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

Bicycle 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 memorandum summarizes the risk factor analysis performed for bicycle crashes. It also represents a model method which can be used throughout the SDI for analyses of infrastructure-based emphasis areas and exposure. Memos for other emphasis areas may describe different methods used to adapt to the needs of those areas.

This memo expands upon the preliminary analysis results summarized in a memo delivered to MassDOT on May 28, 2020. This preliminary analysis memo describes a zonal analysis that identified socioeconomic and demographic risk factors related to bicycle 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 bicycle 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 bicycle crash, nor did it consider bicycle 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 bicycle crashes on roadway segments in Massachusetts.

This memo is separated into two sections. The first section documents a comparison of contributing circumstances in fatal and serious injury bicycle 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 KAB bicycle crash on a given segment of road. The second section of this memo provides the results of this analysis and prioritizes individual risk factors for MassDOT's consideration. This memo recommends possible applications of risk factors identified in both analyses.

Crash Severity Comparisons

MassDOT provided VHB with bicycle 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 bicyclists 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, and crashes with these values in addition to a flagged traffic control device formed an even lower proportion of values. 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 KAB bicyclist crashes and crashes of all severities (KABCO) of each travel mode. The following sections note observations from each analysis.

Bicycle Crash Analysis Results

The project team compared the distribution of crash characteristics between KAB and CO mid-block bicycle crashes to identify potential factors that are overrepresented (or underrepresented) in more severe bicycle crash outcomes. The project team elected to use KAB crashes for comparison due to the very low number of KA bicycle crashes during the study period (246). Through this comparison, the project team made the following observations relevant to road segment-level risk factors:

- KAB bicycle crashes overwhelmingly occur on arterials and major collector streets. Roughly 83 percent of all KAB bicycle crashes occurred on roads classified as a principal arterial, minor arterial, or major collector, as shown in Figure 1. KAB bicycle crashes are highly overrepresented on these functional classifications.

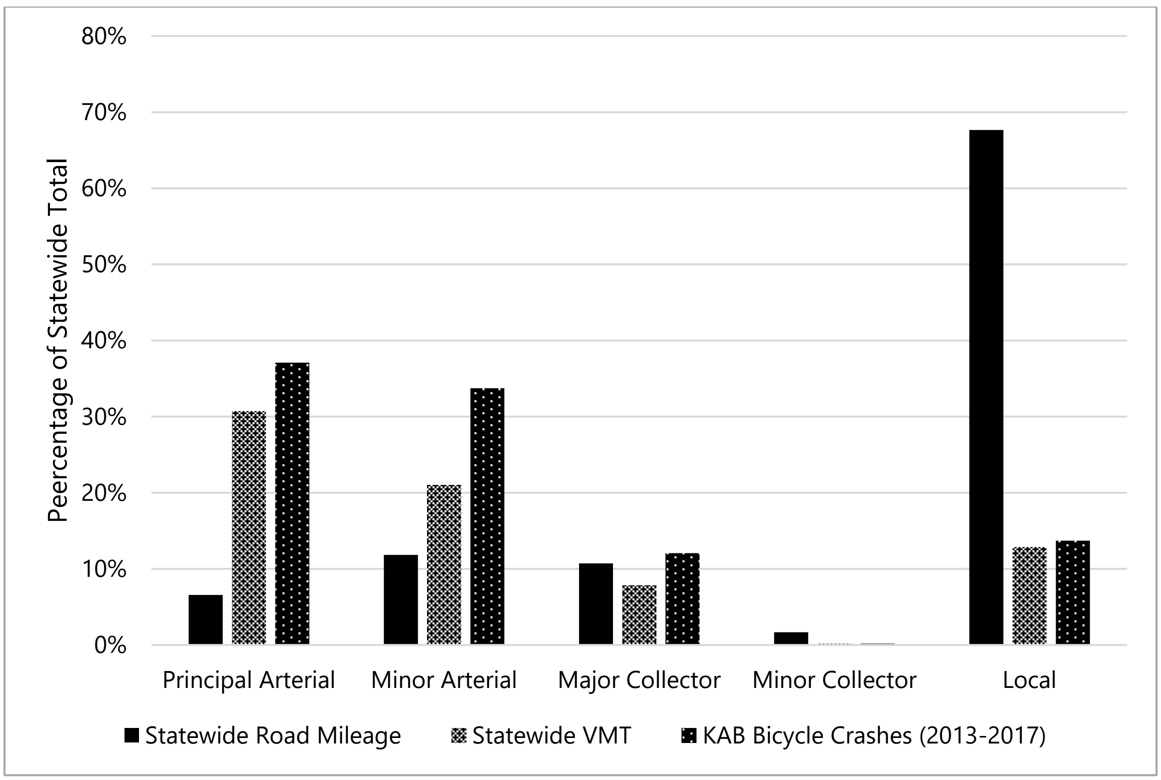

Figure 1. Distribution of statewide road mileage[1], vehicle miles traveled (VMT)[2], and KAB bicycle crashes by functional classification.

[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/

- KAB bicycle crashes tend to occur on roads with AADTs less than 15,000 vehicles per day.
    - This is likely due to cyclist exposure. Few bicyclists may be willing to travel on roads with a high number of vehicles.
- Most KAB bicyclist crashes occurred on two-lane, two-way, undivided roads (59%).
- KAB and CO crashes display a similar distribution within lighting conditions, as roughly 80 percent of both crash severity groupings occurred during daylight hours.
    - This may be related to exposure because bicyclists may generally try to avoid cycling at night if possible.
- Roughly 90 percent of KAB bicyclist crashes occurred on a dry road surface.
    - Bicyclists likely avoid cycling in bad weather (rain, snow etc.).
- Both KAB and CO bicycle crashes tended to be more evenly distributed across all age groups.

Most KAB mid-block bicycle crashes occurred within one mile of a transit stop (bus and rail), as shown in Figure 2.

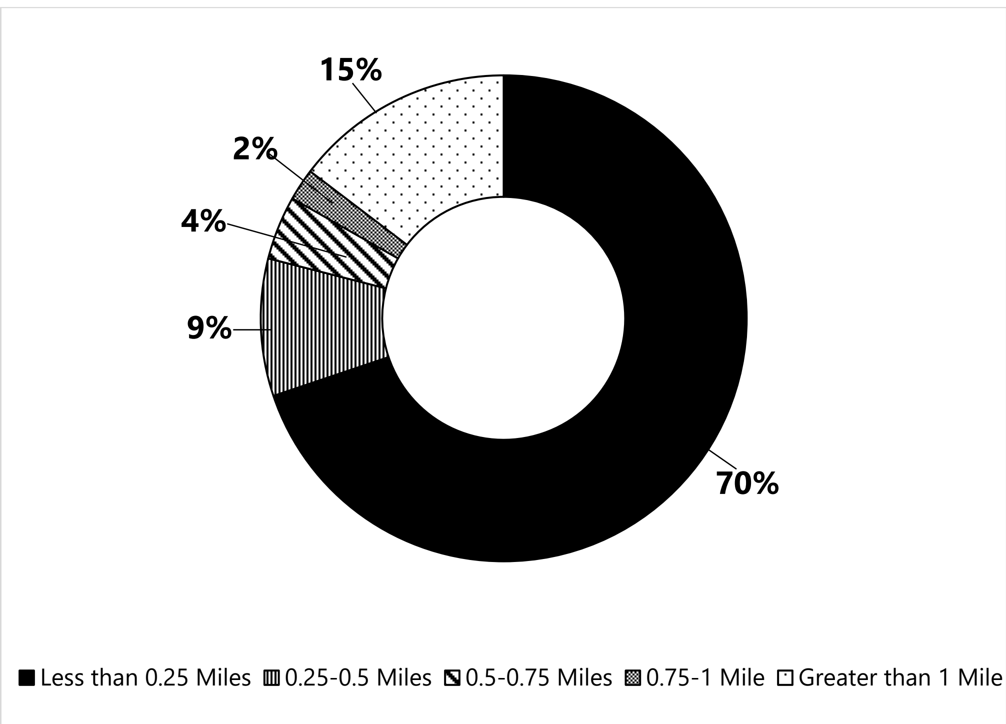

Figure 2. Distribution of KAB midblock bicycle crashes by distance to nearest transit stop.

Summary of Severity Comparisons

The comparison of historic crash characteristics across different severity categories revealed certain trends associated with higher severity bicycle crashes. Bicyclists involved in severe crashes tended to be older than those involved in 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 KAB bicycle crashes meet the following criteria:

Bicyclists

- Principal arterial, minor arterial, and/or a major collector.
- AADT is less than 15,000 vehicles per day.
- Two-way, undivided configuration.
- Two travel lanes.
- 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 memo 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 KAB bicycle crash on road segments in Massachusetts. The project team developed separate models for bicycle 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 bicycle 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 bicycle crashes.

Data

For roadway and traffic characteristics, VHB obtained MassDOT's roadway inventory from the MassDOT open data portal.[3] For contextual factors, VHB collected data from the same sources described in the May 28 memo. In addition to the variables collected for the May 28 memo, 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).[4] 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.

[3] https://geo-massdot.opendata.arcgis.com/datasets/46bb709a682a4373b57dfa832f35ade6

[4] https://www.mass.gov/info-details/environmental-justice-populations-in-massachusetts

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).

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 memo, 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 KAB bicycle crash) on a given segment based on the model inputs. Agresti (2007) provides more background information on this method.[5] 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 KAB bicycle 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 bicyclists.

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

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

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).

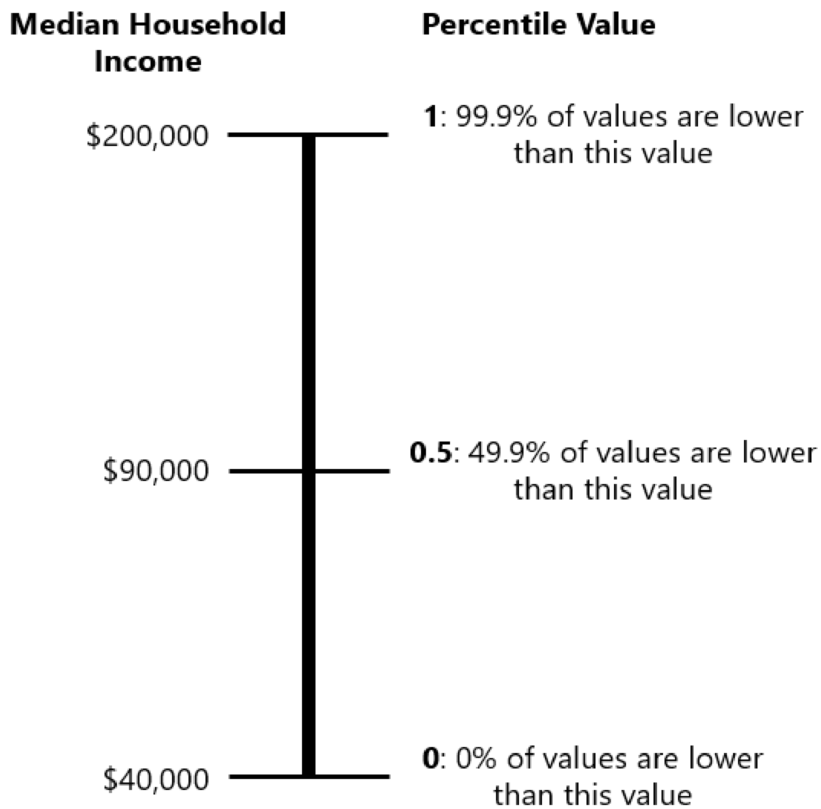

Figure 3. Percentile rank example – not actual data.

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

- KAB bicycle crashes on principal arterials.
- KAB bicycle crashes on minor arterials.
- KAB bicycle crashes on major collectors.

Results

This section reports the modeling results. The correlation of variables with the probability of bicycle 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 memo, all factors should be interpreted as having a correlation with KAB bicycle crashes; the causal relationship for any particular crash may not be captured in the following models.

Binary Logit Models for Bicycle Crashes

The binary logit regression models for bicycle crashes on principal arterials, minor arterials and major collectors are shown in Table 1, Table 2, and Table 3 respectively. As previously mentioned, all three bicycle crash probability models were developed using KAB crashes. This significantly expanded the sample size of segments that experienced at least one bicycle crash across all three functional classifications; this allowed the project team to be more confident that potential causes of bicyclist and motor vehicle have been captured by the sample.

Table 1. Binary logit model for bike KAB crashes on principal arterials.

Variable	Odds Ratio	Standard error	z-value	P>|z|	95% Confidence Interval	
3 or more travel lanes, both directions	0.92	0.07	-1.11	0.27	0.79	1.07
Presence of median	0.52	0.05	-6.82	<0.01	0.43	0.63
AADT over 15,000	1.30	0.08	4.12	<0.01	1.15	1.48
Segment length (miles)	4.45	0.64	10.41	<0.01	3.36	5.90
Transit stop presence (rail and/or bus) on road segment	1.83	0.13	8.73	<0.01	1.60	2.09
Median household income	1.87	0.23	5.14	<0.01	1.47	2.38
Proportion of commuters that walk, bicycle, or take transit	2.60	0.39	6.40	<0.01	1.94	3.49
Proportion of households without a motor vehicle	2.07	0.30	5.00	<0.01	1.56	2.75
Proportion of employment in the accommodation, food services, or retail trades	1.67	0.18	4.63	<0.01	1.34	2.07
Employment density	3.39	0.52	7.91	<0.01	2.50	4.58
Population density	5.57	0.96	9.98	<0.01	3.98	7.81
No shoulder wider than 4 feet on either side of the road segment	1.49	0.13	4.71	<0.01	1.26	1.75
Constant	0.0005	0.0001	-50.2	<0.01	0.0004	0.0007

Note: Number of observations = 81,562; Log likelihood = -5808.2097; Pseudo R2 = 0.1282; LR chi2(12) = 1707.65; Prob > chi2 < 0.0000.

Table 1 documents the results of the binary logistic regression for KAB bicycle crashes on principal arterials. For roadway-based variables, traffic volume and presence of a median are statistically significant indicators of a likelihood of a KAB bicycle crash. AADT greater than 15,000 vehicles per day is positively correlated with bicycle crashes, and the presence of a median is associated with a lower probability of a crash. While the former is in line with engineering expectations, the latter, median presence, may require more detailed investigation. It is possible that medians provide a separated space for bicyclists to ride or that medians help moderate vehicle travel speeds on roads where cyclists are more likely to ride. It is also possible that medians help restrict turning movements, particularly left turns in and out of commercial driveways that can reduce the number of conflicts that bicyclists have with motor vehicles. The project team included a number of lanes variable for MassDOT's reference, but no category for the number of lanes was significantly correlated with bicycle crashes. This is not surprising, as bicyclists traveling in midblock locations are likely on the shoulders of roads (or in a marked bicycle lane) and not crossing the road; therefore, the number of lanes is less important except as a proxy value for total AADT and vehicle speed.

The demographic and socioeconomic variables including households without a motor vehicle, commuters that walk, bicycle, or take transit, proportion of employment in the accommodation, food services, or retail trades, population density, and employment density have odds ratio greater than one. These are strong indicators of high non-motorized travel demand and volumes, and many of these indicators tend to be correlated with downtowns or central business districts. Household income was correlated with an increased likelihood of a bicycle KAB crash. This observation, combined with the significant correlation of transit stops with an increased likelihood of bicycle crashes, seems to indicate an association of bicycle crashes with downtowns and central business districts that have a high level of access for several different travel modes, including walking, biking, and taking transit. These areas may pose a higher risk to cyclists due to greater exposure, especially relative to less accessible areas of the State.

Table 2. Binary logit model for bike KAB crashes on minor arterials.

Variable	Odds Ratio	Standard error	z-value	P>|z|	95% Confidence Interval	
3 or more travel lanes, both directions	1.26	0.16	1.77	0.08	0.98	1.63
AADT over 9,000	1.33	0.08	4.79	<0.01	1.19	1.50
Presence of median	0.75	0.13	-1.72	0.09	0.54	1.04
Segment length (miles)	31.76	7.46	14.73	<0.01	20.04	50.32
Transit stop presence (rail and/or bus) on road segment	1.71	0.13	7.18	<0.01	1.48	1.99
Two or more MassGIS EJ flags	1.16	0.10	1.63	0.10	0.97	1.38
Median household income	1.60	0.22	3.36	<0.01	1.22	2.11
Commuters that walk, bicycle, or take transit	2.21	0.30	5.78	<0.01	1.69	2.88
Proportion of households without a motor vehicle	2.29	0.35	5.51	<0.01	1.71	3.08
Proportion of employment in the accommodation, food services, or retail trades	1.56	0.17	4.23	<0.01	1.27	1.92
Employment density	2.07	0.30	5.05	<0.01	1.56	2.74
Population density	8.53	1.40	13.03	<0.01	6.18	11.78
No shoulder wider than 4 feet on either side of the road segment	1.33	0.16	2.30	0.02	1.04	1.69
Constant	0.0004	0.0001	-44.65	<0.01	0.0003	0.0006

Note: Number of observations = 130,844; Log likelihood = -6193.2282; Pseudo R2 = 0.1172; LR chi2(13) = 1645.11; Prob > chi2 < 0.0000.

Table 2 shows the results of the binary logit models of KAB bicycle crashes on minor arterials. Like principal arterials, minor arterials in highly accessible, multimodal neighborhoods seem to be associated with a higher risk of a KAB bicycle crash. For roadway characteristics, only traffic volume and presence of a median was a significantly associated with an increased likelihood of a crash at the p = 0.05 confidence level; the number of travel lanes is a marginally statistically significant factor, indicating that road segments with three or more travel lanes are correlated with an increased likelihood of a KAB bicycle crash.

Table 3. Binary logit model for bike KAB crashes on major collectors.

Variable	Odds Ratio	Standard error	z-value	P>|z|	95% Confidence Interval	
AADT over 9,000	2.14	0.33	5.03	<0.01	1.59	2.89
Segment length (miles)	20.32	5.34	11.47	<0.01	12.14	33.99
Transit stop presence (rail and/or bus) on road segment	1.30	0.20	1.74	0.08	0.97	1.75
Two or more MassGIS EJ flags	1.53	0.18	3.67	<0.01	1.22	1.91
Commuters that walk, bicycle, or take transit	2.59	0.53	4.62	<0.01	1.73	3.88
Population density	5.06	1.21	6.78	<0.01	3.17	8.08
Employment density	4.33	0.89	7.14	<0.01	2.90	6.47
Constant	0.0007	0.0001	-52.73	<0.01	0.0005	0.0009

Note: Number of observations = 76,819; Log likelihood = -2823.1433; Pseudo R2 = 0.0968; LR chi2(7) = 605.33; Prob > chi2 < 0.0000.

Table 3 presents the binary logistic regression for KAB bicycle crashes on major collectors. Like principal and minor arterials, major collectors in highly accessible, multimodal neighborhoods seem to indicate a higher risk of a KAB bicycle crash. As noted in Table 3, there is not any substantial variation in the roadway configuration of major collectors, and so traffic volume is the only roadway variable left in the model.

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.
- The project team investigated bicycle 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 KAB bicycle 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. Bicyclists gravitate toward marked bicycle lanes, 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 cyclists if these features were not present. However, the results of this analysis indicate that MassDOT should not overlook locations simply because a marked bicycle lane 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 bicyclists 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 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 memo for risk identification.

Based on the results of the negative binomial model, VHB has identified several risk factors associated with severe bicycle 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 memo 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 bicyclists 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. Bicycle 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. No shoulder >4 feet on either side of the road segment **(+)** 4. AADT over 15,000 **(+)**	1. Presence of a Median **(-)** 2. Transit stop presence on a road segment, rail and/or bus **(+)** 3. AADT over 9,000 **(+)** 4. No shoulder >4 feet on either side of the road segment **(+)** 5. 3+ travel lanes in both directions of travel **(+)**	1. AADT over 9,000 **(+)** 2. Transit stop presence on a road segment, rail and/or bus **(+)**
Neighborhood	1. Population density **(+)** 2. Employment density **(+)** 3. Commuters that walk, bicycle, or take transit **(+)** 4. Proportion of households without a motor vehicle **(+)** 5. Median household income **(+)** 6. Ratio of employment in the accommodation, food services, or retail trades **(+)**	1. Population density **(+)** 2. Proportion of households without a motor vehicle **(+)** 3. Commuters that walk, bicycle, or take transit **(+)** 4. Employment density **(+)** 5. Median household income **(+)** 6. Two or more MassGIS EJ flags **(+)** 7. Ratio of employment in the accommodation, food services, or retail trades **(+)**	1. Population density **(+)** 2. Employment density **(+)** 3. Commuters that walk, bicycle, or take transit **(+)** 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. 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

Risk Factor	Risk Factor Scoring Category	Categories and Corresponding Score
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
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

Risk Factor	Risk Factor Scoring Category	Categories and Corresponding Score
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

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 6 and 7 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 6. 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	76.67%	100%	4696	5.4%
	Secondary Risk Site	63.33%	76.00%	9255	10.7%

Table 7. 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%	93.33%	211	7.21%
	Secondary Risk Site	50%	59%	229	7.83%
Boston Region MPO	Primary Risk Site	81.67%	100%	1753	5.24%
	Secondary Risk Site	71.67%	81%	3367	10.06%
Cape Cod Commission	Primary Risk Site	54.17%	79%	245	5.15%
	Secondary Risk Site	45%	54%	498	10.47%
Central Massachusetts Regional Planning Commission	Primary Risk Site	74%	96.67%	450	5.17%
	Secondary Risk Site	58.33%	73.33%	882	10.13%
Franklin Regional Council of Governments	Primary Risk Site	47.5%	68.33%	93	5.07%
	Secondary Risk Site	34.17%	47%	218	11.87%
Martha's Vineyard Commission	Primary Risk Site	48.33%	55.83%	13	6.07%
	Secondary Risk Site	45%	47.5%	20	9.35%
Merrimack Valley Planning Commission	Primary Risk Site	72.5%	93.33%	237	5.31%
	Secondary Risk Site	59.17%	71.67%	443	9.92%
Montachusett Regional Planning Commission	Primary Risk Site	56.67%	85.83%	229	5.82%
	Secondary Risk Site	45.83%	55%	374	9.50%
Nantucket Planning and Economic Development Commission	Primary Risk Site	55%	68.33%	22	9.09%
	Secondary Risk Site	51.67%	51.67%	31	12.81%
Northern Middlesex Council of Governments	Primary Risk Site	70%	95%	212	5.43%
	Secondary Risk Site	58.33%	69.17%	399	10.22%
Pioneer Valley Planning Commission	Primary Risk Site	75.83%	93.33%	436	5.05%
	Secondary Risk Site	65%	75%	924	10.70%
Old Colony Planning Council	Primary Risk Site	68.33%	95%	287	5.30%
	Secondary Risk Site	55%	67.5%	595	10.99%
Southeastern Regional Planning and Economic Development District	Primary Risk Site	70%	96.67%	421	5.48%
	Secondary Risk Site	55%	69.17%	782	10.18%