0.439	-7.585	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.034	0.008	4.467	≤ 0.000	0.008	1.035
Average Annual Daily Pedestrian Traffic (Thousands)	0.922	0.295	3.122	0.002	0.227	2.514
Model Summary						
Number of Observations	340					
Log-likelihood at Zero	-209.156					
Log-likelihood at Convergence	-196.002					
McFadden Pseudo R-squared	0.063					
Dispersion Parameter	0.798					
Standard Error of Dispersion Parameter	0.434					

Table B.22: 4-Leg Unsignalized Intersections, Portland Urban Area, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.862	0.270	-10.599	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.029	0.007	4.003	≤ 0.000	0.006	1.029
Average Annual Daily Pedestrian Traffic (Thousands)	0.674	0.212	3.179	0.001	0.135	1.962
Model Summary						
Number of Observations	565					
Log-likelihood at Zero	-304.625					
Log-likelihood at Convergence	-289.248					
McFadden Pseudo R-squared	0.05					
Dispersion Parameter	1.182					
Standard Error of Dispersion Parameter	0.48					

Table B.23: 4-Leg Unsignalized Intersections, Portland/Corvallis Urban Areas, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.169	0.403	-7.871	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.033	0.007	4.417	≤ 0.000	0.008	1.034
Average Annual Daily Pedestrian Traffic (Thousands)	0.825	0.281	2.939	0.003	0.198	2.282
Model Summary						
Number of Observations	369					
Log-likelihood at Zero	-223.952					
Log-likelihood at Convergence	-210.515					
McFadden Pseudo R-squared	0.06					
Dispersion Parameter	0.835					
Standard Error of Dispersion Parameter	0.438					

Table B.24: 4-Leg Unsignalized Intersections, Statewide, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.434	0.198	-17.311	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.006	0.004	1.407	0.159	0.000	1.006
Average Annual Daily Pedestrian Traffic (Thousands)	0.725	0.127	5.701	≤ 0.000	0.034	2.065

Model Summary						
Number of Observations	1,208					
Log-likelihood at Zero	-235.643					
Log-likelihood at Convergence	-225.389					
McFadden Pseudo R-squared	0.044					

Table B.25: 4-Leg Unsignalized Intersections, Portland City Limits, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-4.224	1.102	-3.832	≤ 0.000		
Average Annual Daily Traffic (Thousands)	-0.010	0.014	-0.707	0.480	-0.001	0.990
Average Annual Daily Pedestrian Traffic (Thousands)	2.576	0.821	3.139	0.002	0.183	13.144
Model Summary						
Number of Observations	127					
Log-likelihood at Zero	-32.823					
Log-likelihood at Convergence	-27.851					
McFadden Pseudo R-squared	0.151					

Table B.26: 4-Leg Unsignalized Intersections, Portland Urban Area, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.079	0.571	-5.393	≤ 0.000		
Average Annual Daily Traffic (Thousands)	-0.005	0.008	-0.651	0.515	0.000	0.995
Average Annual Daily Pedestrian Traffic (Thousands)	1.569	0.553	2.835	0.005	0.140	4.802
Model Summary						
Number of Observations	201					
Log-likelihood at Zero	-62.819					
Log-likelihood at Convergence	-59.121					
McFadden Pseudo R-squared	0.059					

Table B.27: 4-Leg Unsignalized Intersections, Portland/Corvallis Urban Areas, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-4.080	0.923	-4.421	≤ 0.000		

Average Annual Daily Traffic (Thousands)	-0.010	0.013	-0.752	0.452	-0.001	0.990
Average Annual Daily Pedestrian Traffic (Thousands)	2.473	0.733	3.372	≤ 0.001	0.163	11.858
Model Summary						
Number of Observations	152					
Log-likelihood at Zero	-37.213					
Log-likelihood at Convergence	-31.933					
McFadden Pseudo R-squared	0.142					

Table B.28: 4-Leg Unsignalized Intersections, Statewide, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-4.315	0.107	-40.453	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.036	0.004	9.805	≤ 0.000	0.001	1.037
Average Annual Daily Pedestrian Traffic (Thousands)	1.055	0.116	9.059	≤ 0.000	0.042	2.872
Model Summary						
Number of Observations	7,897					
Log-likelihood at Zero	-1336.053					
Log-likelihood at Convergence	-1223.183					
McFadden Pseudo R-squared	0.084					
Dispersion Parameter	0.729					
Standard Error of Dispersion Parameter	0.408					

Table B.29: 4-Leg Unsignalized Intersections, Portland City Limits, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-5.108	0.341	-14.998	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.036	0.006	6.502	≤ 0.000	0.003	1.037
Average Annual Daily Pedestrian Traffic (Thousands)	1.576	0.239	6.606	≤ 0.000	0.110	4.836
Model Summary						
Number of Observations	1,634					
Log-likelihood at Zero	-431.156					
Log-likelihood at Convergence	-382.481					
McFadden Pseudo R-squared	0.113					

Table B.30: 4-Leg Unsignalized Intersections, Portland Urban Area, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-4.205	0.177	-23.763	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.029	0.005	6.341	≤ 0.000	0.002	1.029
Average Annual Daily Pedestrian Traffic (Thousands)	1.079	0.150	7.181	≤ 0.000	0.062	2.942
Model Summary						
Number of Observations	3,065					
Log-likelihood at Zero	-692.371					
Log-likelihood at Convergence	-634.392					
McFadden Pseudo R-squared	0.084					

APPENDIX C

Table C.1: 3-Leg Signalized Intersections, Statewide, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.460	0.442	-3.303	≤ 0.001		
Average Annual Daily Traffic (Thousands)	0.005	0.008	0.627	0.531	0.002	1.005
Average Annual Daily Pedestrian Traffic (Thousands)	0.892	0.303	2.943	0.003	0.396	2.440
Model Summary						
Number of Observations	169					
Log-likelihood at Zero	-144.989					
Log-likelihood at Convergence	-141.968					
McFadden Pseudo R-squared	0.021					
Dispersion Parameter	1.464					
Standard Error of Dispersion Parameter	0.567					

Table C.2: 3-Leg Signalized Intersections, Portland City Limits, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-0.105	1.114	-0.094	0.925		
Average Annual Daily Traffic (Thousands)	-0.019	0.019	-0.994	0.320	-0.008	0.981
Average Annual Daily Pedestrian Traffic (Thousands)	0.252	0.729	0.345	0.730	0.105	1.287
Model Summary						
Number of Observations	36					
Log-likelihood at Zero	-35.404					
Log-likelihood at Convergence	-34.677					
McFadden Pseudo R-squared	0.021					

Table C.3: 3-Leg Signalized Intersections, Portland Urban Area, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.787	0.665	-2.689	0.007		
Average Annual Daily Traffic (Thousands)	0.001	0.012	0.091	0.928	0.000	1.001
Average Annual Daily Pedestrian Traffic (Thousands)	1.919	0.532	3.607	≤ 0.000	0.851	6.814
Model Summary						
Number of Observations	111					
Log-likelihood at Zero	-95.010					
Log-likelihood at Convergence	-89.777					
McFadden Pseudo R-squared	0.055					
Dispersion Parameter	1.430					
Standard Error of Dispersion Parameter	0.682					

Table C.4: 3-Leg Signalized Intersections, Portland/Corvallis Urban Areas, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-0.386	1.05	-0.367	0.713		
Average Annual Daily Traffic (Thousands)	-0.016	0.019	-0.858	0.391	-0.006	0.984
Average Annual Daily Pedestrian Traffic (Thousands)	0.411	0.701	0.586	0.558	0.166	1.508
Model Summary						
Number of Observations	37					
Log-likelihood at Zero	-35.815					
Log-likelihood at Convergence	-35.141					
McFadden Pseudo R-squared	0.019					

Table C.5: 3-Leg Signalized Intersections, Statewide, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.031	0.301	-6.748	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.019	0.006	3.306	≤ 0.001	0.007	1.019
Average Annual Daily Pedestrian Traffic (Thousands)	0.319	0.207	1.543	0.123	0.110	1.376
Model Summary						
Number of Observations	411					
Log-likelihood at Zero	-320.999					
Log-likelihood at Convergence	-308.829					
McFadden Pseudo R-squared	0.038					
Dispersion Parameter	0.801					
Standard Error of Dispersion Parameter	0.320					

Table C.6: 3-Leg Signalized Intersections, Portland City Limits, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.283	0.571	-4.000	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.018	0.009	1.942	0.052	0.006	1.018
Average Annual Daily Pedestrian Traffic (Thousands)	0.491	0.317	1.550	0.121	0.162	1.634
Model Summary						
Number of Observations	160					
Log-likelihood at Zero	-119.941					
Log-likelihood at Convergence	-116.900					
McFadden Pseudo R-squared	0.025					
Dispersion Parameter	0.480					
Standard Error of Dispersion Parameter	0.473					

Table C.7: 3-Leg Signalized Intersections, Portland Urban Area, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.994	0.352	-5.662	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.017	0.006	2.560	0.010	0.006	1.017
Average Annual Daily Pedestrian Traffic (Thousands)	0.350	0.218	1.601	0.109	0.118	1.419
Model Summary						
Number of Observations	297					
Log-likelihood at Zero	-228.993					
Log-likelihood at Convergence	-219.221					
McFadden Pseudo R-squared	0.043					
Dispersion Parameter	0.290					
Standard Error of Dispersion Parameter	0.313					

Table C.8: 3-Leg Signalized Intersections, Portland/Corvallis Urban Areas, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.208	0.496	-4.454	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.019	0.008	2.261	0.024	0.006	1.019
Average Annual Daily Pedestrian Traffic (Thousands)	0.417	0.276	1.509	0.131	0.140	1.517
Model Summary						
Number of Observations	167					
Log-likelihood at Zero	-126.983					
Log-likelihood at Convergence	-123.284					
McFadden Pseudo R-squared	0.029					

Table C.9: 3-Leg Signalized Intersections, Statewide, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.037	0.567	-3.595	≤ 0.000		
Average Annual Daily Traffic (Thousands)	-0.001	0.012	-0.079	0.937	0.000	0.999
Average Annual Daily Pedestrian Traffic (Thousands)	0.352	0.744	0.473	0.636	0.048	1.422
Model Summary						
Number of Observations	147					
Log-likelihood at Zero	-61.684					
Log-likelihood at Convergence	-61.178					
McFadden Pseudo R-squared	0.008					

Table C.10: 3-Leg Signalized Intersections, Portland City Limits, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.060	2.237	-1.368	0.171		
Average Annual Daily Traffic (Thousands)	0.007	0.031	0.238	0.812	0.001	1.007
Average Annual Daily Pedestrian Traffic (Thousands)	1.188	1.231	0.965	0.335	0.170	3.281
Model Summary						
Number of Observations	21					
Log-likelihood at Zero	-9.531					
Log-likelihood at Convergence	-9.092					
McFadden Pseudo R-squared	0.046					

Table C.11: 3-Leg Signalized Intersections, Portland Urban Area, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.559	1.051	-1.483	0.138		
Average Annual Daily Traffic (Thousands)	-0.010	0.019	-0.515	0.606	-0.001	0.990
Average Annual Daily Pedestrian Traffic (Thousands)	0.321	0.928	0.346	0.729	0.048	1.379
Model Summary						
Number of Observations	47					
Log-likelihood at Zero	-21.170					
Log-likelihood at Convergence	-20.818					
McFadden Pseudo R-squared	0.017					

Table C.12: 3-Leg Signalized Intersections, Portland/Corvallis Urban Areas, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.209	2.201	-1.458	0.145		
Average Annual Daily Traffic (Thousands)	0.009	0.031	0.283	0.777	0.001	1.009
Average Annual Daily Pedestrian Traffic (Thousands)	1.264	1.206	1.048	0.295	0.172	3.540
Model Summary						
Number of Observations	22					
Log-likelihood at Zero	-9.670					
Log-likelihood at Convergence	-9.163					
McFadden Pseudo R-squared	0.052					

Table C.13: 3-Leg Signalized Intersections, Statewide, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.901	0.509	-5.700	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.029	0.010	2.872	0.004	0.005	1.029
Average Annual Daily Pedestrian Traffic (Thousands)	0.229	0.468	0.490	0.624	0.041	1.257
Model Summary						
Number of Observations	197					
Log-likelihood at Zero	-109.787					
Log-likelihood at Convergence	-95.061					
McFadden Pseudo R-squared	0.134					

Table C.14: 3-Leg Signalized Intersections, Portland City Limits, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.989	1.717	-1.740	0.082		
Average Annual Daily Traffic (Thousands)	0.013	0.039	0.342	0.733	0.001	1.013
Average Annual Daily Pedestrian Traffic (Thousands)	0.400	0.933	0.428	0.668	0.041	1.492
Model Summary						
Number of Observations	39					
Log-likelihood at Zero	-13.903					
Log-likelihood at Convergence	-13.693					
McFadden Pseudo R-squared	0.015					

Table C.15: 3-Leg Signalized Intersections, Portland Urban Area, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.151	0.756	-2.843	0.004		
Average Annual Daily Traffic (Thousands)	0.010	0.015	0.656	0.512	0.002	1.010
Average Annual Daily Pedestrian Traffic (Thousands)	0.026	0.588	0.045	0.964	0.004	1.026
Model Summary						
Number of Observations	95					
Log-likelihood at Zero	-56.552					
Log-likelihood at Convergence	-45.666					
McFadden Pseudo R-squared	0.192					

Table C.16: 3-Leg Signalized Intersections, Portland/Corvallis Urban Areas, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.231	1.727	-1.871	0.061		
Average Annual Daily Traffic (Thousands)	0.016	0.040	0.413	0.679	0.002	1.016
Average Annual Daily Pedestrian Traffic (Thousands)	0.509	0.934	0.545	0.585	0.048	1.664
Model Summary						
Number of Observations	42					
Log-likelihood at Zero	-14.193					
Log-likelihood at Convergence	-13.927					
McFadden Pseudo R-squared	0.019					

Table C.17: 3-Leg Unsignalized Intersections, Statewide, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.172	0.487	-6.509	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.015	0.011	1.335	0.182	0.002	1.015
Average Annual Daily Pedestrian Traffic (Thousands)	1.135	0.841	1.349	0.177	0.120	3.111
Model Summary						
Number of Observations	151					
Log-likelihood at Zero	-54.413					
Log-likelihood at Convergence	-50.994					
McFadden Pseudo R-squared	0.063					

Table C.18: 3-Leg Unsignalized Intersections, Portland City Limits, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-1.871	1.351	-1.384	0.166		
Average Annual Daily Traffic (Thousands)	0.010	0.022	0.454	0.650	0.001	1.010
Average Annual Daily Pedestrian Traffic (Thousands)	-0.776	1.698	-0.457	0.648	-0.111	0.460
Model Summary						
Number of Observations	28					
Log-likelihood at Zero	-13.170					
Log-likelihood at Convergence	-12.993					
McFadden Pseudo R-squared	0.013					

Table C.19: 3-Leg Unsignalized Intersections, Portland Urban Area, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.408	0.942	-2.556	0.011		
Average Annual Daily Traffic (Thousands)	0.013	0.015	0.865	0.387	0.002	1.013
Average Annual Daily Pedestrian Traffic (Thousands)	0.331	1.145	0.289	0.772	0.063	1.392
Model Summary						
Number of Observations	42					
Log-likelihood at Zero	-23.534					
Log-likelihood at Convergence	-22.878					
McFadden Pseudo R-squared	0.028					

Table C.20: 3-Leg Unsignalized Intersections, Portland/Corvallis Urban Areas, State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.085	1.296	-1.609	0.108		
Average Annual Daily Traffic (Thousands)	0.010	0.021	0.475	0.635	0.001	1.010
Average Annual Daily Pedestrian Traffic (Thousands)	-0.528	1.633	-0.323	0.747	-0.073	0.590
Model Summary						
Number of Observations	29					
Log-likelihood at Zero	-13.310					
Log-likelihood at Convergence	-13.172					
McFadden Pseudo R-squared	0.010					

Table C.21: 3-Leg Unsignalized Intersections, Statewide, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.367	0.200	-16.836	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.027	0.007	3.947	≤ 0.000	0.002	1.027
Average Annual Daily Pedestrian Traffic (Thousands)	0.759	0.193	3.926	≤ 0.000	0.063	2.136
Model Summary						
Number of Observations	1,244					
Log-likelihood at Zero	-360.255					
Log-likelihood at Convergence	-344.611					
McFadden Pseudo R-squared	0.043					
Dispersion Parameter	1.848					
Standard Error of Dispersion Parameter	0.813					

Table C.22: 3-Leg Unsignalized Intersections, Portland City Limits, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.780	0.607	-6.224	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.017	0.011	1.525	0.127	0.002	1.017
Average Annual Daily Pedestrian Traffic (Thousands)	1.161	0.456	2.548	0.011	0.119	3.193
Model Summary						
Number of Observations	254					
Log-likelihood at Zero	-86.952					
Log-likelihood at Convergence	-82.276					
McFadden Pseudo R-squared	0.054					

Table C.23: 3-Leg Unsignalized Intersections, Portland Urban Area, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.248	0.283	-11.489	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.019	0.008	2.505	0.012	0.002	1.019
Average Annual Daily Pedestrian Traffic (Thousands)	0.793	0.284	2.798	0.005	0.076	2.210
Model Summary						
Number of Observations	619					
Log-likelihood at Zero	-202.124					
Log-likelihood at Convergence	-193.773					
McFadden Pseudo R-squared	0.041					

Table C.24: 3-Leg Unsignalized Intersections, Portland/Corvallis Urban Areas, Non-State System, and Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.816	0.574	-6.652	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.017	0.011	1.525	0.127	0.002	1.017
Average Annual Daily Pedestrian Traffic (Thousands)	1.240	0.437	2.839	0.005	0.126	3.456
Model Summary						
Number of Observations	275					
Log-likelihood at Zero	-94.351					
Log-likelihood at Convergence	-88.810					
McFadden Pseudo R-squared	0.059					

Table C.25: 3-Leg Unsignalized Intersections, Statewide, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.964	0.131	-30.315	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.008	0.002	3.895	≤ 0.000	0.000	1.008
Average Annual Daily Pedestrian Traffic (Thousands)	0.804	0.126	6.399	≤ 0.000	0.025	2.234
Model Summary						
Number of Observations	3,993					
Log-likelihood at Zero	-579.509					
Log-likelihood at Convergence	-546.179					
McFadden Pseudo R-squared	0.058					

Table C.26: 3-Leg Unsignalized Intersections, Portland City Limits, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.448	0.356	-6.875	≤ 0.000		
Average Annual Daily Traffic (Thousands)	-0.005	0.005	-1.026	0.305	0.000	0.995
Average Annual Daily Pedestrian Traffic (Thousands)	0.238	0.460	0.517	0.605	0.018	1.269
Model Summary						
Number of Observations	561					
Log-likelihood at Zero	-156.296					
Log-likelihood at Convergence	-155.527					
McFadden Pseudo R-squared	0.005					

Table C.27: 3-Leg Unsignalized Intersections, Portland Urban Area, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-3.496	0.296	-11.815	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.000	0.004	0.110	0.912	0.000	1.000
Average Annual Daily Pedestrian Traffic (Thousands)	1.393	0.390	3.575	≤ 0.000	0.081	4.027
Model Summary						
Number of Observations	1,144					
Log-likelihood at Zero	-250.633					
Log-likelihood at Convergence	-244.832					
McFadden Pseudo R-squared	0.023					
Dispersion Parameter	4.229					
Standard Error of Dispersion Parameter	1.876					

Table C.28: 3-Leg Unsignalized Intersections, Portland/Corvallis Urban Areas, State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-2.860	0.331	-8.635	≤ 0.000		
Average Annual Daily Traffic (Thousands)	-0.004	0.005	-0.740	0.459	0.000	0.996
Average Annual Daily Pedestrian Traffic (Thousands)	0.640	0.423	1.514	0.130	0.043	1.896
Model Summary						
Number of Observations	641					
Log-likelihood at Zero	-162.485					
Log-likelihood at Convergence	-160.728					
McFadden Pseudo R-squared	0.011					

Table C.29: 3-Leg Unsignalized Intersections, Statewide, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-5.303	0.097	-54.467	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.041	0.003	14.068	≤ 0.000	0.001	1.042
Average Annual Daily Pedestrian Traffic (Thousands)	1.351	0.143	9.435	≤ 0.000	0.023	3.861
Model Summary						
Number of Observations	21,134					
Log-likelihood at Zero	-1773.422					
Log-likelihood at Convergence	-1593.586					
McFadden Pseudo R-squared	0.101					
Dispersion Parameter	3.432					
Standard Error of Dispersion Parameter	0.849					

Table C.30: 3-Leg Unsignalized Intersections, Portland City Limits, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-5.592	0.281	-19.888	≤ 0.001		
Average Annual Daily Traffic (Thousands)	0.031	0.005	6.105	≤ 0.001	0.001	1.031
Average Annual Daily Pedestrian Traffic (Thousands)	1.760	0.242	7.271	≤ 0.001	0.058	5.812
Model Summary						
Number of Observations	2,918					
Log-likelihood at Zero	-434.605					
Log-likelihood at Convergence	-381.498					
McFadden Pseudo R-squared	0.122					

Table C.31: 3-Leg Unsignalized Intersections, Portland Urban Area, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-5.398	0.162	-33.411	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.037	0.004	9.977	≤ 0.000	0.001	1.038
Average Annual Daily Pedestrian Traffic (Thousands)	1.512	0.180	8.394	≤ 0.000	0.036	4.536
Model Summary						
Number of Observations	8,341					
Log-likelihood at Zero	-911.325					
Log-likelihood at Convergence	-810.905					
McFadden Pseudo R-squared	0.110					
Dispersion Parameter	3.023					
Standard Error of Dispersion Parameter	0.951					

Table C.32: 3-Leg Unsignalized Intersections, Portland/Corvallis Urban Areas, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-5.687	0.272	-20.886	≤ 0.001		
Average Annual Daily Traffic (Thousands)	0.032	0.005	6.346	≤ 0.001	0.001	1.033
Average Annual Daily Pedestrian Traffic (Thousands)	1.807	0.238	7.588	≤ 0.001	0.057	6.092
Model Summary						
Number of Observations	3,245					
Log-likelihood at Zero	-448.482					
Log-likelihood at Convergence	-388.603					
McFadden Pseudo R-squared	0.134					

Table B.31: 4-Leg Unsignalized Intersections, Portland/Corvallis Urban Areas, Non-State System, and No Marked Crossing Model Specifications

Variable	Coefficient	Std. Error	*z*-statistic	*p*-value	Marginal Effects	IRR
Constant	-4.643	0.292	-15.892	≤ 0.000		
Average Annual Daily Traffic (Thousands)	0.033	0.005	6.012	≤ 0.000	0.002	1.034
Average Annual Daily Pedestrian Traffic (Thousands)	1.302	0.217	5.994	≤ 0.000	0.091	3.677
Model Summary						
Number of Observations	1,788					
Log-likelihood at Zero	-472.021					
Log-likelihood at Convergence	-428.465					
McFadden Pseudo R-squared	0.092					

APPENDIX D

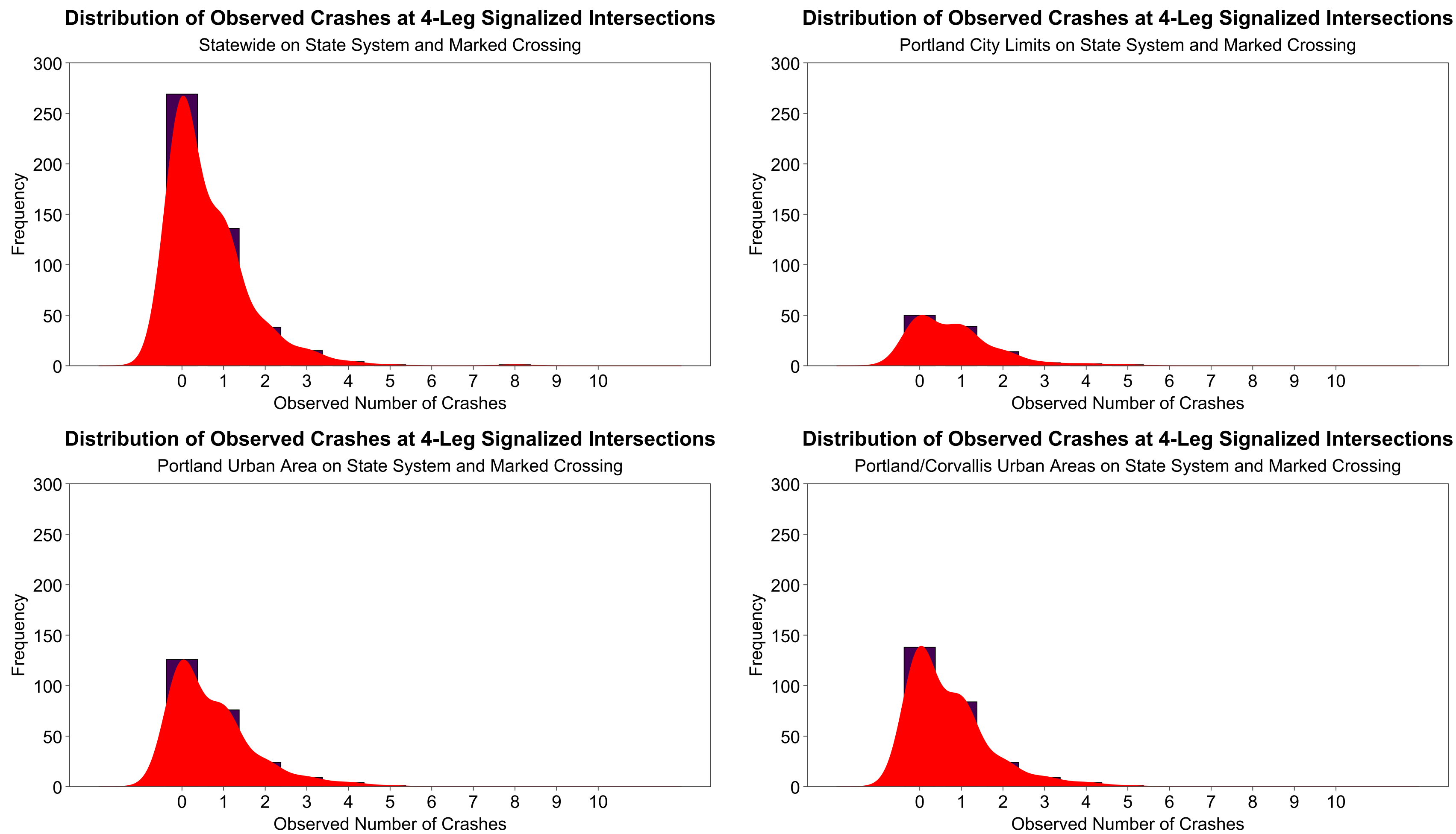


Figure D.1: Distribution of Observed Crashes at 4-Leg Signalized Intersections on State System with Marked Crossing

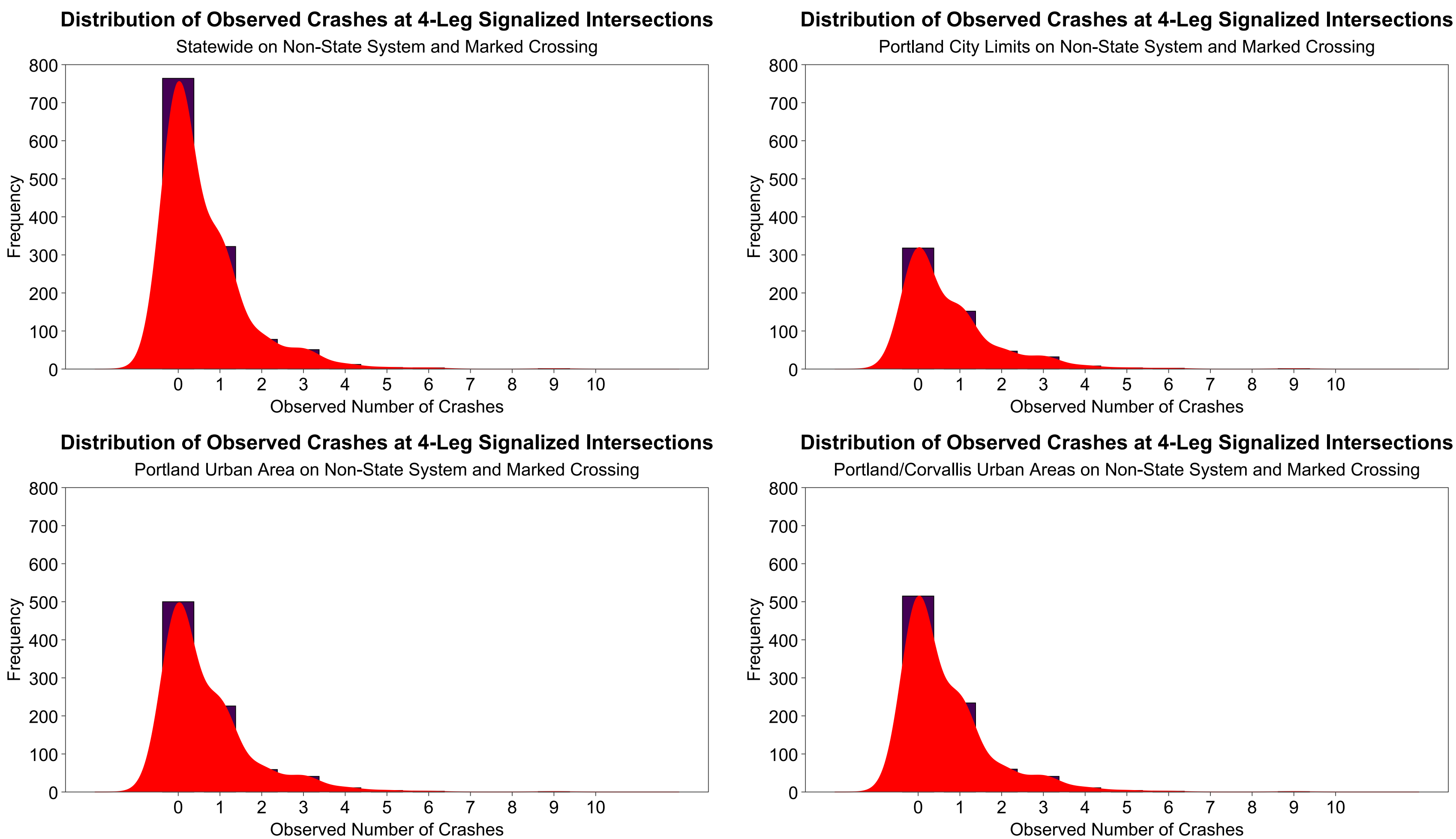


Figure D.2: Distribution of Observed Crashes at 4-Leg Signalized Intersections on Non-State System with Marked Crossing

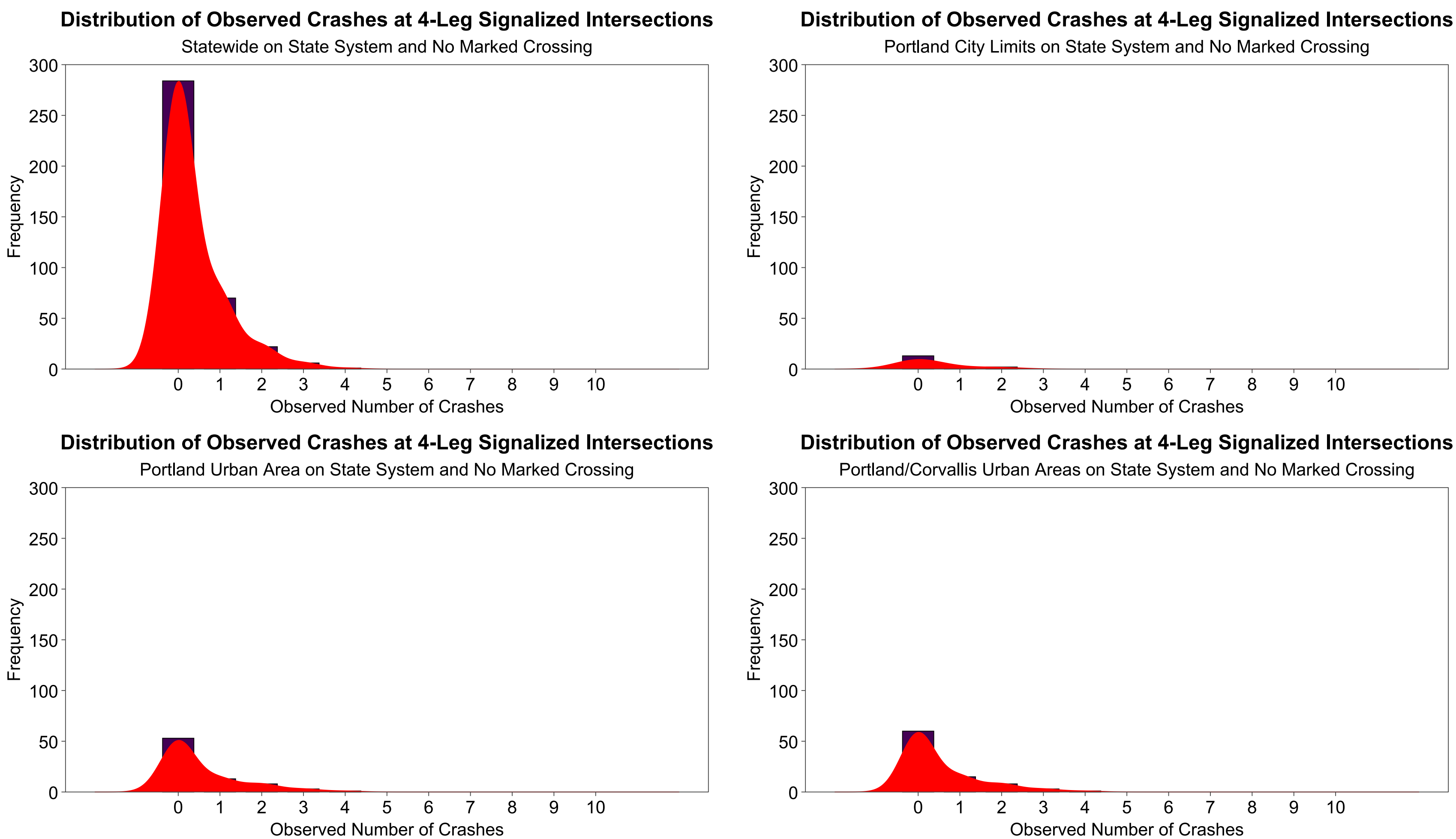


Figure D.3: Distribution of Observed Crashes at 4-Leg Signalized Intersections on State System with No Marked Crossing

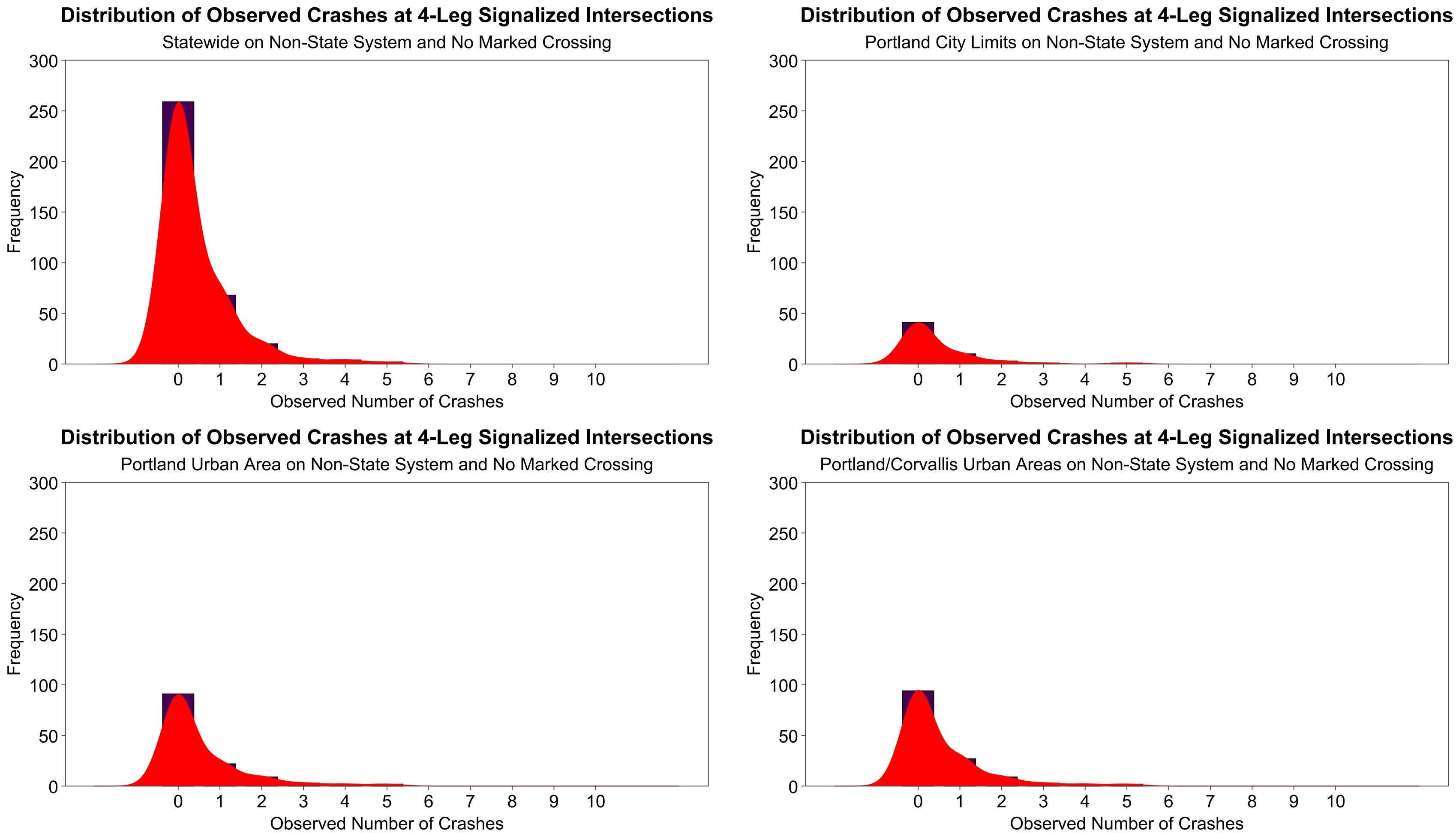


Figure D.4: Distribution of Observed Crashes at 4-Leg Signalized Intersections on Non-State System with No Marked Crossing

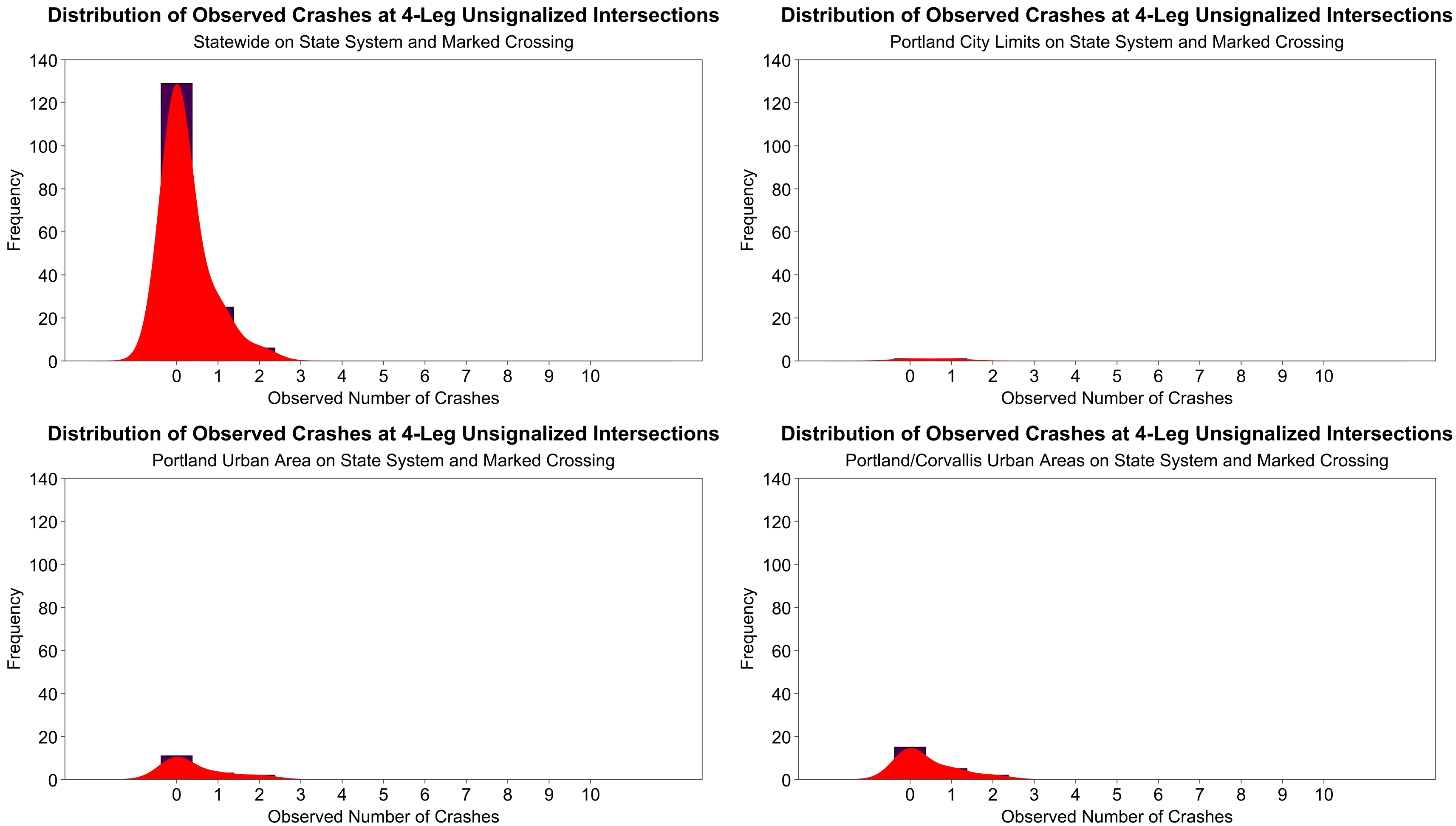


Figure D.5: Distribution of Observed Crashes at 4-Leg Unsignalized Intersections on State System with Marked Crossing

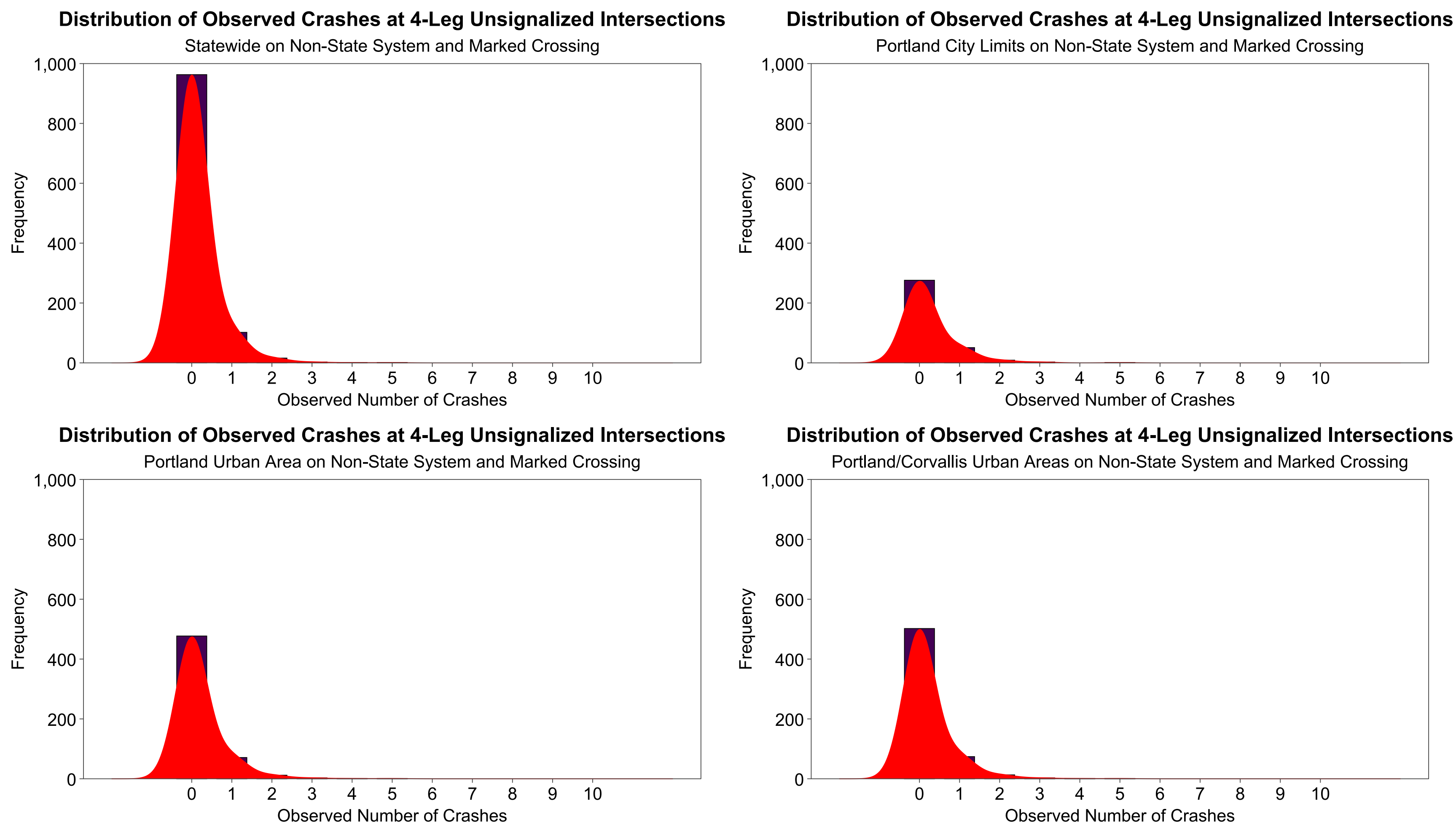


Figure D.6: Distribution of Observed Crashes at 4-Leg Unsignalized Intersections on Non-State System with Marked Crossing

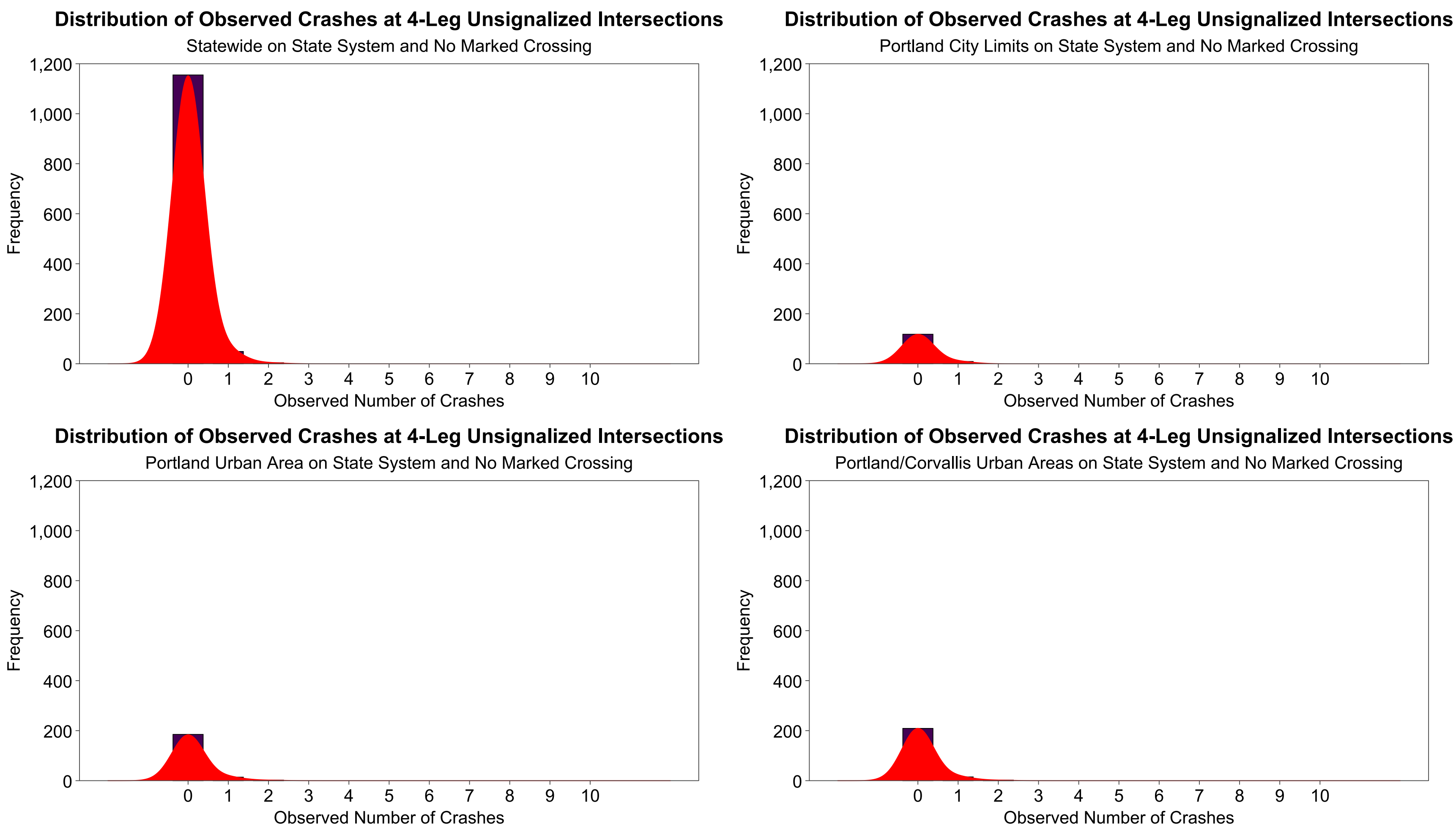


Figure D.7: Distribution of Observed Crashes at 4-Leg Unsignalized Intersections on State System with No Marked Crossing

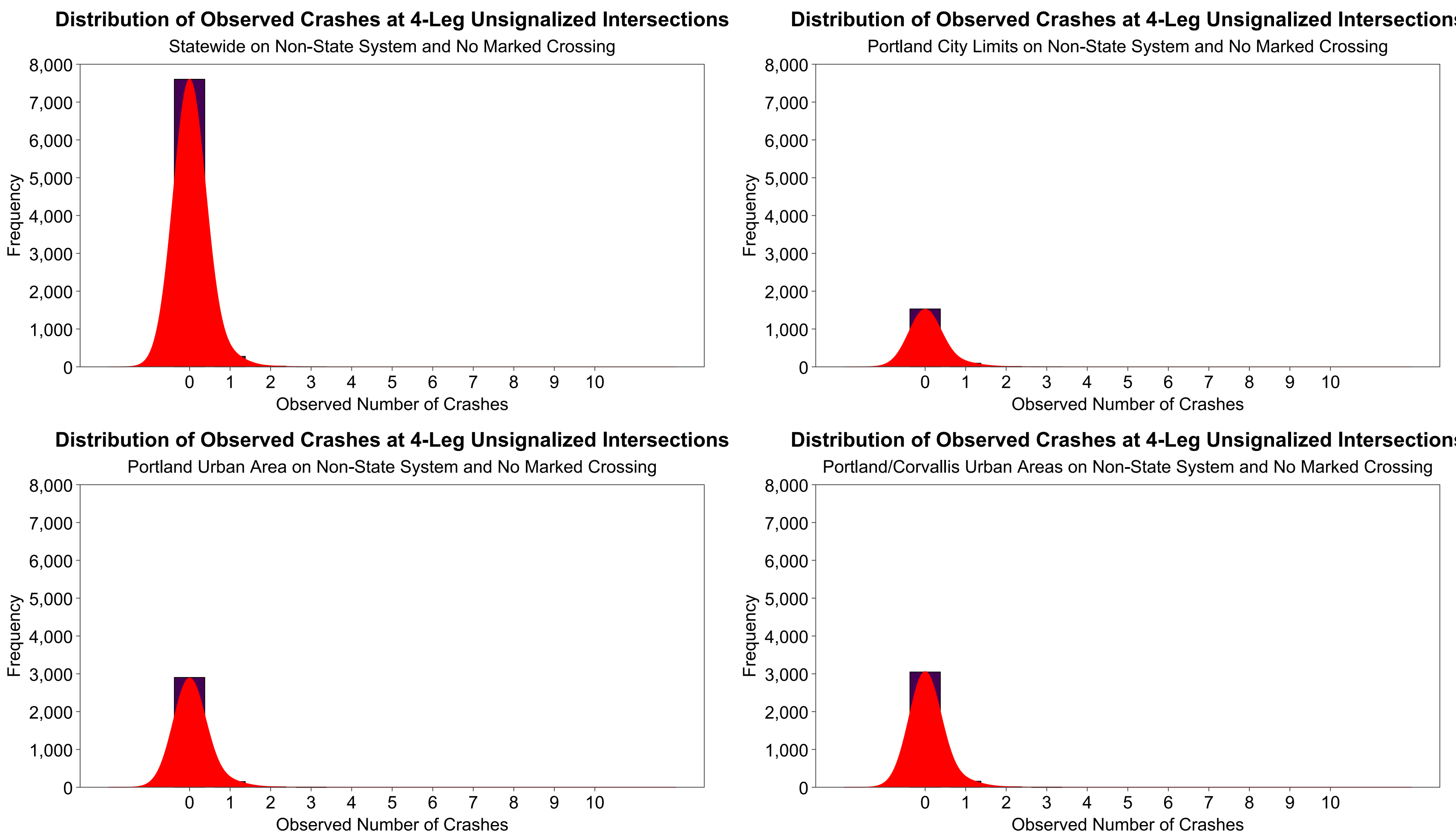


Figure D.8: Distribution of Observed Crashes at 4-Leg Unsignalized Intersections on Non-State System with No Marked Crossing

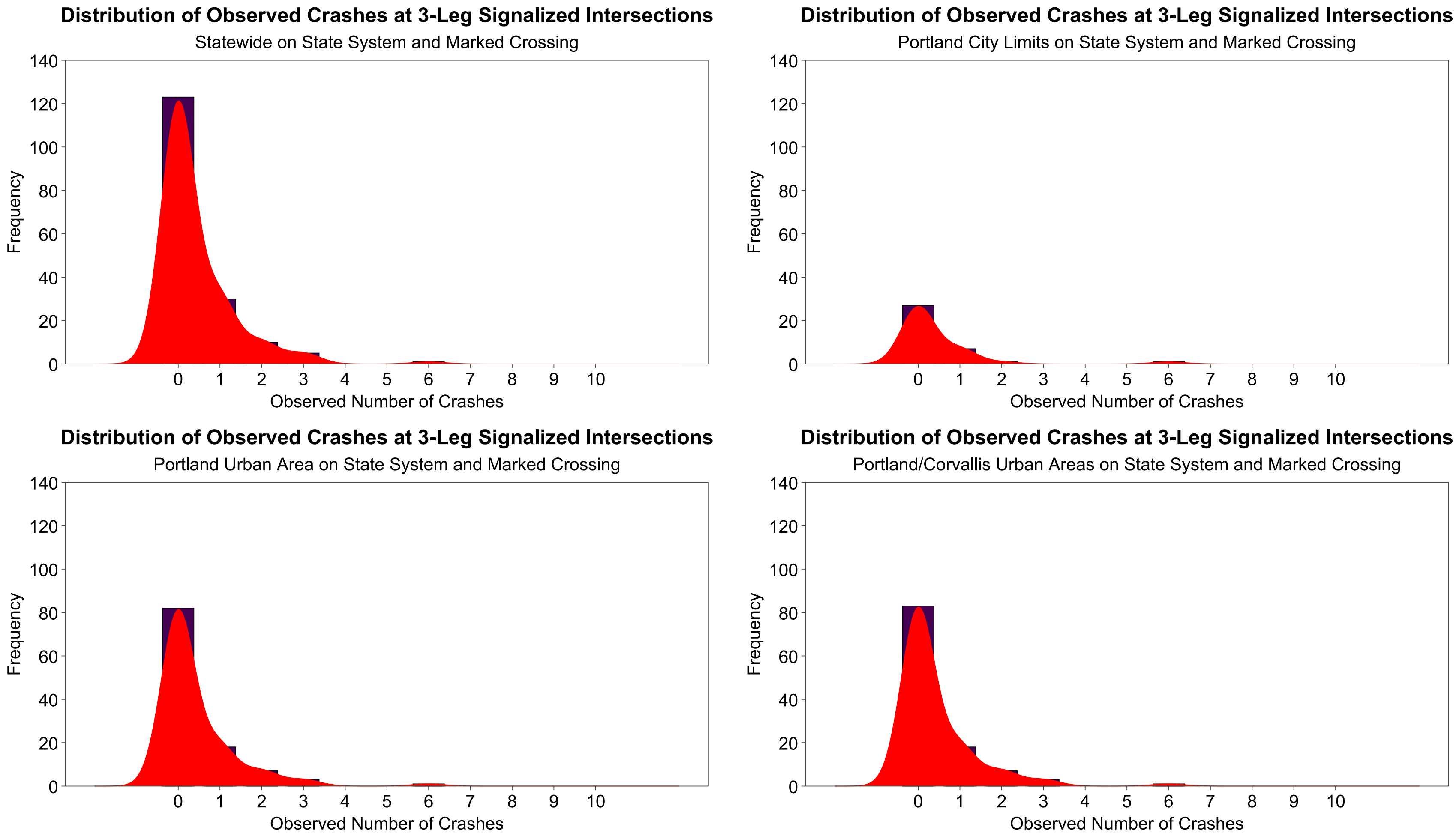


Figure D.9: Distribution of Observed Crashes at 3-Leg Signalized Intersections on State System with Marked Crossing

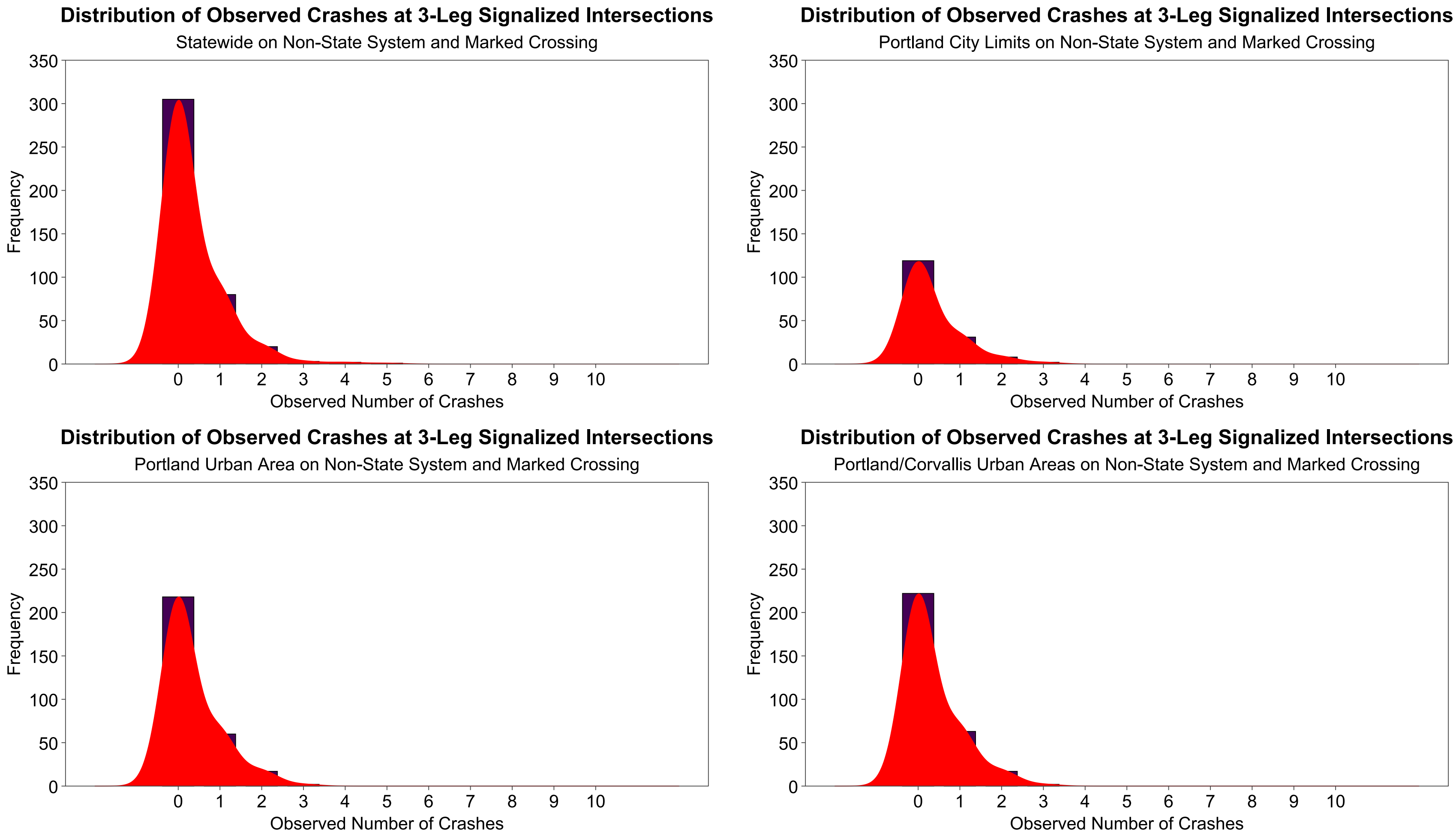


Figure D.10: Distribution of Observed Crashes at 3-Leg Signalized Intersections on Non-State System with Marked Crossing

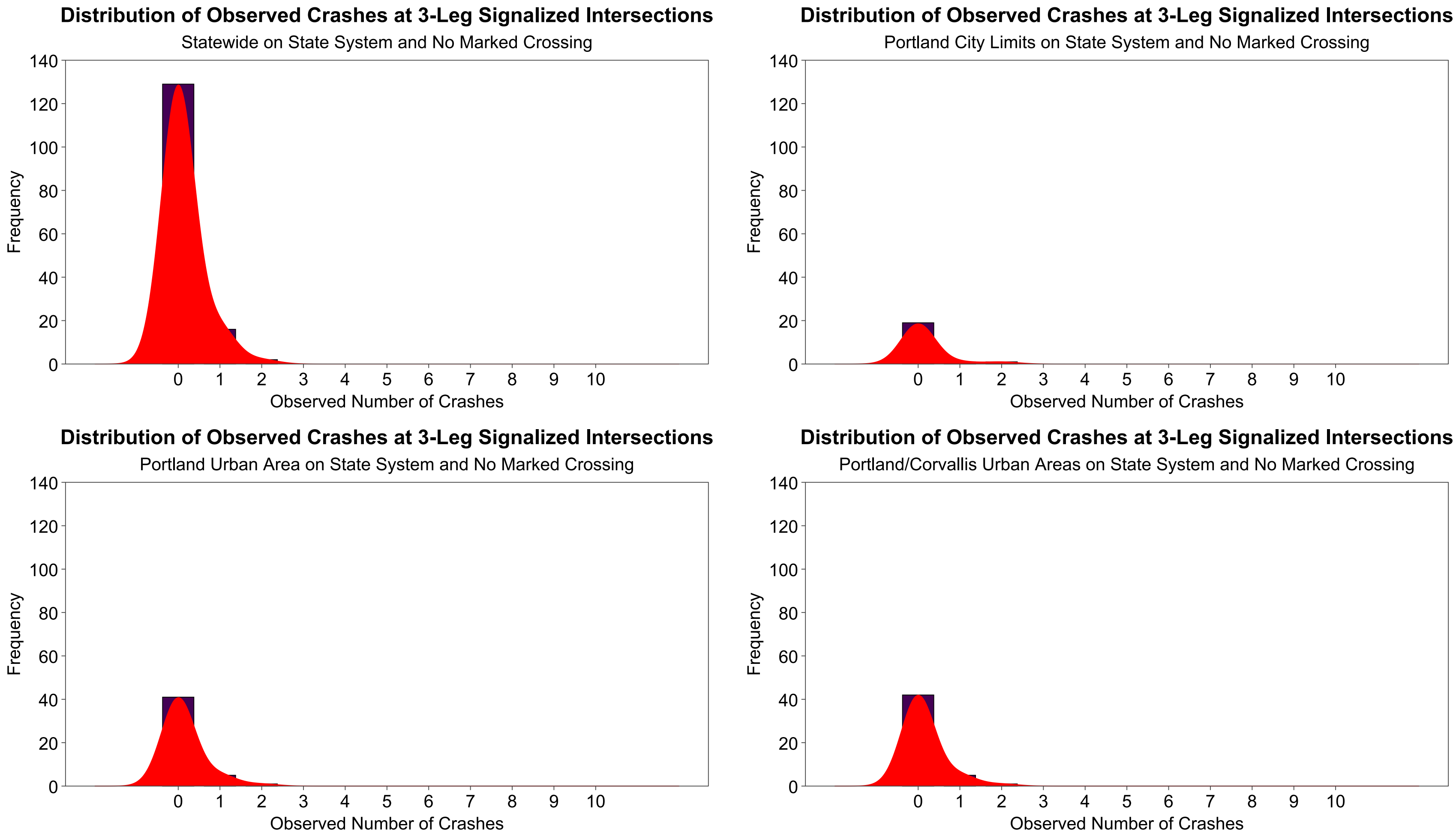


Figure D.11: Distribution of Observed Crashes at 3-Leg Signalized Intersections on State System with No Marked Crossing

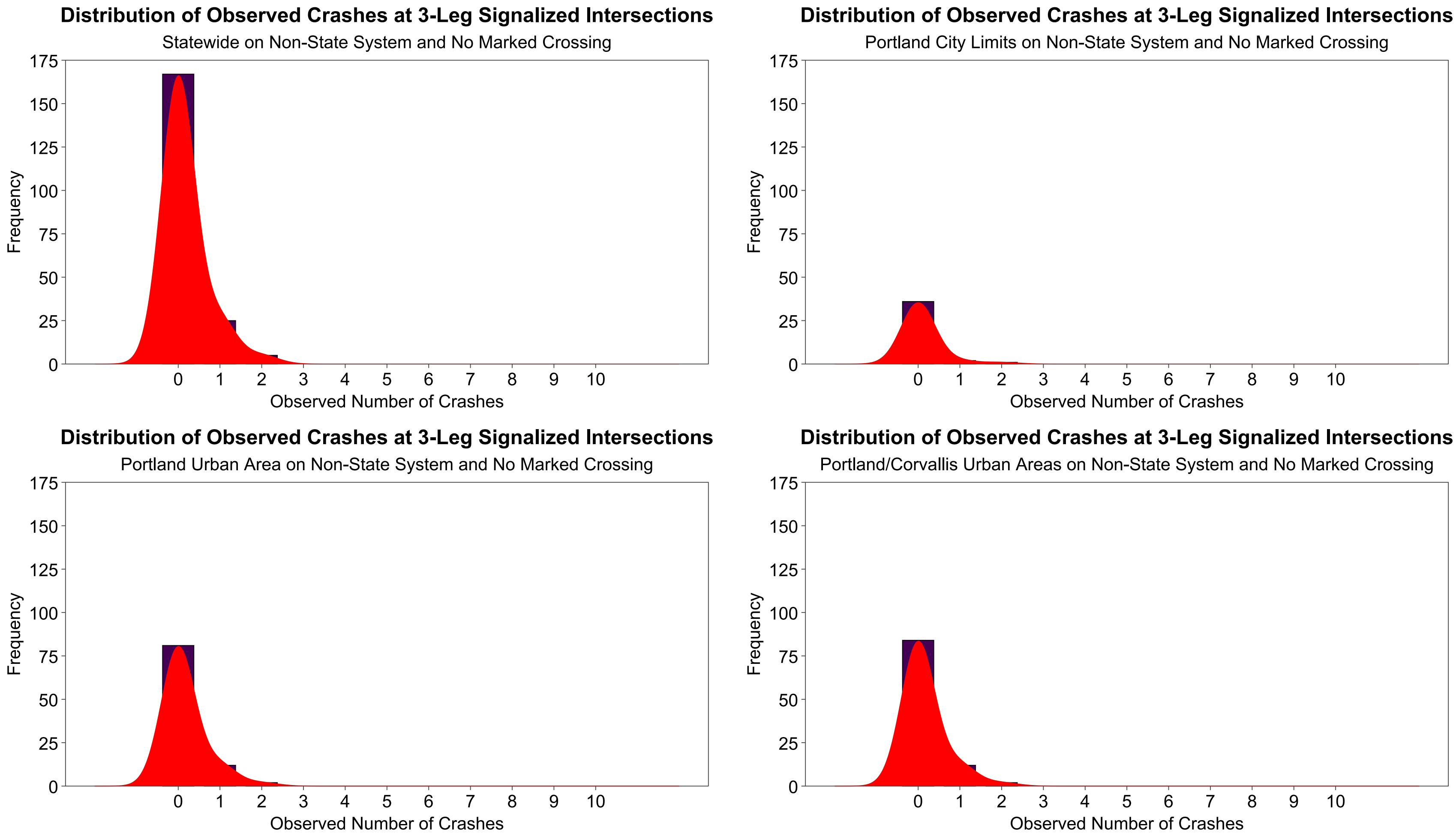


Figure D.12: Distribution of Observed Crashes at 3-Leg Signalized Intersections on Non-State System with No Marked Crossing

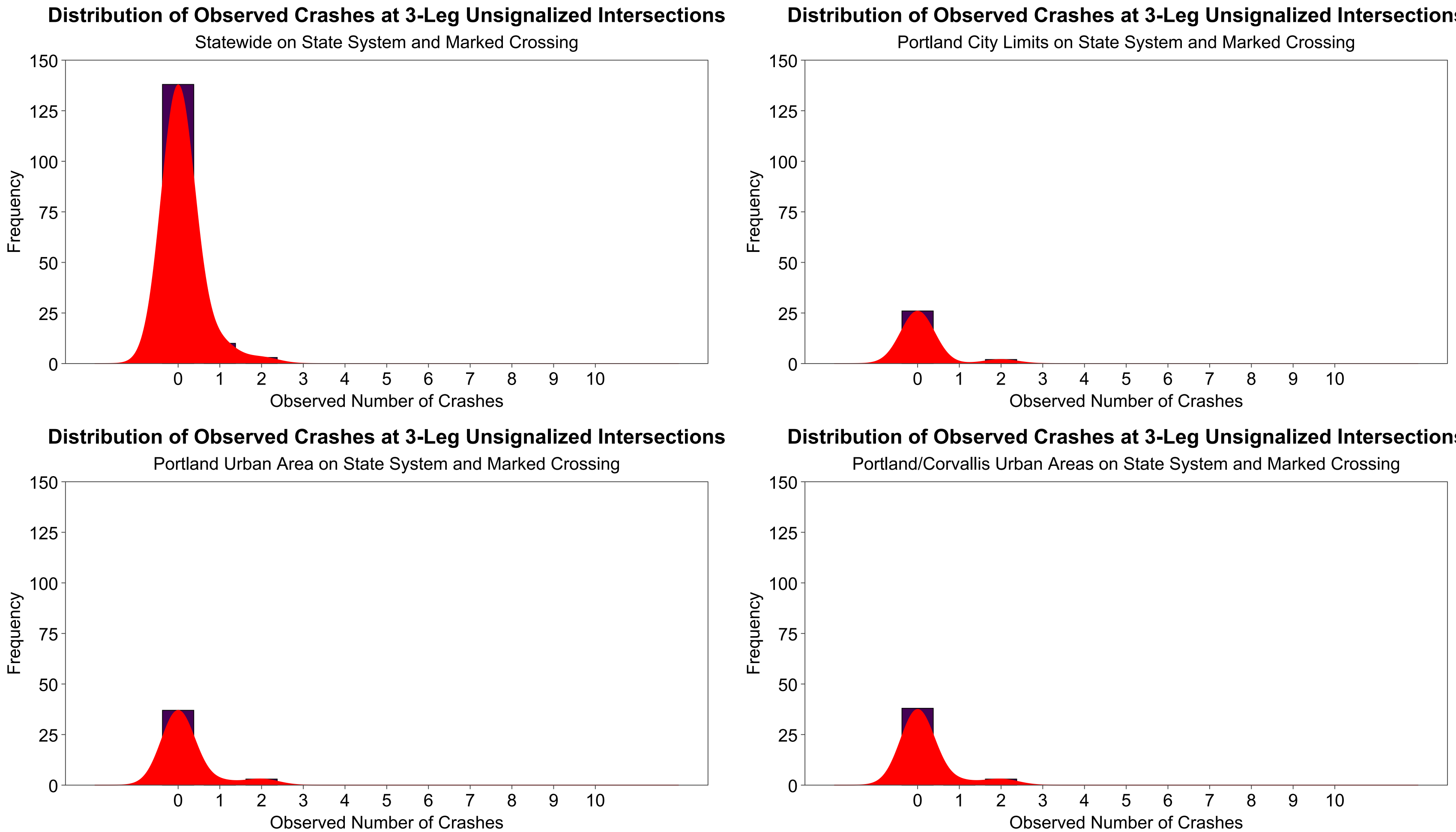


Figure D.13: Distribution of Observed Crashes at 3-Leg Unsignalized Intersections on State System with Marked Crossing

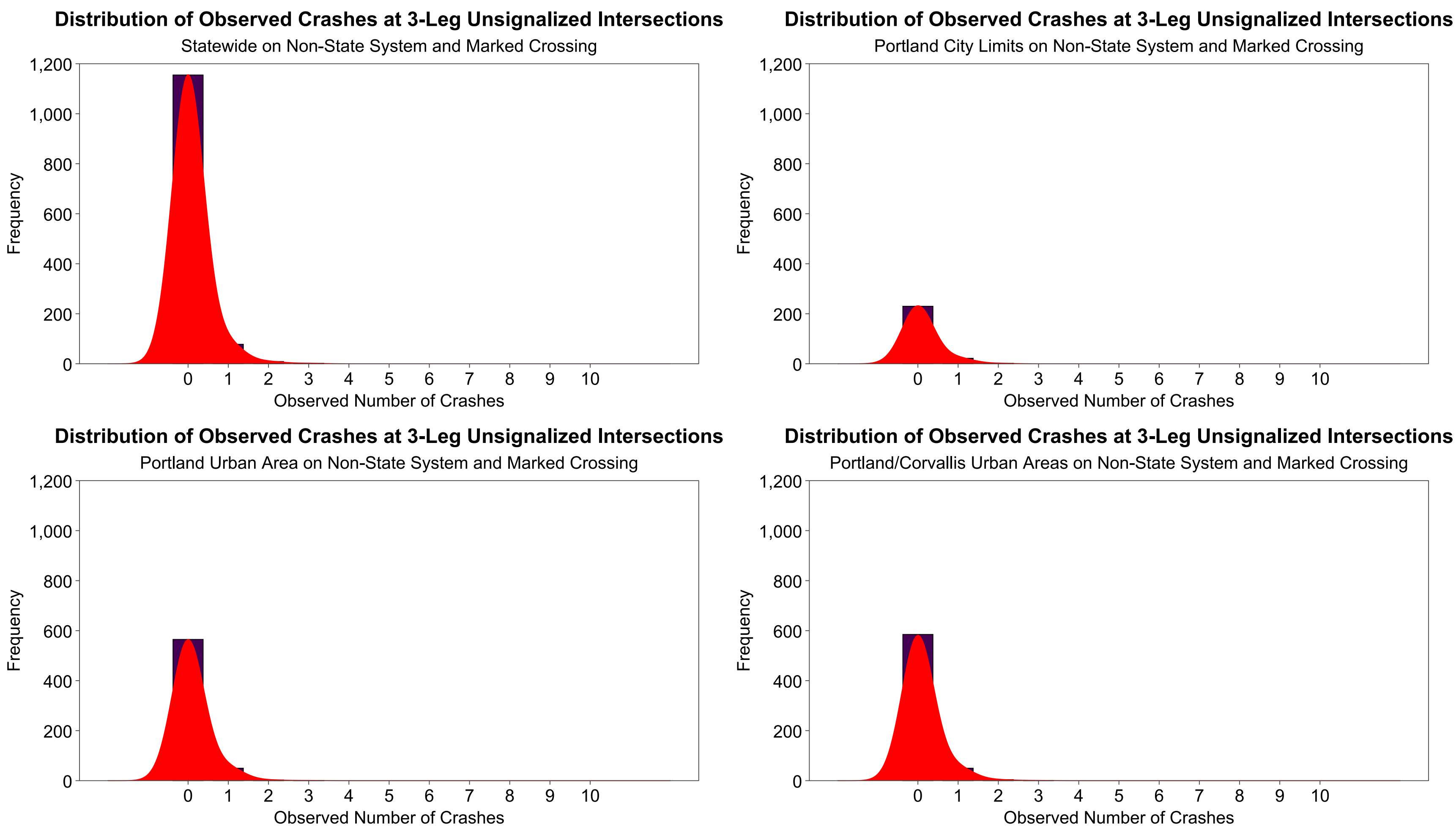


Figure D.14: Distribution of Observed Crashes at 3-Leg Unsignalized Intersections on Non-State System with Marked Crossing

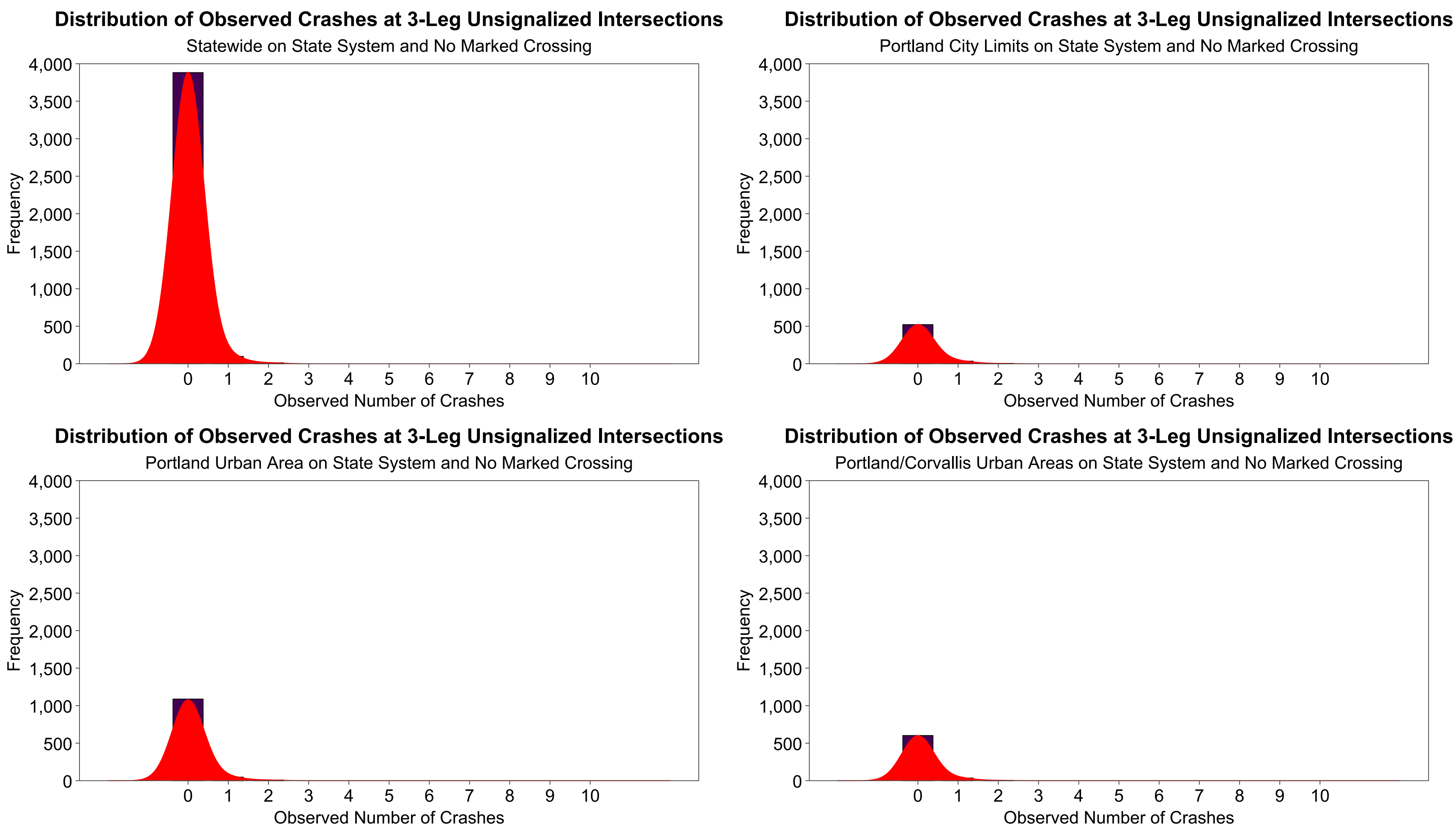


Figure D.15: Distribution of Observed Crashes at 3-Leg Unsignalized Intersections on State System with No Marked Crossing

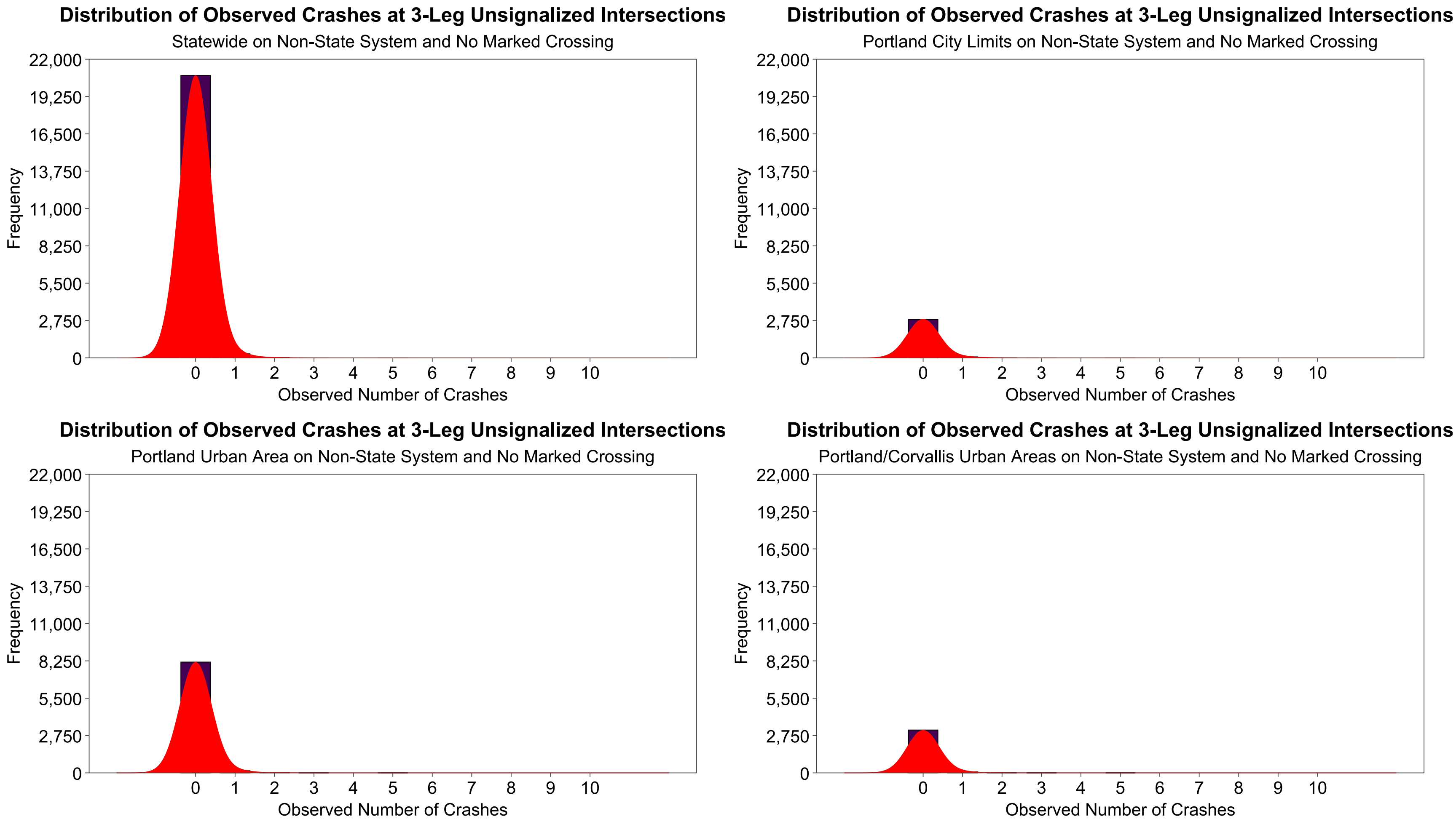


Figure D.16: Distribution of Observed Crashes at 3-Leg Unsignalized Intersections on Non-State System with No Marked Crossing