MassDOT IMPACT Phase II - Identification of Risk Factors for SHSP Emphasis Areas

Pedestrian Crashes

PREPARED FOR

Massachusetts Department of Transportation

PREPARED BY

vhb

REPORT DATE: AUGUST 2021

Purpose & Background

MassDOT was awarded a grant by the United States Department of Transportation (USDOT) under its Safety Data Initiative (SDI) competition. As part of this work, MassDOT is identifying focus crash types, facility types, and risk factors for their Strategic Highway Safety Plan (SHSP) Emphasis Areas. This report summarizes the risk factor analysis performed for pedestrian crashes. It also represents a model method which can be used throughout the SDI for analyses of infrastructure-based emphasis areas and exposure. Reports for other emphasis areas may describe different methods used to adapt to the needs of those areas.

This report expands upon the preliminary analysis results summarized in a report delivered to MassDOT on May 28, 2020. This preliminary analysis report describes a zonal analysis that identified socioeconomic and demographic risk factors related to pedestrian fatal (K), serious injury (A), and non-incapacitating (B) injury crashes that occurred between 2013-2017. The previous analysis provided some insight on geographic locations where high-severity pedestrian crashes occur; however, it did not consider any roadway or traffic factors that may lead to an increased likelihood of a fatal or serious injury pedestrian crash, nor did it consider pedestrian exposure due to a lack of such data.

The objective of this additional analysis is to identify a series of potential risk factors, including demographic, socioeconomic, and road-based, that lead to non-intersection, mid-block non-motorized crashes. Once approved by MassDOT, these factors can be applied by MassDOT to assess the risk of severe pedestrian crashes on roadway segments in Massachusetts.

This report is separated into two sections. The first section documents a comparison of contributing circumstances in fatal and serious injury pedestrian crashes to crashes of all severities between 2013 and 2017. Based on these observations, VHB performed a binary logistic regression of principal arterials, minor arterials, and major collectors in Massachusetts to assess the impact of road, traffic, and socioeconomic characteristics on the probability of a KA pedestrian crash on a given segment of road. The second section of this report provides the results of this analysis and prioritizes individual risk factors for MassDOT's consideration. This report recommends possible applications of risk factors identified in both analyses.

Crash Severity Comparisons

MassDOT provided VHB with pedestrian crash data for a five-year period between 2013 and 2017. Intersection-related crashes were excluded for the purposes of this analysis, leaving only mid-block crashes. VHB defined a midblock crash as those with the following codes in the "RDWY_JNCT_TYPE_DESCR" field within MassDOT's crash data:

- Not at Junction.
- Driveway.
- Not Reported.
- Unknown.

The project team felt it was appropriate to include driveway crashes in this midblock study due to the potential risk posed to pedestrians at commercial and mixed-use driveways. These conflicts are inherent to midblock crossings, and segments with a high frequency of driveways and high non-motorized user volumes could potentially have the highest frequency of fatal and serious injuries. Furthermore, "Not Reported" and "Unknown" crashes accounted for a small percentage of total observations (e.g., 5.8% of KABCO pedestrian crashes), and crashes with these values in addition to a flagged traffic control device formed an even lower proportion of values (e.g., 0.7% of KABCO pedestrian crashes). Given this exceptionally low number of potentially misclassified crashes, particularly given the potential for misclassified crashes in the opposite fashion, the project team believed it was acceptable to include these crashes for further analysis.

The project team compared the proportional distribution of contributing circumstances between KA pedestrian crashes and crashes of all severities (KABCO). The following sections note observations from each analysis.

Pedestrian Crash Analysis Results

The project team compared the distribution of crash characteristics between KA and BCO mid-block pedestrian crashes to identify potential factors that are overrepresented (or underrepresented) in more severe pedestrian crash outcomes. Through this comparison, the project team made the following observations relevant to road segment-level risk factors:

- The most pronounced difference between KA and BCO pedestrian crashes is lighting. Fatal and serious injury pedestrian crashes tended to occur at night, in dark lighting conditions (46 percent of KA crashes as opposed to 29 percent of BCO crashes).
    - 59 municipalities in Massachusetts had above average frequency and rate of KABCO pedestrian crashes that occurred in dark lighting conditions; these municipalities met the following criteria:
        - At least three crashes that occurred in dark lighting conditions between 2013 and 2017.
        - The proportion of pedestrian crashes that occurred in dark lighting conditions is higher than the statewide proportion.
        - The rate of crashes that occurred in dark lighting conditions (per 1,000 residents) is higher than the statewide rate.
- KA pedestrian crashes typically occur on arterial and major collector classified roads; nearly 80 percent of all KA pedestrian crashes occurred on principal arterials, minor arterials, and major collectors.
    - KA crashes are significantly overrepresented on principal and minor arterials relative to both the statewide road mileage and total statewide vehicle miles traveled (VMT); crashes may be slightly overrepresented on major collectors relative to statewide VMT on those facilities (Figure 1).
    - KA crashes may be overrepresented on local roads as a proportion of statewide VMT, they are disproportionately infrequent when compared to statewide road mileage.

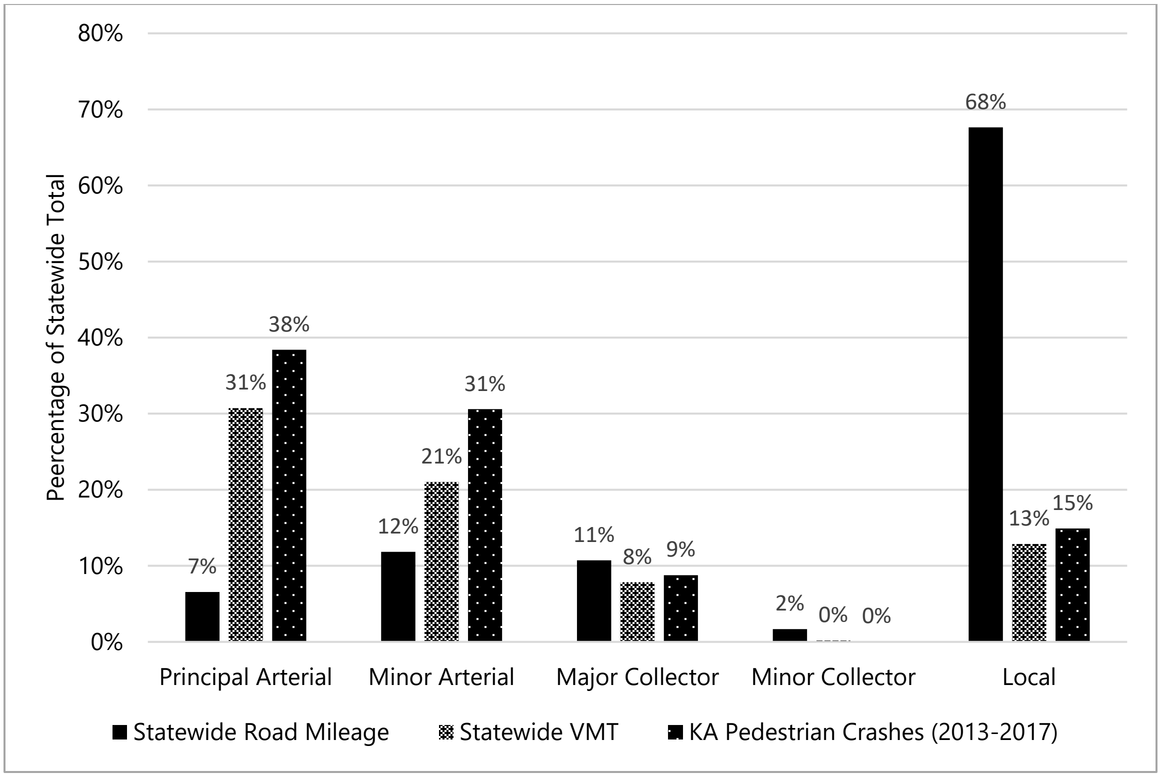

Figure 1. Proportion of statewide road mileage[1], statewide VMT[2], and KA pedestrian crashes by functional classification.

- The majority of KA pedestrian crashes did not have a speed limit reported on the crash report; however, of the KA pedestrian crashes with a known posted speed limit, the majority occurred on roads posted 35 mph and below (71%).
- The majority of KA pedestrian crashes occurred on two-lane, two-way, undivided roads.
- KA pedestrian crashes were overrepresented on roads with an annual average daily traffic (AADT) count of more than 9,000 vehicles per day. The AADT category groupings were based on categories developed for the Federal Highway Administration's (FHWA) *Guide for Improving Pedestrian Safety at Uncontrolled Crossing Locations.*[3]
- Most pedestrian KA crashes occurred during dry road conditions (78%).
- Pedestrians involved in KA crashes tended to be older than those in BCO crashes as a whole.
    - Of pedestrians with a known age, roughly 49 percent of pedestrians in KA crashes were 45 and older as opposed to 41 percent of pedestrians involved in BCO crashes.

In addition to crash contributing factors, the project team assessed the distribution of pedestrian crashes according to their proximity to a transit stop (bus and rail). For the purposes of this analysis, proximity was defined as the straight-line distance to the nearest transit stop; this distance does not necessarily reflect

---

[1] FHWA 2018 Highway Statistics; HM-20 Tables. https://www.fhwa.dot.gov/policyinformation/statistics/2018/
[2] FHWA 2018 Highway Statistics; VM-2 Table. https://www.fhwa.dot.gov/policyinformation/statistics/2018/
[3] https://safety.fhwa.dot.gov/ped_bike/step/docs/STEP_Guide_for_Improving_Ped_Safety_at_Unsig_Loc_3-2018_07_17-508compliant.pdf

the road or pedestrian network nor the shortest or most desirable path. Figure 2 illustrates that most KA mid-block pedestrian crashes occur within one mile of a transit stop.

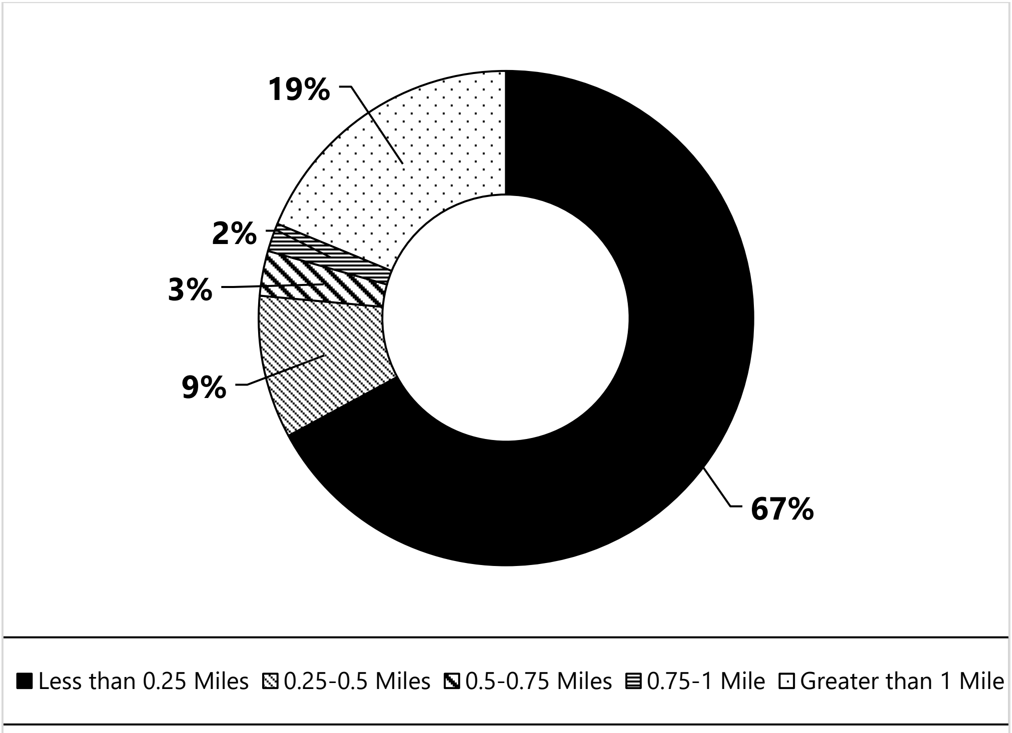

Figure 2. Distribution of KA midblock pedestrian crashes by distance to nearest transit stop.

Summary of Severity Comparisons

The comparison of historic crash characteristics across different severity categories revealed certain trends associated with higher severity pedestrian crashes. Pedestrians involved in severe crashes tended to be older than those involved in crashes of all severities, and severe pedestrian crashes skewed slightly more towards darker lighting conditions than pedestrian crashes of all severities. However, these risk factors do not necessarily allow MassDOT to target specific corridors for future improvements. Based on the distribution of crash data characteristics, roads that experience the majority of KA pedestrian crashes meet the following criteria:

Pedestrians

- Principal arterial, minor arterial, and/or a major collector.
- AADT is more than 9,000 vehicles per day.
- Two-way, undivided configuration.
- Two travel lanes, although 4- and 6-lane roads are overrepresented in the KA dataset.
- Posted speed limit of 35 mph or less.

Identifying High Risk Locations

While these characteristics may be indicative of high severe crash locations, it is difficult to assess priorities to specific segments that meet these criteria without a consistent measure of non-motorized exposure. As the May 28 report demonstrated, demographic and socioeconomic factors can be useful indicators of high traffic locations where MassDOT can proactively plan improvements where they are needed most. To underscore contextual risk factors on these priority facilities, VHB developed binary logistic regression models to predict the likelihood of a KA pedestrian crash on road segments in Massachusetts. The project team developed separate models for pedestrian crashes on principal arterials, minor arterials, and major collectors. The goal of these models is to identify key risk factors, both segment-based and contextual factors, for severe pedestrian crashes for systemic project screening in Massachusetts. The comparison of historic crash data demonstrated that these functional classifications experience a far greater share of KA or KAB non-motorized crashes than their share of statewide road mileage and vehicle traffic. By targeting specific characteristics of these facilities that indicate high risk environments, MassDOT can systemically prioritize locations for improvement.

Arterial and Major Collector Risk Factor Analysis

This section outlines the binary logistic modeling to develop functional classification-specific risk factors for severe pedestrian crashes.

Data

For roadway and traffic characteristics, VHB obtained MassDOT's roadway inventory from the MassDOT open data portal.[4] For contextual factors, VHB collected data from the same sources described in the May 28 report. In addition to the variables collected for the May 28 report, VHB developed two additional data elements at the request of MassDOT. These two elements are updated versions of the environmental justice (EJ) population indicators developed by the Massachusetts Bureau of Geographic Information (MassGIS).[5] MassGIS developed this geographic information systems (GIS) layer based on 2010 United States Census data for three indicators of high environmental justice need neighborhoods:

- **Proportion of non-white population**: Block groups with a proportion of non-white population greater than 25 percent are flagged in this category.
- **Limited English proficiency (LEP) households**: Block groups with a proportion of limited English-speaking households greater than 25 percent are flagged in this category.
- **Median household income**: Block groups with a median household income below $40,673 are flagged in this category.

VHB incorporated the MassGIS EJ data layer by identifying road segments that are located within block groups that have at least two of these three EJ flags. The final set of socioeconomic and demographic risk factor variables included:

- Number of employees per square mile (employment density).
- Number of residents per square mile (population density).
- Proportion of households without a motor vehicle.
- Proportion of commuters that walk, bike, or take transit.
- Proportion of employment in the accommodation, food services, or retail trades.
- Ratio of population living in poverty (relative to total population for which poverty status has been determined).
- Median household income.
- Two or more MassGIS EJ flags.
- Bus and rail stops per square mile (transit stop density).

---

[4] https://geo-massdot.opendata.arcgis.com/datasets/46bb709a682a4373b57dfa832f35ade6
[5] https://www.mass.gov/info-details/environmental-justice-populations-in-massachusetts

Based on the correlation between transit stop presence and non-motorized crashes observed in the previous section, VHB developed an additional measure of transit access for risk factor analysis:

- Transit stop presence (rail and/or bus) on a road segment.

Like the analysis summarized in the May 28 report, Boston city block groups were flagged and excluded from the analysis due to concerns with the completeness of crash record data within the Boston city limits.

Method

Due to the binary nature of the crash severity outcome of interest, the project team used binary logistic regression. This probabilistic modeling technique assesses the probability that an event has occurred (i.e., a KA pedestrian crash) on a given segment based on the model inputs. Agresti (2007) provides more background information on this method.[6] VHB obtained road segment data from MassDOT and separated the three functional classes mentioned in previous sections—principal arterials, minor arterials, and major collectors—into separate datasets. If a single KA pedestrian crash occurred on a given segment (e.g., within 25 feet as calculated in GIS) at any time between 2013 and 2017, VHB assigned that segment with a "1"; those segments without an observed midblock crash received a value of "0."

MassDOT's GIS road inventory contained all relevant roadway characteristics (e.g., number of lanes, posted speed limits, and traffic volumes). VHB spatially joined transit stops to this network in GIS, as well as all socioeconomic and neighborhood characteristics based on Census block group locations described in the previous section. Road segments that crossed block group boundaries (within 25 feet) were double counted, with an entry into the dataset for the characteristics of each block group the segment crossed. While the project team did not consider spatial effects in this modeling effort, repeating boundary road segments with values for both adjacent zones allowed the models to consider relevant factors that are in close proximity to a road segment, but are separated by a largely invisible boundary to pedestrians.

VHB normalized Census data based on block group values using the percentile rank function in Microsoft Excel. This allowed data to be categorized according to its value relative to the State as a whole, rather than an absolute measure. For instance, if the median income in a State ranges from $40,000/year to $200,000/year, the lowest value, $40,000, would receive a value of 0, while the highest value, $200,000, would receive a value of 1; in other words, zero percent of values in the State are below $40,000, while 100% of values in the State are below $200,000. If the median, median income was $90,000, then that block group would receive a value of 0.5 indicating that 50% of values are below $90,000/year. Figure 3 is a visual representation of this concept. This allows the model to assess the risk associated with being a high or low value, rather than assessing the risk associated with each additional $1 of income. This allows for a more direct comparison between different risk factors, as opposed to different units of measurement between each factor (e.g., % of households with zero vehicles vs. $ of income).

[6] Agresti, A. (2007). An Introduction to Categorical Data Analysis. Second Edition. John Wiley & Sons, Inc., New York.

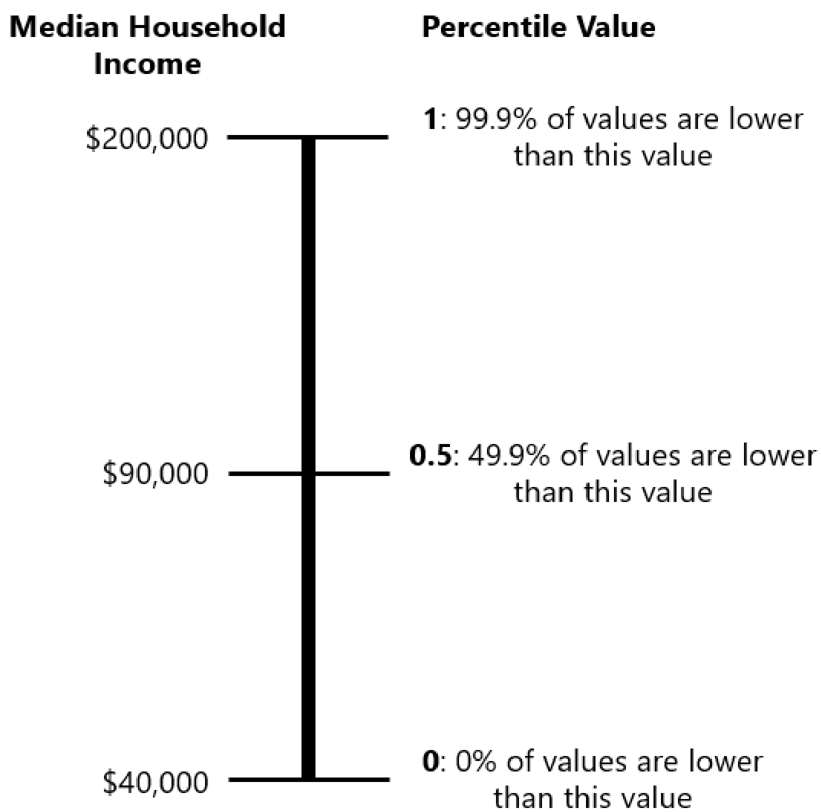

Figure 3. Percentile rank example – not actual data.

The following section reports the results of six models developed by VHB:

- KA pedestrian crashes on principal arterials.
- KA pedestrian crashes on minor arterials.
- KA pedestrian crashes on major collectors.

Results

This section reports the modeling results. The correlation of variables with the probability of pedestrian crashes is represented by odds ratios. An odds ratio greater than 1 indicates that an increase in that variable is associated with a higher probability of a crash occurring on that segment (all other things held equal), while an odds ratio of less than one indicates that a decrease in that variable is associated with a lower probability of a crash occurring on that segment. Each model includes the length of the segment to account for potential differences in the likelihood that a crash occurred on longer segments as opposed to shorter ones. As with all results reported in this report, all factors should be interpreted as having a correlation with KA pedestrian crashes; the causal relationship for any particular crash may not be captured in the following models.

Binary Logit Models for Pedestrian Crashes

The binary logit regression models for pedestrian crashes on principal arterials, minor arterials and major collectors are shown in Table 1, Table 2, and Table 3 respectively. All three probability models were developed using fatal and serious injury crashes.

Table 1. Binary Logit Model for Pedestrian KA Crashes on Principal Arterials.

Variable	Odds Ratio	Standard error	z-value	P>|z|	95% Confidence Interval	
3 or more travel lanes, both directions	1.72	0.12	7.63	<0.01	1.49	1.97
Presence of median	0.37	0.03	-11.00	<0.01	0.31	0.44
AADT over 15,000	1.53	0.10	6.55	<0.01	1.35	1.74
Segment length (miles)	4.28	0.51	12.08	<0.01	3.38	5.41
Transit stop presence (rail and/or bus) on road segment	1.88	0.14	8.58	<0.01	1.62	2.17
Two or more MassGIS EJ flags	1.24	0.11	2.47	0.01	1.04	1.46
Median household income	0.49	0.06	-5.88	<0.01	0.38	0.62
Transit stop density	1.90	0.23	5.18	<0.01	1.49	2.42
Proportion of employment in the accommodation, food services, or retail trades	1.45	0.16	3.30	<0.01	1.16	1.80
Employment density	2.65	0.39	6.66	<0.01	1.99	3.53
Population density	2.13	0.34	4.68	<0.01	1.55	2.92
Constant	0.004	0.0005	-43.80	<0.01	0.003	0.005

Note: Number of observations = 81,562; Log likelihood = -6005.2443; Pseudo R2 = 0.0913; LR chi2(11) = 1206.21; Prob > chi2 < 0.0000.

The binary logit model in Table 1 presents the pedestrian crash probability model for fatal and injury crashes on principal arterials in Massachusetts. The results show that the number of lanes (three or more in both directions) has a statistically significant correlation. The odds ratio is greater than one which indicates that the probability of at least one KA pedestrian crash occurring on a segment is higher for principal arterials with three or more lanes. This is consistent with engineering expectations, as more lanes and a wider travel way provide longer crossing distances (and greater exposure) for pedestrians. Another significant and positive correlation is the traffic volume of the roadway. According to Table 1, the odds of at least one KA pedestrian crash occurring on a principal arterial increase as the AADT rises above 15,000. Conversely, the presence of a median lowers the probability of a pedestrian crash relative to the observed traffic volume and number of lanes. The odds of at least one pedestrian crash occurring on a segment are significantly lower when there is a median present in the roadway. This is in line with engineering expectations and guidance, as medians provide shorter crossing distances for pedestrians and can provide refuge, depending on the type of median.

Several demographic and the socioeconomic variables are also statistically significant at 95% confidence level. Employment density, population density, and the proportion of employment in accommodation, food services, or retail trades are likely associated with higher pedestrian volumes, and therefore increase the likelihood of a crash occurring on a given segment. The two remaining variables, an indicator that a road segment is in a Census block group with two or more MassGIS EJ flags and median household income, suggest that there may be socioeconomic disparities with regards to pedestrian safety in Massachusetts. Median income is negatively associated with the probability of a KA pedestrian crash (i.e., lower incomes indicate a higher likelihood of a crash), which is highly consistent with the literature of pedestrian safety and neighborhood context.7 Alternatively, the EJ indicator is positively correlated with the probability of a KA pedestrian crash, meaning that a combination of low household income, high non-white population, and high levels of limited English-language proficiency indicate a higher likelihood of a crash occurring. While this is also highly consistent with past research on pedestrian safety and an important consideration in neighborhood planning and policy, MassDOT should consider how best to apply this information from an infrastructure investment standpoint. The Recommendations and Next Steps section of this report outline additional considerations for MassDOT.

Finally, there are two variables related to transit access (e.g., transit stop density and transit stop presence on a road segment), that both have a positive correlation with the probability of a KA pedestrian crash. This is also in-line with engineering and planning expectations, as the number of pedestrians is likely higher near any bus or rail transit stop due to boarding and alighting riders and the placement of stops near key community features and resources.

[7] https://www.transportation.gov/sites/dot.gov/files/docs/mission/office-policy/transportation-policy/328686/effects-roadway-and-built-environment-characteristics-pedestrian-fatality-risk-mansfield-et-al.pdf

Table 2. Binary logit model for pedestrian KA crashes on minor arterials.

Variable	Odds Ratio	Standard error	z-value	P>|z|	95% Confidence Interval	
3 or more travel lanes, both directions	1.81	0.23	4.65	<0.01	1.41	2.32
Presence of median	0.42	0.08	-4.46	<0.01	0.28	0.61
AADT over 9,000	1.47	0.10	5.64	<0.01	1.28	1.68
Segment length (miles)	34.62	8.42	14.57	<0.01	21.49	55.76
Transit stop presence (rail and/or bus) on road segment	1.73	0.15	6.27	<0.01	1.46	2.05
Two or more MassGIS EJ flags	1.41	0.13	3.71	<0.01	1.18	1.69
Median household income	0.32	0.05	-7.53	<0.01	0.23	0.43
Proportion of commuters that walk, bicycle, or take transit	1.28	0.18	1.75	0.08	0.97	1.68
Transit stop density	1.69	0.23	3.93	<0.01	1.30	2.20
Proportion of employment in the accommodation, food services, or retail trades	1.65	0.20	4.13	<0.01	1.30	2.10
Employment density	2.45	0.40	5.46	<0.01	1.77	3.37
Population density	3.32	0.63	6.33	<0.01	2.29	4.82
Constant	0.002	0.0002	-43.80	<0.01	0.001	0.002

Note: Number of observations = 130,844; Log likelihood = -5132.9286; Pseudo R2 = 0.1093; LR chi2(12) = 1259.88; Prob > chi2 < 0.0000.

Table 2 shows the binary logistic regression for KA pedestrian crashes on minor arterials. Like the results for principal arterials in Table 1, most of the results in Table 2 are highly consistent with planning and engineering expectations. The only substantial difference in this model involves the AADT threshold of 9,000 vehicles per day in the minor arterial model as opposed to 15,000 vehicles per day in the principal arterial model. The better performance of this lower threshold (e.g., higher odds ratio and improved model fit), indicates that traffic volumes on minor arterials tend to be lower, and so traffic volumes above a lower AADT threshold on minor arterials is positively associated with an increased likelihood of a pedestrian KA crash.

Table 3. Binary logit model for pedestrian KA crashes on major collectors.

Variable	Odds Ratio	Standard error	z-value	P>|z|	95% Confidence Interval	
AADT over 9,000	2.51	0.44	5.28	<0.01	1.78	3.53
Segment length (miles)	12.97	4.35	7.64	<0.01	6.72	25.03
Transit stop presence (rail and/or bus) on road segment	2.22	0.35	5.08	<0.01	1.63	3.02
Two or more MassGIS EJ flags	1.53	0.24	2.68	0.01	1.12	2.10
Median household income	0.68	0.15	-1.72	0.09	0.43	1.06
Employment density	4.21	1.04	5.84	<0.01	2.60	6.83
Population density	4.96	1.33	5.97	<0.01	2.93	8.39
Constant	0.0009	0.0002	-33.25	<0.01	0.0006	0.001

Note: Number of observations = 75,104; Log likelihood = -2036.4336; Pseudo R2 = 0.0839; LR chi2(7) = 372.86; Prob > chi2 < 0.0000.

Table 3 documents the results for the major collector model. Note that the relatively small amount of KA pedestrian crashes on major collectors during the study period (116) limits the number of variables that retained statistical significance. Traffic volume, represented by AADT over 9,000 vehicles per day, is the only roadway variable included in the major collector crash model. This is due to the lack of variation in the number of lanes and presence of a median on major collectors; 97% of major collectors have two travel lanes and 98% are undivided. Like principal and minor arterials, the demographic, socioeconomic, and transit access variables are all in line with engineering expectations.

Recommendations and Next Steps

As MassDOT considers the results of this analysis, the project team would like to note a few recommendations for application in systemic project screening and countermeasure implementation:

- The project team did not include street light presence as a systemic risk factor due to the lack of a lighting inventory that would cover the relevant facilities in the State.
- While the screening process for municipalities with an overrepresentation of pedestrian crashes in dark lighting conditions had three criteria, VHB recommends that roads in these 59 municipalities only receive one flag for dark lighting conditions.
- The project team investigated pedestrian infrastructure (on-street bike lanes and sidewalks, respectively) as a part of the binary logistic regression analysis. Both variables were associated with an increased likelihood of a KA pedestrian crash occurring on almost all functional classifications. While these variables were not included in the final models (Tables 1-3), the correlation between these infrastructure and non-motorized crashes is common in safety analyses; however, this does not necessarily reflect the safety effectiveness of these infrastructure. Pedestrians gravitate toward sidewalks, leading to generally higher exposure at these locations

than similar locations without these features. Furthermore, the binary logistic regression analysis does not consider the change in safety performance of these sites relative to the same locations without the infrastructure; these locations would be riskier to pedestrians if these features were not present. However, the results of this analysis indicate that MassDOT should not overlook locations simply because a sidewalk is already present.

- The project team did not include posted speed limit as a final factor in the systemic analysis for two reasons:
  - Posted speed limits produced results that did not necessarily reflect the speed of the moving vehicle (i.e., higher probabilities of a crash were associated with lower posted speed limits). This is likely not an indicator of lower speeds being more dangerous, but rather that posted speed limits reflect regulatory conditions. For instance, cities typically have lower posted speed limits than unincorporated areas, even for separate segments of the same road.
  - While probe speed measures or observed speeds from corridor studies could be useful tools for determining unsafe speeds for pedestrians at the site-level, posted speed limits were not appropriate for this statewide screening.
- Many of the neighborhood-level factors have a logical relationship to pedestrian and cyclist volumes; since there is no statewide count program for these modes, these factors are used as surrogates for exposure.
  - The project team recommends that MassDOT consider developing a statewide or regional non-motorized exposure model to fill gaps in existing count databases or probe data and better understand risk across the State.
  - This model could replace many of the surrogates identified in this report for risk identification.

Based on the results of the negative binomial model, VHB has identified several risk factors associated with severe pedestrian crashes. Table 4 categorizes these risk factors by their category (roadway or neighborhood factors) and prioritizes them in order of relative importance according to the model results. MassDOT has several options to apply these risk factors for network screening purposes:

- MassDOT could use the probabilistic model to directly score individual road segments. Each factor identified in this report can be applied to individual road segments, and the prioritization algorithm can calculate the probability for a particular crash type and severity based on the appropriate functional class model from Tables 1-6. However, VHB does not recommend this approach as the models incorporate segment length in the result. While it is intuitive to observe a higher likelihood of a crash on longer segments, this does not make a segment inherently riskier to pedestrians on the ground. Therefore, the project team recommends that MassDOT apply each variable as an individual risk factor to be summed on each road segment.

- As a more flexible screening approach, each road segment can be flagged with the relevant risk factors in GIS (e.g., a binary 0 or 1 classification). For continuous, non-binary data (e.g., median household income), the percentile values of each variable can allow MassDOT to easily categorize data inputs. For instance, percentiles can be combined into more discrete categories:

- Top 10% of block group values: 1
- 80-89th percentile of block group values: 0.9
- 70-79th percentile of block group values: 0.8
- 60-69th percentile of block group values: 0.7
- 50-59th percentile of block group values: 0.6
- 40-49th percentile of block group values: 0.5
- 30-39th percentile of block group values: 0.4
- 20-29th percentile of block group values: 0.3
- 10-19th percentile of block group values: 0.2
- Bottom 10% of block group values: 0.1

VHB recommends that MassDOT consider using all factors presented in Table 4 for segment scoring and prioritization.

Table 4. Pedestrian systemic risk factors by functional classification.

Category	Principal Arterials	Minor Arterials	Major Collectors
Roadway	1. Presence of a Median **(-)** 2. Transit stop presence on a road segment, rail and/or bus **(+)** 3. 3+ travel lanes in both directions of travel **(+)** 4. AADT over 15,000 **(+)**	1. Presence of a Median **(-)** 2. 3+ travel lanes in both directions of travel **(+)** 3. Transit stop presence on a road segment, rail and/or bus **(+)** 4. AADT over 9,000 **(+)**	1. AADT over 9,000 **(+)** 2. Transit stop presence on a road segment, rail and/or bus **(+)**
Neighborhood	1. Employment density **(+)** 2. Median household income **(-)** 3. Population density **(+)** 4. Transit stop density **(+)** 5. Ratio of employment in the accommodation, food services, or retail trades **(+)** 6. Two or more MassGIS EJ flags **(+)**	1. Median household income **(-)** 2. Population density **(+)** 3. Employment density **(+)** 4. Ratio of employment in the accommodation, food services, or retail trades **(+)** 5. Transit stop density **(+)** 6. Two or more MassGIS EJ flags **(+)** 7. Commuters that walk, bicycle, or take transit **(+)**	1. Population density **(+)** 2. Employment density **(+)** 3. Median household income **(-)** 4. Two or more MassGIS EJ flags **(+)**

(+) = Odds ratio >1

(-) = Odds ratio <1

MassDOT can calculate the risk score by assigning a point for every risk factor present on a segment. MassDOT can consider different weights for different risk factors or consider each factor equally. When selecting countermeasures and developing deployment plans, MassDOT should prioritize segments with the highest risk scores first. Once those sites have been treated, MassDOT should then proceed to developing plans to address sites with the next lowest scores, and so on. Table 5 shows the recommended scoring system for each risk factor, and Table 6 provides an example of how each relevant factor can be equally applied to generate a hypothetical pedestrian risk score for a principal arterial segment. The risk score can then be normalized by dividing by the total possible risk score, then multiplying by 100.

Table 5. Risk factor assessment example for a principal arterial.

Risk Factor	Risk Factor Scoring Category	Categories and Corresponding Score
Presence of a Median	Categorical	• Undivided road = 1 • Median is present = 0
Transit stop presence on a road segment, rail and/or bus	Categorical	• Transit stop is present = 1 • No transit stop present = 0
3+ travel lanes in both directions of travel	Categorical	• Road has 3 or more travel lanes = 1 • Road has 1 or 2 travel lanes = 0
AADT over 15,000	Categorical	• Observed AADT is 15,000 or higher = 1 • Observed AADT is less than 15,000 = 0
AADT over 9,000	Categorical	• Observed AADT is 9,000 or higher = 1 • Observed AADT is less than 9,000 = 0
No shoulder >4 feet on either side of the road segment	Categorical	• Both outside shoulders are less than 4 feet wide = 1 • At least one outside shoulder is wider than 4 feet = 0
Two or more MassGIS EJ flags	Categorical	• Segment is in a block group with at least 2 EJ flags = 1 • Segment is in a block group with 0 or 1 EJ flags = 0
Employment density	Continuous	• Top 10% of block group values: 1 • 80-89th percentile of block group values: 0.9 • 70-79th percentile of block group values: 0.8 • 60-69th percentile of block group values: 0.7 • 50-59th percentile of block group values: 0.6 • 40-49th percentile of block group values: 0.5 • 30-39th percentile of block group values: 0.4 • 20-29th percentile of block group values: 0.3 • 10-19th percentile of block group values: 0.2 • Bottom 10% of block group values: 0.1
Population density	Continuous	• Top 10% of block group values: 1 • 80-89th percentile of block group values: 0.9 • 70-79th percentile of block group values: 0.8 • 60-69th percentile of block group values: 0.7 • 50-59th percentile of block group values: 0.6 • 40-49th percentile of block group values: 0.5 • 30-39th percentile of block group values: 0.4 • 20-29th percentile of block group values: 0.3 • 10-19th percentile of block group values: 0.2 • Bottom 10% of block group values: 0.1
Median household income	Continuous	• Top 10% of block group values: 0.1 • 80-89th percentile of block group values: 0.2 • 70-79th percentile of block group values: 0.3 • 60-69th percentile of block group values: 0.4 • 50-59th percentile of block group values: 0.5 • 40-49th percentile of block group values: 0.6 • 30-39th percentile of block group values: 0.7 • 20-29th percentile of block group values: 0.8 • 10-19th percentile of block group values: 0.9 • Bottom 10% of block group values: 1

Risk Factor	Risk Factor Scoring Category	Categories and Corresponding Score
Transit stop density	Continuous	• Top 10% of block group values: 1 • 80-89th percentile of block group values: 0.9 • 70-79th percentile of block group values: 0.8 • 60-69th percentile of block group values: 0.7 • 50-59th percentile of block group values: 0.6 • 40-49th percentile of block group values: 0.5 • 30-39th percentile of block group values: 0.4 • 20-29th percentile of block group values: 0.3 • 10-19th percentile of block group values: 0.2 • Bottom 10% of block group values: 0.1
Ratio of employment in the accommodation, food services, or retail trades	Continuous	• Top 10% of block group values: 1 • 80-89th percentile of block group values: 0.9 • 70-79th percentile of block group values: 0.8 • 60-69th percentile of block group values: 0.7 • 50-59th percentile of block group values: 0.6 • 40-49th percentile of block group values: 0.5 • 30-39th percentile of block group values: 0.4 • 20-29th percentile of block group values: 0.3 • 10-19th percentile of block group values: 0.2 • Bottom 10% of block group values: 0.1
Commuters that walk, bicycle, or take transit	Continuous	• Top 10% of block group values: 1 • 80-89th percentile of block group values: 0.9 • 70-79th percentile of block group values: 0.8 • 60-69th percentile of block group values: 0.7 • 50-59th percentile of block group values: 0.6 • 40-49th percentile of block group values: 0.5 • 30-39th percentile of block group values: 0.4 • 20-29th percentile of block group values: 0.3 • 10-19th percentile of block group values: 0.2 • Bottom 10% of block group values: 0.1
Proportion of households without a motor vehicle	Continuous	• Top 10% of block group values: 1 • 80-89th percentile of block group values: 0.9 • 70-79th percentile of block group values: 0.8 • 60-69th percentile of block group values: 0.7 • 50-59th percentile of block group values: 0.6 • 40-49th percentile of block group values: 0.5 • 30-39th percentile of block group values: 0.4 • 20-29th percentile of block group values: 0.3 • 10-19th percentile of block group values: 0.2 • Bottom 10% of block group values: 0.1

Table 6. Risk factor assessment example for pedestrians on a principal arterial.

Risk Factor	Risk Factor Scoring Category	Segment Value	Risk Score
Presence of a Median	Categorical	Undivided	1
Transit stop presence on a road segment, rail and/or bus	Categorical	Transit stop is present	1
3+ travel lanes in both directions of travel	Categorical	2 lanes	0
AADT over 15,000	Categorical	13,500 AADT	0
Employment density	Continuous	0.65	0.7
Median household income	Continuous	0.32	0.7
Population density	Continuous	0.47	0.5
Transit stop density	Continuous	0.88	0.9
Ratio of employment in the accommodation, food services, or retail trades	Continuous	0.56	0.6
Two or more MassGIS EJ flags	Categorical	1 EJ flag	0
Total Risk Score			**5.4**
Normalized Risk Score			**54%**

In order to finalize the data, MassDOT dissolved the road inventory based on the risk factor inputs to generate uniform corridors. These corridors can be used to identify targeted safety improvement projects. Additionally, MassDOT identified the closest address geospatially to the beginning and end of each corridor as reference points. The addresses include the street number, street name, and town of the address. Note these are the closest addresses geospatially, so the reference address may not be on the same street as the corridor itself, and the beginning and end reference address may be the same. MassDOT continues to provide mileposts for MassDOT routes and encourages users to use both mileposts and address points as references.

The segments were then ranked at both the Statewide and MPO levels using the normalized risk score and the percentile of score ranking (rank kind equal to weak) function in ArcGIS. For each normalized risk score, a percentile rank for the given score was computed relative to all the normalized risk scores. If there are repeated occurrences of the same normalized risk score, then the percentile rank corresponds to values that are less than or equal to the given score. The advantage of the weak ranking approach is that it guarantees that the highest normalized score will receive a percentile rank of 100%. The risk categories were then determined using the computed ranks. For example, sites ranked in the top 5 percentile (95 through 100) were categorized as "Primary Risk Site," sites ranked in the next 10 percentile (85 through

95) were categorized as "Secondary Risk Site," and the remaining sites were not categorized. In instances where there are large repeated occurrences of the same normalized risk score, the percentage of segments computed for top 5% or next 10% may not be equal to 5 or 10%. This is a byproduct of the weak ranking approach used. Table 7 and 8 show the distribution of focus facility type segments with the normalized risk score (presented as percentages) across these categories for Statewide and MPO rankings, respectively.

Table 7. Statewide risk categories.

State	Risk Category	Minimum Normalized Risk Score Percentage	Maximum Normalized Risk Score Percentage	Number of Segments	Percent of Scored State Segments
MA	Primary Risk Site	69.09%	98.33%	4917	5.3%
	Secondary Risk Site	56.67%	69%	9066	9.8%

Table 8. MPO risk categories.

MPO	Risk Category	Minimum Normalized Risk Score Percentage	Maximum Normalized Risk Score Percentage	Number of Segments	Percent of Scored MPO Segments
Berkshire Regional Planning Commission	Primary Risk Site	60%	90%	166	5.16%
	Secondary Risk Site	46%	59%	331	10.28%
Boston Region MPO	Primary Risk Site	72.73%	97%	2002	5.57%
	Secondary Risk Site	61.82%	72%	3434	9.55%
Cape Cod Commission	Primary Risk Site	51.82%	81%	260	5.11%
	Secondary Risk Site	41.82%	51%	534	10.50%
Central Massachusetts Regional Planning Commission	Primary Risk Site	69.09%	98.33%	506	5.42%
	Secondary Risk Site	54%	69%	909	9.74%
Franklin Regional Council of Governments	Primary Risk Site	44.55%	65%	118	5.88%
	Secondary Risk Site	36.36%	43.64%	186	9.26%
Martha's Vineyard Commission	Primary Risk Site	45.45%	54.54%	16	6.99%
	Secondary Risk Site	40%	44.55%	20	8.73%
Merrimack Valley Planning Commission	Primary Risk Site	67.27%	82.73%	242	5.10%
	Secondary Risk Site	56.36%	65.45%	486	10.24%
Montachusett Regional Planning Commission	Primary Risk Site	54.55%	70%	254	5.93%
	Secondary Risk Site	41.82%	54%	390	9.11%
Nantucket Planning and Economic Development Commission	Primary Risk Site	59.09%	68.18%	24	9.16%
	Secondary Risk Site	50%	58.33%	26	9.92%

MPO	Risk Category	Minimum Normalized Risk Score Percentage	Maximum Normalized Risk Score Percentage	Number of Segments	Percent of Scored MPO Segments
Northern Middlesex Council of Governments	Primary Risk Site	66.36%	85.45%	208	5.09%
	Secondary Risk Site	54.55%	65.45%	413	10.11%
Pioneer Valley Planning Commission	Primary Risk Site	70%	96.67%	496	5.37%
	Secondary Risk Site	59.09%	69.09%	902	9.77%
Old Colony Planning Council	Primary Risk Site	65.45%	95%	296	5.17%
	Secondary Risk Site	51.82%	65%	618	10.80%
Southeastern Regional Planning and Economic Development District	Primary Risk Site	66.36%	84.55%	428	5.27%
	Secondary Risk Site	56.67%	66%	798	9.83%