Systemic Predictive Safety Tools for Pedestrians: Benefits and Challenges

Wesley Kumfer, PhD, RSP1

THE UNIVERSITY OF NORTH CAROLINA
HIGHWAY SAFETY
RESEARCH CENTER

www.hsrc.unc.edu

January 12, 2023

Overview

- How is a systemic approach different than a traditional approach?
- What do the models show?
- How can the models be used to identify risks?
- What systemic tools do agencies use?
- What data are needed?
- What are the benefits and challenges of a systemic safety approach?

Overview

- The new FHWA VRU rules recommend the use of predictive safety tools and systemic safety tools.
- HSRC and colleagues have developed systemic safety analysis tools for pedestrians that combine:
  - Historic crash data
  - Predictive models
  - Network data for risk identification
- The work in these slides was developed during two projects:
  - NCHRP 17-73: Systemic Pedestrian Safety Analysis (Report 893)
  - Predictive Safety Analysis for Montgomery County

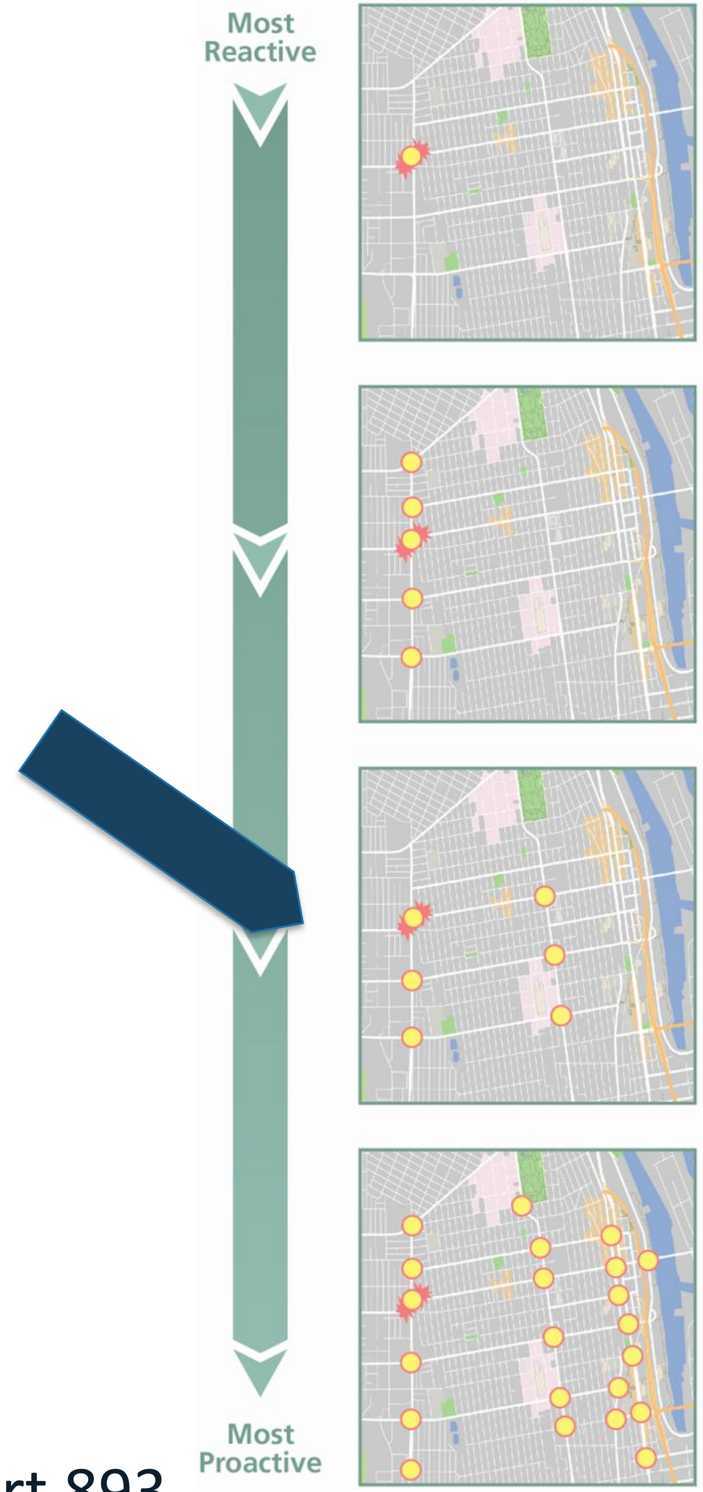
How is a systemic approach different than a traditional approach?

The Traditional Approach

- The traditional transportation safety management system tends to rely on “hot spot” (high crash location) identification.
- While this approach is important, there are several limitations:
  - A hot spot approach is inherently reactive and based on crash history.
  - Pedestrian crashes may be rare or widely dispersed across a network, making a hot spot approach unreliable and cost-ineffective.
  - Crash risk factors for pedestrians may be different than for motor vehicles.
  - The process needs to be tailored to data related to pedestrians, and to provide guidance on how to gather needed data.
- We need a method to identify locations of risk.

A Risk-Based Approach

- The systemic approach is an effective tool for identifying locations of risk proactively.
  - Identifies a safety concern based on an evaluation of data at the system (or network) level.
  - Establishes common characteristics (risk factors) of locations where severe crashes occur.
  - Emphasizes low-cost safety countermeasures to address the risk factors for high severity types of crashes.
  - Prioritizes locations across the entire roadway network where treatable risk factors are present, with or without a prior crash history.
- One way to identify risks is to develop Safety Performance Functions (SPFs) to predict crashes at different locations.



Source: NCHRP Report 893

The Systemic Approach

- Seven-step process

Systemic Pedestrian Safety Analysis Process

Step 1
Define Study Scope

Step 2
Compile Data

Step 3
Determine Risk Factors

Step 4
Identify Potential Treatment Sites

Step 5
Select Potential Countermeasures

Step 6
Refine and Implement Treatment Plan

Step 7
Evaluate Project and Program Impacts

Source: NCHRP Report 893

What do the models show?

Seattle Systemic Safety Analysis

- Statistically significant variables ($p<0.05$)

Exposure Measures and Activity surrogates – scale measures *	MV Straight Through struck pedestrian crossing at segment	All types MV struck ped. at Night - segments
MV VOLUME 1 - ADA AADT estimate log-transformed	+	n.s.
MV VOLUME 2 - RF Regression Estimate of AADT – scaled /10,000	n.s.	+
BUS Traffic - No. of buses stopping within 150 feet	+	+
Pedestrians Walking Along the segment (AADP raw estimate)	+	+
Pedestrians Walking Along the segment (AADP log)	–	–
Mean income area residents / 10,000	–	–
Commercial density within 1/10 mi	+	+
Light pole density (no. per mile)	+	+

Source: Kumfer et al., 2019

Seattle Systemic Safety Analysis

- Statistically significant variables ($p<0.05$)

Roadway and Built Environment – categorical variables*	**MV Straight Through struck pedestrian crossing at segment**	**All types MV struck ped. at Night - segments**
Speed Limit 30 or 35 compared to 25	**n.s.**	**+**
4 or 5 plus through lanes compared to 1 lane	**+**	**n.s.**
Two-way continuous left turn lane	**+**	**+**
One or two striped on-street parking lanes (compared to no striped parking)	**+**	**n.s.**
Dedicated right turn lane at either adjacent intersection	**+**	**n.s.**
Marked midblock crosswalk (one or more)	**+**	**+**
One-way traffic flow	**n.s.**	**–**
Urban village designation (increasing intensity/uses)	+	+

Source: Kumfer et al., 2019

Montgomery County Systemic Safety Analysis

- Crashes between motor vehicles and pedestrians at night at intersections (iPedDark).

Variable	Category (if applicable)	β Coefficient	Standard Error	p-value
Intercept	n.a.	-11.4039	0.786	<.0001
Average annual daily pedestrians	n.a.	-0.0003	0.0001	<.0001
Average annual daily traffic (categorical)	>= 10,000	1.5455	0.2634	<.0001
	5,000-9,999	1.5139	0.2473	<.0001
	< 5,000 (base)	n.a.		
Logarithmic transform of the AADP	n.a.	0.4843	0.0979	<.0001
Number of bus stops (1/10 mile)	n.a.	0.0185	0.0081	0.0227
Maximum number of through lanes on any leg	n.a.	0.1578	0.0507	0.0018
Number of metro stations (1/4 mile)	n.a.	0.2415	0.0669	0.0003
Total number of legs at the intersection	n.a.	0.4448	0.1271	0.0005
Number of marked crosswalks	n.a.	0.2069	0.0502	<.0001
Max speed limit among intersection legs	n.a.	0.0305	0.0128	0.0175
Intersection is signalized (categorical)	Yes	1.0537	0.1714	<.0001
	No (base)	n.a.		
Density of transportation facilities per square mile (1/2 mile)	n.a.	0.0185	0.0085	0.0295
Population density (people per square mile) (1/4 mile)	n.a.	0.2329	0.0906	0.0101
Percent of households with income of $100,000 or more (1/4 mile)	n.a.	-0.0216	0.004	<.0001
Dispersion	n.a.	0.8049	0.1982	n.a.

Source: Kumfer et al., in press

Montgomery County Systemic Safety Analysis

- Crashes between motor vehicles traveling straight and pedestrians at segments (SPMVStrt).

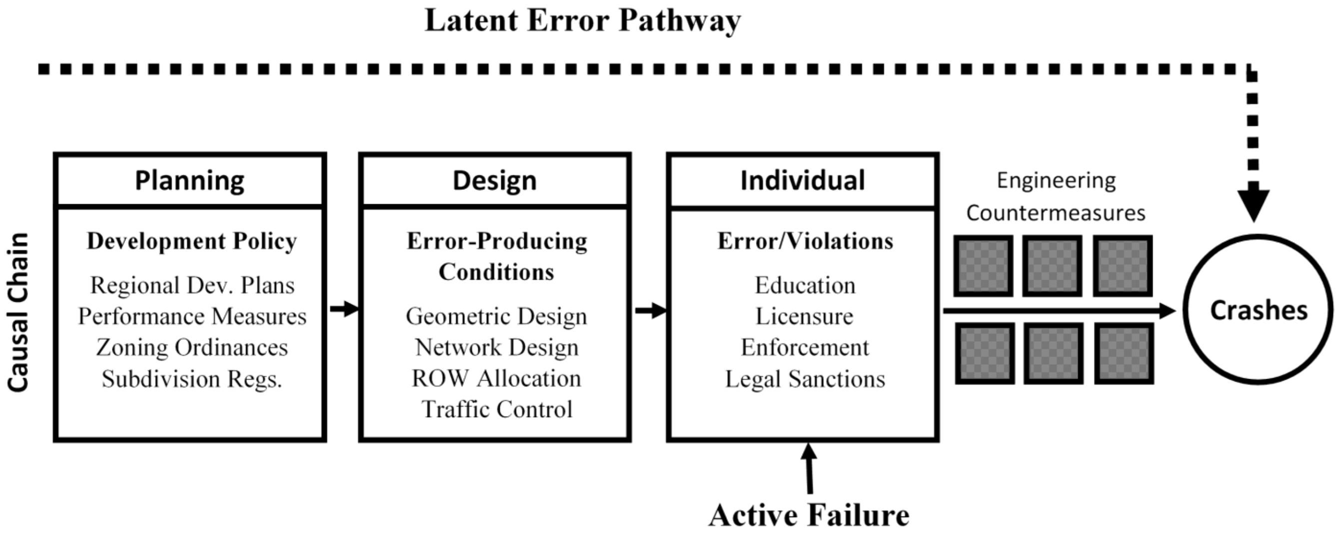
Variable	Category (if applicable)	β Coefficient	Standard Error	p-value
Intercept	n.a.	-5.2822	0.5197	<0.0001
Average annual daily pedestrians	n.a.	0.0002	0.0001	0.0013
Average annual daily traffic (categorical)	>= 10,000	1.1101	0.2007	<0.0001
	5,000-9,999	0.6876	0.1993	0.0006
	< 5,000 (base)	n.a.		
Number of bus stops (1/10 mile)	n.a.	0.0309	0.0054	<0.0001
Segment ends in a dead end (categorical)	Yes	-1.3952	0.5119	0.0064
	No (base)	n.a.		
Number of marked crosswalks	n.a.	0.2533	0.0897	0.0047
Percent of area within 500 feet in parking lots	n.a.	0.0257	0.0047	<0.0001
Length of the segment (in miles)	n.a.	0.7773	0.2047	0.0001
Roadway classification (categorical)	Major arterial	0.9682	0.5053	0.0553
	Minor arterial	0.9206	0.4976	0.0643
	Local street	0.0359	0.4746	0.9397
	Other (base)	n.a.		
Density of off-premise alcohol locations (1/10 mile)	n.a.	0.0108	0.0044	0.0139
Density of businesses per square mile (1/2 mile)	n.a.	-0.0065	0.0028	0.0218
Density of recreational facilities per square mile (1/2 mile)	n.a.	-0.0023	0.001	0.0259
Percent of households with income of $100,000 or more (1/4 mile)	n.a.	-0.0312	0.0036	<0.0001
Dispersion	n.a.	0.7788	0.297	n.a.

Source: Kumfer et al., in press

How can the models be used to identify risks?

Examining Risk

- Systemic approaches can help us better understand where latent risks exist in our transportation network.
- We can use the coefficients of the SPF models to identify “risk factors” for treatment.
- We can scan the entire roadway network for individual risk factors, or we can package risk factors based on how they can be treated.



Latent Error Pathway

Causal Chain

Planning

Development Policy

Regional Dev. Plans
Performance Measures
Zoning Ordinances
Subdivision Regs.

Design

Error-Producing Conditions

Geometric Design
Network Design
ROW Allocation
Traffic Control

Individual

Error/Violations

Education
Licensure
Enforcement
Legal Sanctions

Engineering Countermeasures

Crashes

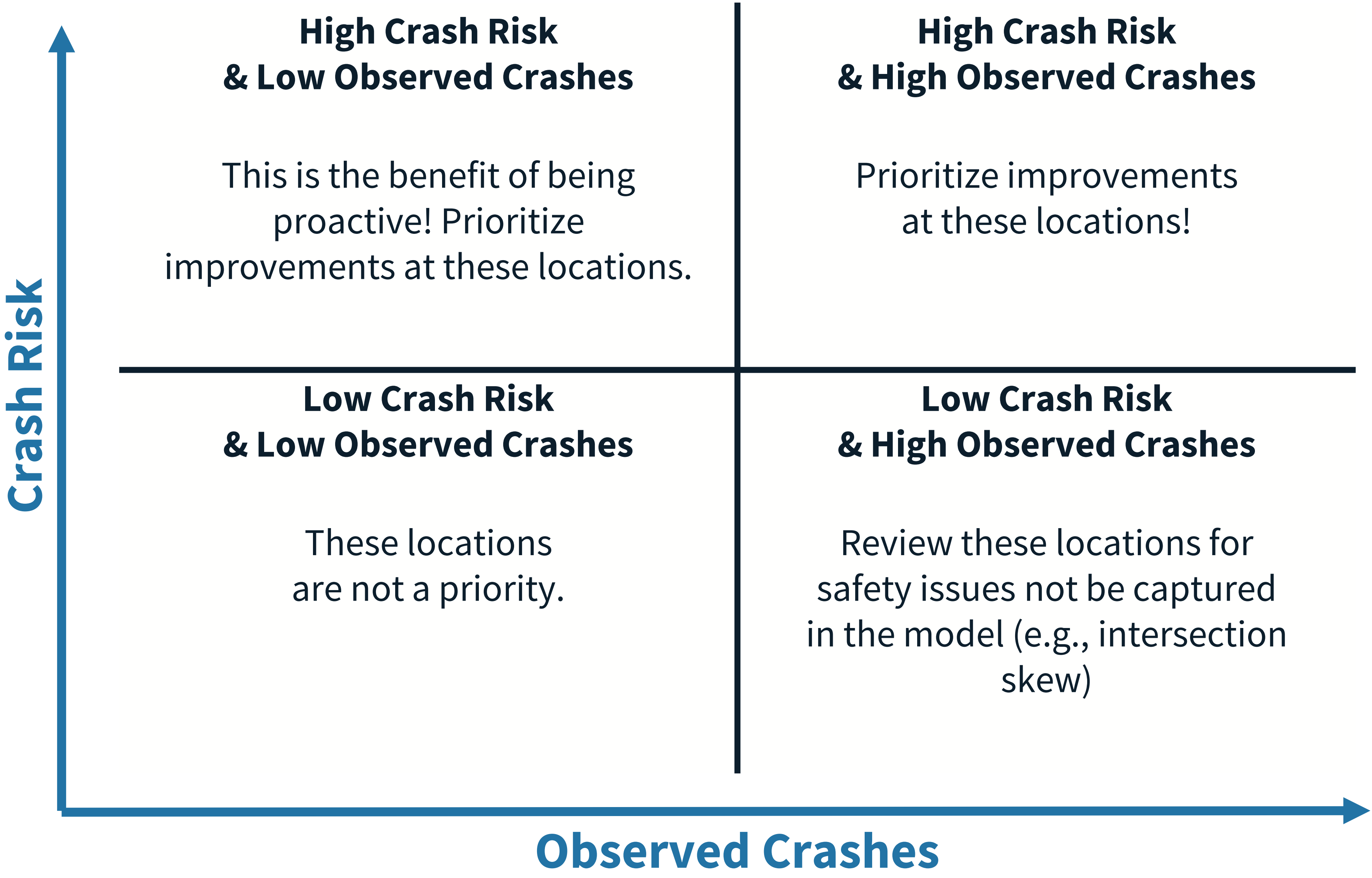
Active Failure

Source: Dumbaugh et al., 2019

Methods for Ranking Sites by Risk Factor

- **Observed Crashes** are the historical crashes. These are the basis of most crash analysis but are biased by the random nature of crashes.
- **Predicted Crashes** are the outcome of the SPFs and account for the characteristics in the SFP equation. They are useful for identifying sites which may not have many observed crashes but have the potential to be high-crash sites based on their characteristics.
- **Empirical-Bayes (EB) Crashes ("Expected Crashes")** weighs both observed and predicted crashes based on 1) how well the SPF predicts crashes and 2) the number of predicted crashes at the specific location. EB crashes are the most reliable estimate of the underlying crash frequency at a given location based on all available information.

High Crash Location Identification



High Crash Risk & Low Observed Crashes

This is the benefit of being proactive! Prioritize improvements at these locations.

High Crash Risk & High Observed Crashes

Prioritize improvements at these locations!

Low Crash Risk & Low Observed Crashes

These locations are not a priority.

Low Crash Risk & High Observed Crashes

Review these locations for safety issues not be captured in the model (e.g., intersection skew)

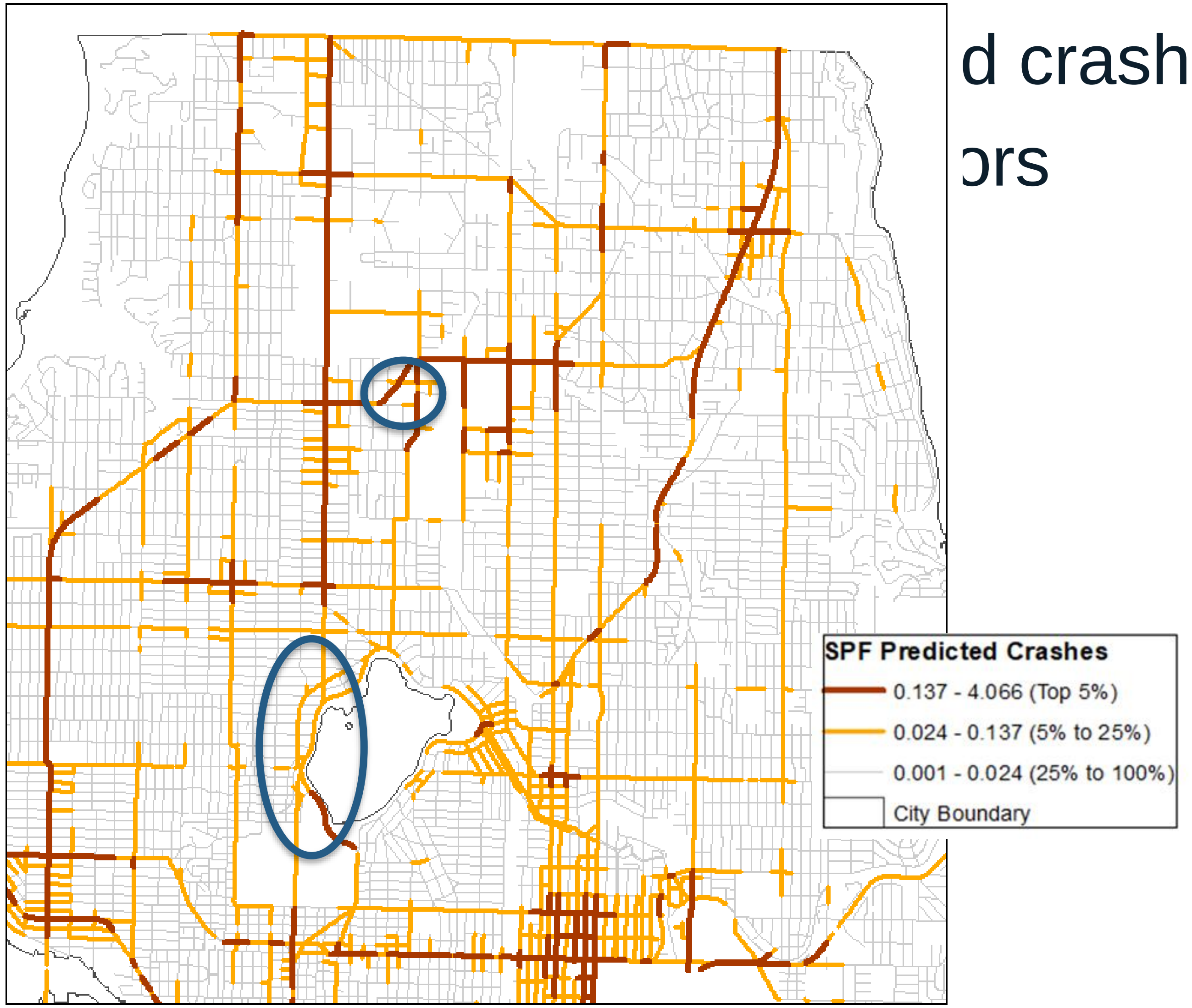
Source: Montgomery County Planning Department, 2022

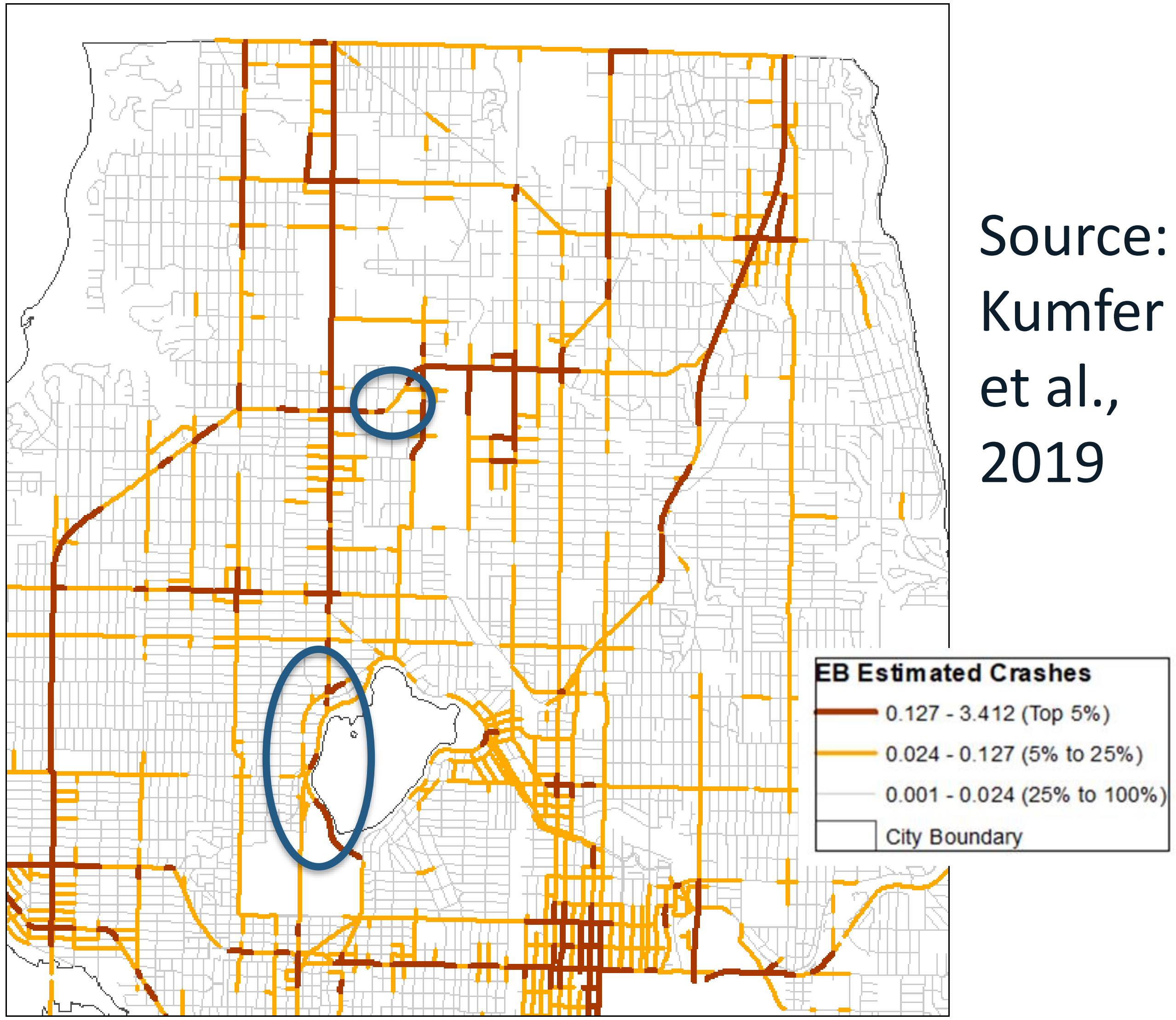
Seattle Site Ranking

- We worked with the City of Seattle to map our results.
- Segments ranked based on two criteria

d crash risk

ors





Source: Kumfer et al., 2019

Seattle Risk Package Identification

- Review of potential countermeasures to pedestrian risk.

Suitable for Signalized Intersections Only (or where signal is added)	Suitable for Unsignalized Locations Only (midblock or intersection)	Suitable for Either Signalized or Unsignalized Crossing Locations (including midblock)
• Leading pedestrian interval • Longer pedestrian phase • Restricted left turn (protected crossing phase)	• In-roadway yield-to-pedestrian (R1-6) sign/gateway • Advance stop/yield bar and R1-5/5a sign • Pedestrian hybrid beacon (sometimes called a HAWK signal)	• High visibility crosswalk • Traffic calming (raised device) • Median crossing island • Reduce number of lanes road diet • Curb extension and parking restriction • Location-specific lighting improvement

Source: NCHRP Report 893

Seattle Risk Package Identification

- Countermeasures can be selected based on ranking by known risk factors, matched with relevant treatments

Risk Characteristics	Number of Segments (of top 500 SPF-pred.)	SPF-Prediction (average SPF-predicted crashes per site per year)	Prior Observed Crashes (average observed crashes per site per year)	Potential Countermeasures
4+ thru lanes (29.4 mi)	357	0.061	0.064	Road diets and/or median islands
TWLTL (15.3 mi)	152	0.053	0.067	Median island (with/without road diet)
4+ thru lanes & TWLTL (12.7 mi)	129	0.054	0.067	Road diets and/or median islands
4+ lanes & Parking (9.5 mi)	102	0.074	0.060	Road diets and/or median islands AND Curb extension + parking restriction
4+ lanes, TWLTL & Parking subset (3.1 mi)	25	0.055	0.085	Road diets and median islands; AND Curb extension + parking restriction

Source: Kumfer et al., 2019

Seattle Risk Package Identification

- Countermeasures can be selected based on ranking by known risk factors, matched with relevant treatments

Risk Characteristics	Number of Segments	Average SPF-Prediction	Average Prior Crashes	Potential Countermeasures
MB Xwalk present (6.4 mi of segments)	50	0.094	0.090	• High visibility crosswalks; potentially many others including: • In-Rdway Yield to Peds signs (2-3 lane sites with lower volume/speed), and others for higher volume/speed rds.
MB Xwalk & 4+ ThrLns (2.2 mi)	18	0.118	0.079	• High visibility Xwalks • Road diets / median islands • Advance stop/yield markings + signs • PHBs (high vol./high speed)
MB Xwalk, 3+ ThrLns, & On-Street Parking subset (1.6 mi)	12	0.169	0.104	• High vis Xwalks • Road diets and/or median islands • Advance stop/yield markings + signs • PHBs (high vol./high speed); AND • Curb extension + parking restriction
4+ Lns & Right turn lane at adjacent intersection subset (4.2 mi)	40	0.076	0.088	Investigate safety issues/conflicts.

Source: Kumfer et al., 2019

Montgomery County High Crash Location Identification

Equity Emphasis Areas vs. Non-Equity Emphasis Areas

EEA	# Ints.	Intersection Crashes				# Segs.	Segment Crashes	
		Ped Dark	Bike	Left Turn	Angle		Ped Seg	Single Veh
Total Expected Crashes (# Annual Crashes)								
EEA	3,087	49	25	253	280	5,049	32	125
Non-EEA	13,606	58	62	482	595	26,033	51	663
High Expected Crash Locations (# Locations within the Top 200)								
EEA	3,087	107	67	80	75	5,049	133	26
Non-EEA	13,606	93	133	120	125	26,033	67	174
Average Expected Crashes (# Annual Crashes per Location)								
EEA	3,087	0.02	0.01	0.08	0.36	5,049	.007	.025
Non-EEA	13,606	0.00	0.00	0.04	0.20	26,033	.002	.026

Highlighted cells have the highest value for any column.

Source: Montgomery County Planning Department, 2022

What systemic tools do agencies use?

Overview of Systemic Safety Analysis Methods

- Multiple systemic safety analyses have recently focused on pedestrian safety and used a variety of methods.
  - Grembek et al. (2013) – Systemic hot spot identification matrix analysis for California arterials.
  - Kimley-Horn and Associates (2017) – Risk factor identification and benefit-cost analysis for state highway system in Arizona.
  - Thomas et al. (2018) and Kumfer et al. (2019) – Safety performance functions and Empirical Bayes analysis across Seattle network.
  - Foster et al. (2020) – Equivalent property damage only risk weighting across urban roadways in Oregon.
  - Gooch et al. (2022) – Binary logistic regression and odds ratios for midblock locations (arterials and collectors) in Massachusetts
- The Montgomery Planning analysis generally followed the methods described by NCHRP Report 893 (Thomas et al., 2018).

1. Compile data

2. Estimate volumes

3. Identify key crash types

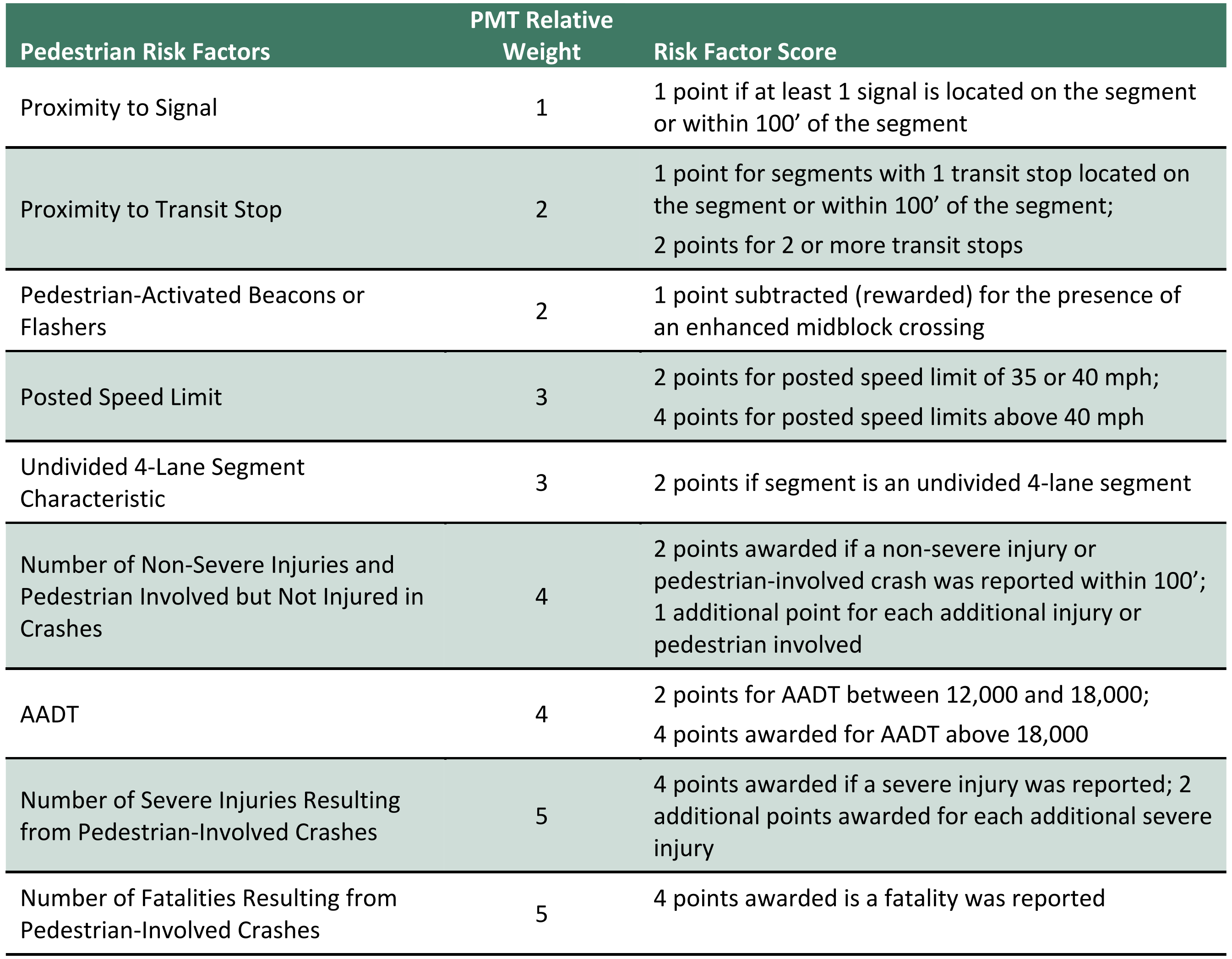
4. Develop Safety Performance Functions

5. Identify high-risk locations

6. Identify countermeasures

Oregon DOT

- ODOT has a less data-intensive approach that relies on an expert panel to weight risks.

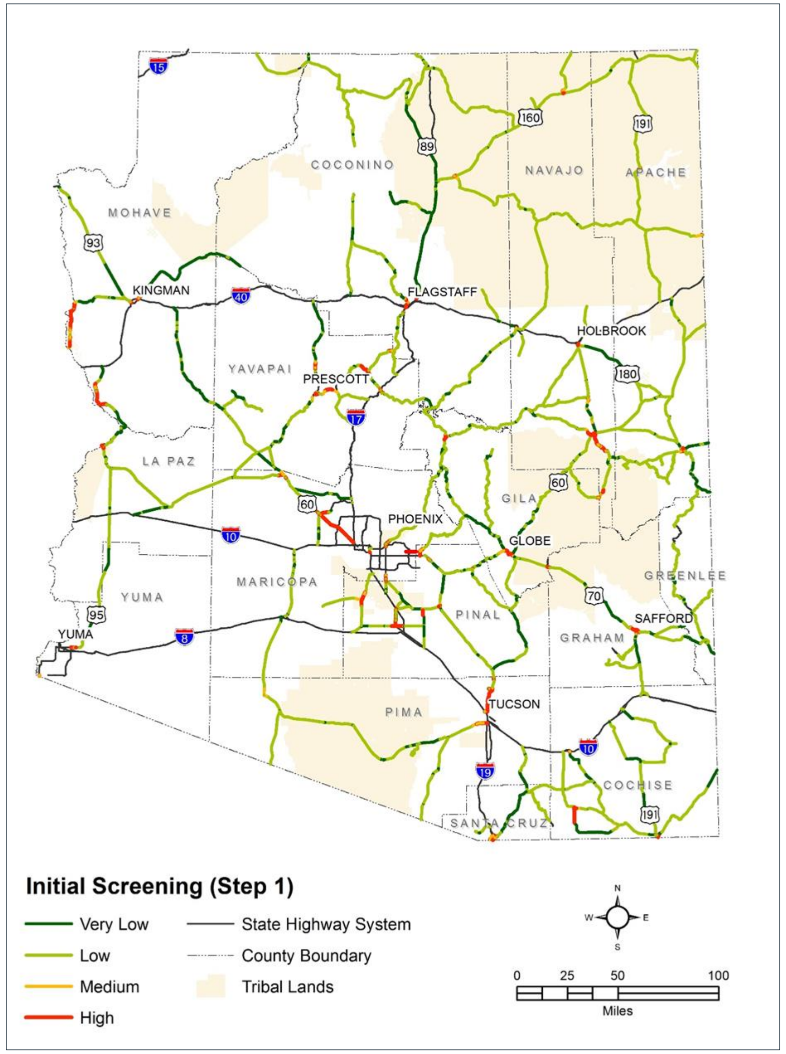


Pedestrian Risk Factors	PMT Relative Weight	Risk Factor Score
Proximity to Signal	1	1 point if at least 1 signal is located on the segment or within 100’ of the segment
Proximity to Transit Stop	2	1 point for segments with 1 transit stop located on the segment or within 100’ of the segment; 2 points for 2 or more transit stops
Pedestrian-Activated Beacons or Flashers	2	1 point subtracted (rewarded) for the presence of an enhanced midblock crossing
Posted Speed Limit	3	2 points for posted speed limit of 35 or 40 mph; 4 points for posted speed limits above 40 mph
Undivided 4-Lane Segment Characteristic	3	2 points if segment is an undivided 4-lane segment
Number of Non-Severe Injuries and Pedestrian Involved but Not Injured in Crashes	4	2 points awarded if a non-severe injury or pedestrian-involved crash was reported within 100’; 1 additional point for each additional injury or pedestrian involved
AADT	4	2 points for AADT between 12,000 and 18,000; 4 points awarded for AADT above 18,000
Number of Severe Injuries Resulting from Pedestrian-Involved Crashes	5	4 points awarded if a severe injury was reported; 2 additional points awarded for each additional severe injury
Number of Fatalities Resulting from Pedestrian-Involved Crashes	5	4 points awarded is a fatality was reported

Source: NCHRP Report 893

Arizona DOT

- ADOT used a less data-intensive approach that relied on past research to assign weights to pedestrian risk factors.
- For example:
  - Width of Roadway
    - 6-Lane Highway = 6 points
    - 4- or 5-Lane Undivided Highway = 3 points
    - 2- or 3-Lane Undivided Highway = 2 points
    - 2- or 3-Lane Divided Highway = 1 point
  - Posted Travel Speed
    - >45 miles-per-hour (mph) = 6 points
    - 35 to 45 mph = 4 points
    - 25 to 35 mph = 2 points
    - <25 mph = 0 points



Source: Kimley Horn et al., 2017

What data are needed?

Seattle Data Collection and Cleaning

- Compile crash data, exposure (volume) data, roadway data, and context data

Data	Source
Comprehensive crash database	City of Seattle
Roadway network geodatabase	
Generalized land use	
Building footprints	
University locations (volume estimate model only)	
Schools	
Short-term, quarterly, and continuous user count data used in pedestrian volume estimation	
Census blocks and demographic/employment data	U.S. Census Bureau
National elevation dataset	U.S. Geological Survey
Transit stop location and schedule data	Google Transit Feed Specification, Sound Transit

Source: NCHRP Report 893

Montgomery County Data Collection and Cleaning

The databases developed for pedestrian crashes consisted of 170 segment variables and 282 intersection variables.

Transportation Characteristics

- Speed limit
- Number of lanes
- Roadway slope
- Presence and type of crosswalk
- Presence and type of bicycle facility
- Roadway classification
- Intersection control
- Lighting
- Transit service

Land Use Characteristics

- Parks
- Hospitals
- Gas stations
- Parking lots
- Schools
- Government facilities
- Shopping centers
- Alcohol-serving locations
- Population density
- Employment density

Demographic Characteristics

- Equity Emphasis Areas
- Income distribution
- Race/ethnicity distribution
- Age distribution

All linked to crash data!

Source: Montgomery County Planning Department, 2022

What are the benefits and challenges of a systemic safety approach?

Benefits

- Systemic tools range in complexity from weighted models to rigorous statistical models that can reveal where crashes are expected to occur, even when historic crash data aren’t available.
- Systemic tools can build on predictive models (SPFs) that can be locally calibrated.
- Systemic tools built on SPFs can account for regression to the mean through the application of the EB method.
- Systemic tools can be used to make good use of available project funding.

Montgomery County Countermeasure Selection

Example Scenarios for Reducing Pedestrian Crashes along Segments with $200,000 (10-Year Impact)

Scenarios	High-Visibility Crosswalks	Speed Humps	Pedestrian Hybrid Beacon
Number of Locations	65	40	1
Total Estimated Cost	$199,600	$200,000	$175,000
Predicted Crash Reduction	22	129	4
Crash Reduction per Location	<1	3	4
Cost per Crashes Reduced	$9,200	$1,600	$44,600
% of Locations in Equity Emphasis Areas	54%	53%	100%

Source: Montgomery County Planning Department, 2022

Challenges

- Collecting exposure data can be challenging.
- Local expertise may not be available to develop SPFs.
- Data requirements for calibration may be intensive.
- It may be difficult to get good model fit for rare crash types (like crashes involving bicyclists), so there may be a greater need for reliance on risk factors identified in the literature.

Montgomery County Volume Estimation Models

- Although the databases contained a substantial amount of data, exposure estimates were missing.
- We used negative binomial regression to develop exposure estimates by:
  - Compiling counts from development projects, MCDOT, and SHA
  - Standardizing counts based on time of day, day of week, and season
  - Estimating counts at all intersections and segments based on transportation, land use, and demographic attributes.

Montgomery County Volume Estimation Models

- AADT and AADP estimates are used in pedestrian SPFs.

Transportation Variables

Variable Description	Type	AADT	AADP
Density of bikeshare per square mile	Numeric	-	X
Segment has separated bikeway or sidepath	Categorical	X	-
Segment has striped bikeway	Categorical	X	-
Density of bus routes along a segment	Numeric	X	-
Number of bus stops	Numeric	X	-
Highest category bikeway	Categorical	X	X
Segment ends in a dead end	Categorical	X	-
Percent of area within 75 feet with driveways	Numeric	X	-
Maximum slope along the segment	Numeric	X	-
Segment has a median	Categorical	X	-
Derived interaction variable for the effect of medians and speed limits	Categorical	X	-
Number of metro stations	Numeric	-	X
Percent of area within 500 feet comprised of parking	Numeric	X	X
Speed limit	Categorical	X	X
Roadway classification	Categorical	-	X
Percent coverage of sidewalk	Categorical	X	X

Source: Kumfer et al., in press

Montgomery County Volume Estimation Models

- AADT and AADP estimates are used in pedestrian SPFs.

Land Use Variables			
Variable	Type	AADT	AADP
Density of on-premise alcohol locations	Numeric	X	-
Density of on-premise alcohol locations	Numeric	-	X
Area type	Categorical	X	X
Policy area category	Categorical	X	X
Employment density	Numeric	-	X
Located within a municipality	Categorical	X	X
Number of signals adjacent to the segment	Categorical	-	X
Number of lanes along segment	Numeric	-	X
Percent of area within 100 feet as park land	Numeric	X	-
Percent of area within 500 feet as park land	Numeric	-	X
Density of apartments per square mile	Numeric	X	X
Density of educational facilities per square mile	Numeric	X	-
Density of educational facilities per square mile	Numeric	-	X
Density of emergency services per square mile	Numeric	X	-
Density of medical facilities per square mile	Numeric	-	X
Density of shopping centers per square mile	Numeric	-	X

Source: Kumfer et al., in press

Montgomery County Volume Estimation Models

- AADT and AADP estimates are used in pedestrian SPFs.

Demographic Variables			
Variable Description	Type	AADT	AADP
Percent African American population in EEAs	Numeric	-	X
Percent Asian population in EEAs	Numeric	-	X
Located within an EEA	Categorical	-	X
Median household income	Numeric	-	X
Percent of households with income of $200,000 or more	Numeric	-	X
Percent of population under 18 years old	Numeric	X	-
Percent of population 65 years and up	Numeric	-	X
Percent of population 65 years and up	Numeric	X	-

Source: Kumfer et al., in press

Conclusions

- Our roadway environments contain many risks for pedestrians.
- Pedestrian crashes are widely dispersed, so identifying treatment locations can be difficult.
- A systemic, risk-based approach can be useful for leveraging crash, exposure, and context data to uncover sites for treatment.
- Achieving a safe system requires us to be proactive in addressing risks.
- Systemic safety tools can be data intensive to build and apply.

Questions?

Thank you!

THE UNIVERSITY OF NORTH CAROLINA
HIGHWAY SAFETY
RESEARCH CENTER

www.hsrc.unc.edu

January 12, 2023