Updates to Risk Factors for SHSP Emphasis Areas

Lane Departure Crashes

PREPARED FOR

Massachusetts Department of Transportation

PREPARED BY

REPORT DATE: November 2023

UPDATED: January 23, 2024

Purpose & Background

The Massachusetts Department of Transportation (MassDOT) is updating the risk-based network screening maps in the IMPACT tool to incorporate recent crash data and build on lessons learned from previous analyses. This document describes the updated systemic analysis performed by VHB for lane departure crashes using crash data from 2017 through 2021. For this analysis, VHB used the default "Lane Departure" query[1] in the MassDOT IMPACT tool. The definition lists several "First Vehicle Event" conditions which qualify a crash for this query including "Collision with curb", "Collision with tree", "Collision with utility pole", "Collision with light pole or other post/support", "Collision with guardrail", "Collision with highway traffic sign post", "Collision with fence", "Collision with mail box", "Collision with impact attenuator/crash cushion", "Collision with bridge", "Collision with other fixed object (wall, building, tunnel, etc.)", "Collision with unknown fixed object", "Ran off the road right", " Ran off the road left", or "Cross median/centerline".[2] VHB then queried down to crashes which were not reported as an intersection per the "Roadway Junction Type" field.

Data Analysis and Focus Crash Types

To establish context, VHB first used the MassDOT IMPACT "Test of Proportions" tool[3] to summarize fatal injury (K) and suspected serious injury (A) lane departure crashes. To identify overrepresented crash attributes, VHB compared KA lane departure segment crashes to all KA segment crashes in the State. Where the proportion for a given attribute is statistically larger than the proportion for the comparison group, that attribute is flagged as a potential area of consideration. Statistical overrepresentation is checked by building 95 percent confidence intervals around the proportion using sampling errors. Figure 1 and Figure 2 show how the lower and upper bounds, respectively, are calculated based on the proportion of crashes (*p*) and the number of crashes in the sample (*N*). If the lower bound of lane departure KA crashes is larger than the upper bound of the comparison group, the attribute was considered "overrepresented" for the data.

$$95\%\ Confidence\ Interval, Lower\ Bound\ =\ p - 1.96 * \sqrt{\frac{p(1-p)}{N}}$$

Figure 1. Calculation of the lower bound of the 95 percent confidence interval for the proportion of crashes with an attribute.

$$95\%\ Confidence\ Interval, Upper\ Bound\ =\ p + 1.96 * \sqrt{\frac{p(1-p)}{N}}$$

Figure 2. Calculation of the upper bound of the 95 percent confidence interval for the proportion of crashes with an attribute.

Table 1 summarizes notable overrepresentations found in the analysis. VHB included the following data elements in their analysis:

---

[1] https://www.mass.gov/info-details/impact-emphasis-area-definitions

[2] MassDOT. *Impact Emphasis Area Definitions.* Available at: https://www.mass.gov/info-details/impact-emphasis-area-definitions. Accessed March 2023.

[3] https://apps.impact.dot.state.ma.us/sat/TestofProportions

- Access Control.
- Age of Driver – Oldest known.
- Age of Driver – Youngest Known.
- Age of Non-Motorist – Oldest Known.
- Age of Non-Motorist – Youngest Known.
- County Name.
- Crash Day of Week.
- Crash Month.
- Curb.
- Driver Contributing Circumstances.
- Driver Distracted By.
- Facility Type.
- Federal Functional Class.
- First Harmful Event.
- First Harmful Event Location.
- FMCSA Reportable.
- Functional Class.
- Jurisdiction.
- Left Shoulder Type-linked.
- Left Shoulder Width-linked.
- Light Conditions.
- Manner of Collision.
- Max Injury Severity Reported.
- Median Type.
- Operation.
- Opposite Number of Travel Lanes.
- Right Shoulder Type-linked.
- Right Shoulder Width-linked.
- Road Contributing Circumstance.
- Road Surface Condition.
- Roadway Junction Type.
- Speed Limit.

- Terrain Type.
- Total Lanes.
- Traffic Control Device Type.
- Trafficway Description.
- Urban Type.
- Weather Conditions.

Table 1. Summary of Key Overrepresentation Findings

Crash Field	Crash Attribute	Percent of Lane Departure KA Crashes	Percent of All KA Crashes
Age of Driver – Oldest Known	18-34	43.3%	30.6%
County Name	Hampshire County	3.3%	2.3%
Weekday	Sunday	18.0%	15.5%
	Saturday	19.3%	16.7%
Curb	No Curb	60.8%	47.9%
Driver Contributing Circumstances	Driving too fast for conditions	5.4%	3.0%
	Exceeded authorized speed limit	8.0%	4.7%
	Failure to keep in proper lane or running off road	12.8%	8.0%
	Fatigued/asleep	4.0%	1.7%
	History heart/epilepsy/fainting	1.4%	0.8%
	Illness	2.8%	1.6%
	Operating vehicle in erratic, reckless, careless, negligent or aggressive manner	16.2%	10.2%
	Over-correcting/over-steering	2.5%	1.4%
	Physical impairment	3.9%	2.1%
Driver Distracted By	Other activity (searching, eating, personal hygiene, etc.)	4.8%	3.3%
Federal Functional Class	Local	17.1%	14.0%
	Major Collector	14.7%	10.7%
	Minor Collector	27.2%	0.7%
First Harmful Event	Collision with curb	6.7%	2.6%
	Collision with guardrail	12.4%	5.4%
	Collision with other light pole or other post/support	4.0%	1.7%
	Collision with tree	27.3%	10.7%
	Collision with unknown fixed object	6.0%	2.4%
	Collision with utility pole	18.1%	6.9%
Jurisdiction	City or Town accepted road	59.6%	55.1%
Light Conditions	Dark – roadway not lighted	18.3%	13.4%
Manner of Collision	Single vehicle crash	85.1%	50.8%
Right Shoulder Type – Linked	Stable – Unruttable compacted subgrade	17.6%	11.4%
	Unstable shoulder	9.2%	6.1%
Road Contributing Circumstance	Road surface condition (wet, icy, snow, slush, etc.)	7.2%	5.2%
Road Surface Condition	Ice	2.3%	1.5%
Total Lanes	2	68.2%	64.1%
Trafficway Description	Two-way, not divided	63.5%	59.2%
Urban Type	Rural	7.8%	5.1%

From a safety management perspective, it is notable that overrepresented features predominantly describe two-lane, undivided, rural highways of low functional classes. Additionally, undesirable driver behaviors are overrepresented, including speeding and aggressive driving, distracted driving, and health issues (fatigue, illness, health history). Interestingly, though likely correlated with poor driving behaviors, severe lane departure crashes are overrepresented on weekends (Saturday and Sunday).

MassDOT should consider these findings when identifying potential lane departure countermeasures. The Federal Highway Administration (FHWA)'s *Proven Countermeasures*[4] includes several effective strategies following three objectives that include keep vehicles in the lane, reduce the potential for crashes if vehicles do leave their lane, and minimize severity if a crash does happen. The strategies include improving curve delineation, installing edgeline, installing rumble strips, maintaining clear zones, and installing safety edge.

While these are notable results, they should not restrict the analysis from focusing on all lane departure crashes. Ultimately, the focus crash type for this analysis is all lane departure crashes on segments.

[4] https://highways.dot.gov/safety/rwd/forrrwd/proven-countermeasures

Crash Tree and Focus Facility Type

After concluding that the lane departure focus crash type should include all segment crashes, VHB developed a crash tree to identify the roadway conditions under which severe lane departure crashes tend to occur most often. Figure 3 shows the crash tree.

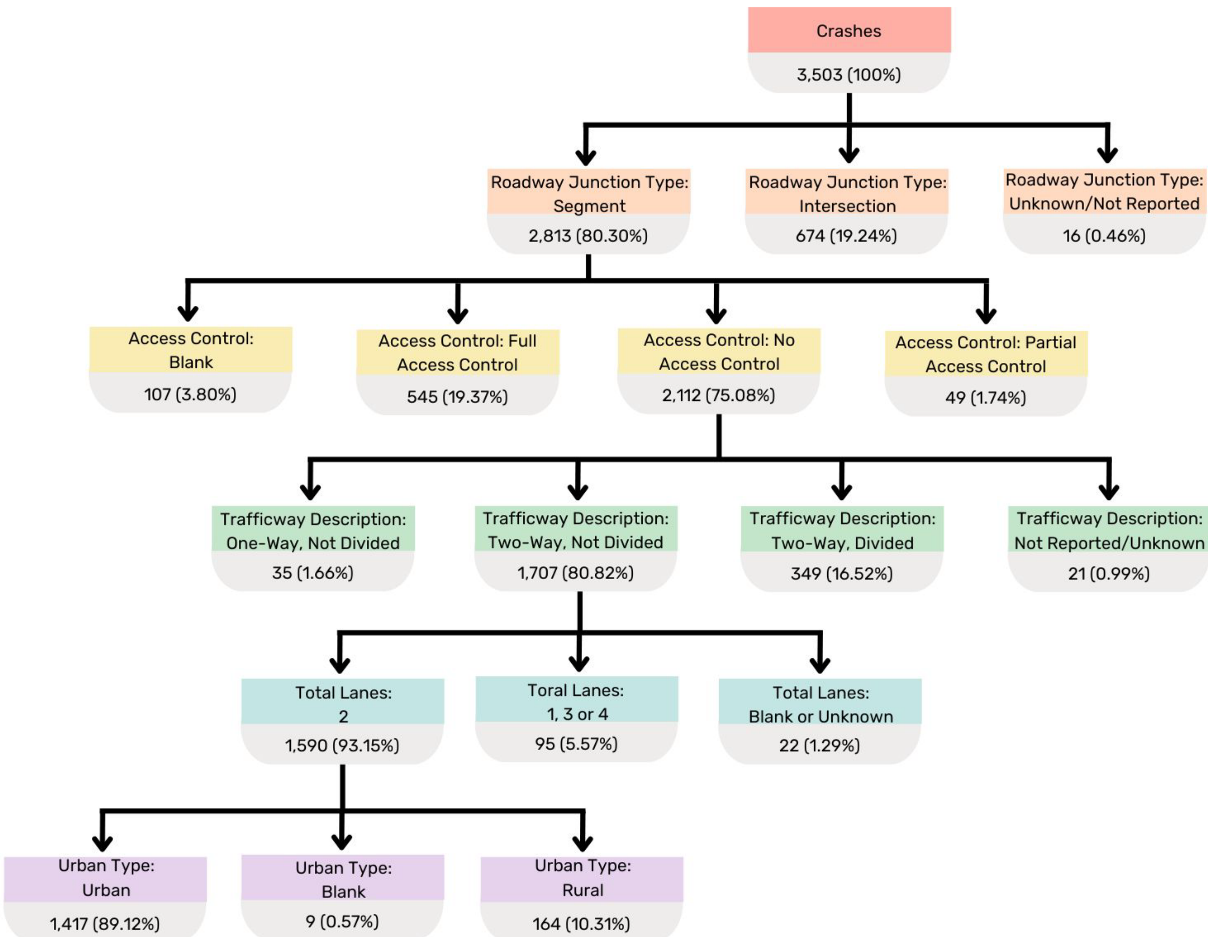

Figure 3. Crash tree summarizing KA lane departure crashes in Massachusetts.

It is evident that the majority of KA crashes related to lane departure were on segments with no access control. Of these crashes, two-lane undivided two-way roads experienced the highest proportion of severe lane departure crashes. Based on the crash tree in Figure 3, VHB recommends the following focus facility types:

- Urban two-lane, two-way undivided highways.
- Rural two-lane, two-way undivided highways.

Risk Factor Analysis

After identifying focus crash types and trends, VHB proceeded with the risk factor analysis. The following sections describe the methodology, data, and results of this analysis.

Methodology

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 lane departure) 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 identified the segments which fit into the focus facility characteristics. If a single KA lane departure crash occurred on a given segment (e.g., within 15 feet as calculated in GIS) at any time between 2017 and 2021, VHB assigned that segment with a "1"; those segments without an observed crash received a value of "0."

When modeling, VHB began with road exposure variables and added additional variables one at a time, monitoring the coefficients to ensure the inclusion of a variable did not result in large changes in magnitude. Additionally, VHB included variables with p-values upwards of 0.25 assuming the magnitude of the results made sense. VHB did not select a strict level of significance, as Hauer notes this could lead to misunderstanding or outright disregard for potentially noteworthy results.[6] The model estimates coefficients for each independent variable which are used to calculate Odds Ratios. An Odds Ratio greater than 1.0 indicates a positive correlation between the variable and the probability of a crash; an Odds Ratio less than 1.0 indicates a negative correlation between the variable and the probability of a crash.

Data

VHB used ArcGIS to manage and integrate data for this analysis. VHB aggregated data at the segment level. For data sources at larger levels (e.g., census block, town, MPO, or county), VHB used spatial data tools to join those data to segments where the center of the segment fell within the geographic polygon. VHB used the "Dissolve" spatial tool to join segments with similar roadway characteristics. Due to limitations with crash data acquisition, VHB excluded segments in the City of Boston from the analysis. Additionally, to avoid modeling complications, VHB only included segments at least 0.05 miles in length for analysis. MassDOT provided VHB with various sources of data, as described in the following sections.

Roadway Data

VHB downloaded the Massachusetts statewide Road Inventory 2020 file, available at https://geo-massdot.opendata.arcgis.com/datasets/342e8400ba3340c1bf5bf2b429ad8294/about. Based on discussions with MassDOT, VHB filtered the roadway data in ArcGIS using mileage counted (equal to 1), jurisdiction (not equal to null), and facility type (less than 7) to identify unique segments that were counted for the Highway Performance Monitoring System (HPMS). Filtering the roadway inventory in this way prevented potential double-counting of mileage and VMT for divided roads and roads with overlapping route numbers. MassDOT provided VHB with updated traffic volume data, which VHB integrated using GIS. VHB somewhat simplified the roadway data by dissolving on common roadway characteristics, including route and street name, town, surface width, shoulder width and type, presence of curbing, traffic volume, etc. Finally, VHB integrated vertical and horizontal geometric data, using the Identity function in GIS to distribute the segments by Horizontal Curve Class and Vertical Grade Class.

---

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

[6] Hauer, E. (2004). The harm done by tests of significance. *Accident Analysis & Prevention*, *36*(3), 495-500.

Crash Data

VHB obtained geolocated KA crashes on segments using the MassDOT IMPACT Query and Visualization (Q&V) Tool for the years 2017-2021. VHB then obtained the vehicle-level data for those crashes from the Q&V tool and queried those data to identify any vehicle where the first event matched MassDOT's Lane Departure query. Finally, VHB flagged any crash involving a vehicle that was found to meet the query – these became the KA lane departure sample. VHB used ArcGIS spatial tools to assign crashes to the nearest segment within 15 feet. VHB then used this join to flag whether a focus crash had occurred on the segment or not.

Driver License Data

MassDOT provided driver license data by age, town, and zip code for 2021. VHB used spatial analysis to assign driver license zip codes to the relevant town, joining the driver license totals by age. VHB then assigned relevant driver's license data to each segment by joining on the town name.

School Location Data

VHB obtained primary and secondary school location data from the Massachusetts Bureau of Geographic Information (MassGIS) open data portal (https://www.mass.gov/info-details/massgis-data-massachusetts-schools-pre-k-through-high-school). VHB identified if any schools were present within several geographical boundaries of each segment.

College and University Data

VHB accessed college and university location data from the MassGIS open data portal (https://www.mass.gov/info-details/massgis-data-colleges-and-universities). Although these data contain several categories of trade schools and other atypical technical training institutions, VHB only included "Colleges, universities, and professional schools," "Fine arts schools," "Junior colleges," and "Other technical and trade schools" for the purposes of this analysis. VHB identified if any schools were present within several geographical boundaries of each segment.

Citation Data

VHB obtained traffic citation count data by town for a five-year period between 2015 and 2019. These data included total citations, as well as subsets of counts for speeding-, seat belt-, impaired driving-, and distraction-related traffic citations. VHB then assigned relevant citation data to each segment by joining on the town name.

Horizontal and Vertical Alignment Data

MassDOT provided horizontal and vertical alignment data which included horizontal curve radius, vertical grade, and classifications of those features. Rather than integrating these data into the modeling dataset, VHB chose to use the "Overrepresentation" approach to assessing the risk of the focus crash type related to these characteristics. As such, VHB identified the alignment data components which fell along focus facility segments. Further, VHB spatially joined focus crashes to the horizontal and vertical segment elements within 15 feet of the crash. For scoring, VHB assigned the most severe feature present along the segment as the relevant risk feature.

The curve and grade data were broken into classes based on degree of curvature and percent absolute grade. Table 2 summarizes the ranges of degree of curvature and percent absolute grade for each class. When integrating the data for scoring, VHB found the most severe curve class and most severe grade class along each dissolved segment and assigned this attribute to the segment.

Table 2. Summary of Curve and Grade Classes

Curve Class	Degree of Curvature	Grade Class	Percent Absolute Grade
A	Less than 3.5 degrees	A	0.0 percent to 0.4 percent
B	3.5 degrees to 5.4 degrees	B	0.5 percent to 2.4 percent
C	5.5 degrees to 8.4 degrees	C	2.5 percent to 4.4 percent
D	8.5 degrees to 13.9 degrees	D	4.5 percent to 6.4 percent
E	14.0 degrees to 27.9 degrees	E	6.5 percent to 8.4 percent
F	28 degrees or more	F	8.5 percent or grater

Additional Data

VHB obtained several additional data sources for integration into the data set, including census and American Community Survey (ACS) data, public health data from the Massachusetts Department of Public Health (DPH), seatbelt use survey data at the county level, and environmental justice (EJ) data provided by MassDOT. These were joined to segments that fell within the relevant geographical polygon.

Modeling Data Sets

The two focus facility types dictated a minimum of two modeling data sets – rural and urban segments. Additionally, after consideration of the role of AADT in crash prediction modeling, VHB split the rural and urban segments into two further datasets based on whether the segment has a known AADT value, as opposed to unknown or default values:

- Urban segments with AADT.
- Urban segments without AADT.
- Rural segments with AADT.
- Rural segments without AADT.

Note that the segments with known AADT are predominantly MassDOT owned roads, while those without AADT are predominantly local roads.

Results

The following sections describe the results of the binary logistic regression modeling effort. For each dataset, VHB established a base model that included the natural log of the mile years (i.e., the product of five years of data and the length of the segment in miles) – this accounts for exposure. Additionally, VHB added the natural log of segment AADT for the models of AADT segments. Before including additional variables in the model, VHB developed a correlation matrix of input variables. Highly correlated variables are indicators of potential complications in the model development process. The following subsections include correlation matrices for each model.

Urban Segments with Known AADT

The binary logistic regression model for urban focus facility segments with known AADT is summarized in Table 3. As expected, crash probability increases with increased exposure, as shown by the odds ratios for segment mile-years and AADT. The model shows odds ratios greater than one for federal functional classes larger than local, indicating that the higher functional class roads are at an elevated risk for severe lane departure crashes. In terms of cross-sectional characteristics, roads with narrow shoulders were found to be at a slightly elevated risk of severe lane departure crashes compared to wider shoulders. In contrast, roads with wider traveled way width (wider than 23 feet) were found to be at an elevated risk – likely correlating with higher speed travel. These findings align with prior study evaluating severe lane departure crashes. [7] Segments with 1 or fewer ABCC licenses nearby were found to be at a higher risk of a crash, suggesting a correlation with the need to drive further distances to obtain alcohol. Finally, several MPOs (Middlesex, Pioneer, Old Colony, Southeast, Cape Cod, and Montachusett) were found to be at an elevated risk compared to other MPOs.

Table 3. Binary Logistic Regression Results for Urban Segments with Known AADT.

Variable (Number)	Odds Ratio	Standard Error	z-value	P>|z|	95% Confidence Interval	
Natural log of segment mile-years. (1)	2.743	0.129	21.42	<0.001	2.501	3.008
Natural log of AADT (2)	1.548	0.084	8.04	<0.001	1.392	1.723
Federal functional class is principal arterial – other or minor arterial (3)	2.719	0.711	3.83	<0.001	1.628	4.538
Federal functional class is major collector or minor collector (4)	2.323	.580	3.38	0.001	1.424	3.789
Average shoulder width is 4 feet or less (5)	1.137	0.142	1.02	0.306	0.889	1.453
Surface width is 23 feet or wider (6)	1.092	0.071	1.36	0.175	0.962	1.240
MPO is Middlesex or Pioneer (7)	1.364	0.121	3.51	<0.001	1.147	1.623
MPO is Old Colony or Southeast (8)	1.712	0.135	6.84	<0.001	1.468	1.998
MPO is Cape Cod or Montachusett (9)	1.232	0.123	2.10	0.036	1.014	1.498
1 or fewer ABCC Licenses within a quarter mile of the segment (10)	1.146	0.089	1.76	0.079	0.984	1.335
Constant (11)	0.0002	0.0001	-17.56	<0.001	0.00008	0.0005

Note: Number of observations = 55,540; Log likelihood = -4992.0683; Pseudo R2 = 0.0645; LR chi2(10) = 688.95; Prob > chi2 = <0.0001.

[7] Sawtelle et al. (2023). Driver, roadway, and weather factors on severity of lane departure crashes in Maine. Journal of Safety Research, Volume 84, https://doi.org/10.1016/j.jsr.2022.11.006.

Table 4 is a correlation matrix identifying correlation between any two variables. There is significant correlation between variables 3 and 4 (the Federal Function Class indicators); however, model results were stable when included, so VHB elected to keep both variables in the model.

Table 4. Correlation Matrix for Binary Logistic Regression Model of Urban Segments with Known AADT

	2	3	4	5	6	7	8	9	10
2	1.000								
3	0.618	1.000							
4	-0.283	-0.813	1.000						
5	-0.213	-0.189	0.156	1.000					
6	0.182	0.166	-0.116	-0.223	1.000				
7	0.037	0.066	-0.016	-0.010	-0.005	1.000			
8	0.059	0.033	0.046	-0.036	0.017	-0.176	1.000		
9	-0.086	0.075	-0.046	0.021	-0.030	-0.159	-0.167	1.000	
10	-0.132	-0.062	0.075	0.020	-0.013	0.011	0.025	0.029	1.000

Urban Segments with Unknown AADT

The binary logistic regression model for urban focus facility segments with unknown or default AADT is summarized in Table 5. The model shows odds ratios greater than 1.0 when there is either no curb, or just curb on one side of the roadway, as opposed to curb on both sides of the roadway, suggesting elevated crash risk when there is not curb to prevent vehicles from redirecting the roadside. Crash risk is also increased when a segment is near an ABCC license venue and on roads with narrow traveled ways. While these findings are contrary to those found in the previous model, this is likely due to the differing nature of the lower-class roads for which MassDOT does not have AADT. Finally, segments in the Southeast Massachusetts MPO were found to be at an increased risk, while those in the Cape Cod MPO were found to be at a decreased risk.

Table 5. Binary Logistic Regression Results for Urban Segments with Unknown AADT.

Variable (Number)	Odds Ratio	Standard Error	z-value	P>|z|	95% Confidence Interval	
Natural log of segment mile-years. (1)	3.911	0.310	17.18	<0.001	3.347	4.569
No curb present on roadway (2)	1.295	0.196	1.71	0.088	0.963	1.743
Curb present on one side of roadway (3)	2.788	0.691	4.14	<0.001	1.716	4.531
At least one ABCC license within a quarter mile of the segment (4)	1.216	0.145	1.65	0.100	0.963	1.535
Surface width is 22 feet or less (5)	1.770	0.219	4.61	<0.001	1.389	2.256
MPO is Southeast (6)	2.160	0.271	6.14	<0.001	1.689	2.761
MPO is Cape Cod (7)	0.551	0.138	-2.38	0.017	0.337	0.899
Constant (8)	0.003	0.0004	-44.21	<0.001	0.002	0.004

Note: Number of observations = 110,260; Log likelihood = -2129.266; Pseudo R2 = 0.0855; LR chi2(7) = 398.12; Prob > chi2 = <0.0001.

Table 6 summarizes the correlation between variables in the Urban Segments with Unknown AADT model. Note that the largest correlation value is 0.400, which was between 2 (the lack of curb presence) and 5 (surface width of 22 feet or less).

Table 6. Correlation Matrix for Binary Logistic Regression Model of Urban Segments with Unknown AADT

	2	3	4	5	6	7
2	1.000					
3	-0.262	1.000				
4	-0.105	0.084	1.000			
5	0.400	-0.046	-0.038	1.000		
6	0.009	-0.005	0.009	-0.021	1.000	
7	0.200	-0.033	-0.045	0.326	-0.115	1.000

Rural Segments with Known AADT

The binary logistic regression model for rural focus facility segments with known AADT is summarized in Table 7. As expected, crash risk increases as AADT increases, even more so when the AADT is greater than 4,000 vehicles per day. Additionally, the model shows elevated risk for segments on rolling or mountainous terrain, suggesting potential difficulty for drivers navigating those segments as opposed to level terrain. Aggregate shoulders (unruttable compacted subgrade) also show an elevated risk compared to other shoulder types. A posted speed limit of 50 MPH suggests that higher speeds may be correlated with more severe crash outcomes. Finally, the Old Colony MPO's segments are at an elevated risk.

Table 7. Binary Logistic Regression Results for Rural Segments with Known AADT.

Variable (Number)	Odds Ratio	Standard Error	z-value	P>|z|	95% Confidence Interval	
Natural log of segment mile-years. (1)	2.449	0.340	6.45	<0.001	1.866	3.216
Natural log of AADT (2)	1.286	0.199	1.63	0.104	0.950	1.741
Terrain is Rolling or Mountainous (3)	1.711	0.468	1.96	0.049	1.001	2.924
MPO is Old Colony (4)	4.979	2.824	2.83	0.005	1.638	15.134
Right Shoulder Type is Stable - Unruttable compacted subgrade (5)	1.408	0.327	1.47	0.141	0.893	2.219
AADT is greater than 4,000 vpd (6)	1.949	0.723	1.80	0.072	0.943	4.031
Posted Speed Limit is 50 MPH (7)	2.033	1.252	1.15	0.249	0.608	6.797
Constant (8)	0.002	0.002	-5.70	<0.001	0.0002	0.015

Note: Number of observations = 5,273; Log likelihood = -419.30215; Pseudo R2 = 0.0713; LR chi2(7) = 64.43; Prob > chi2 = <0.0001.

Table 8 shows the correlation matrix for variables included in the Rural Segments with Known AADT model. The highest correlation found was between variable 2 (natural log of AADT) and variable 6 (AADT greater than 4,000 veh/day), which is expected. However, this correlation was not found to affect the stability of the model.

Table 8. Correlation Matrix for Binary Logistic Regression Model of Rural Segments with Known AADT

	2	3	4	5	6	7
2	1.000					
3	-0.080	1.000				
4	0.059	-0.033	1.000			
5	-0.089	-0.001	0.100	1.000		
6	0.557	-0.075	0.031	-0.025	1.000	
7	0.081	-0.017	-0.011	-0.071	0.029	1.000

Rural Segments with Unknown AADT

Table 9 summarizes the model for rural segments with unknown AADT. Unfortunately, given the small sample size and data availability for these segments, VHB was only able to include two risk factors in the model, one of which is an MPO (Nantucket) and the other being Mountainous terrain, neither of which occur concurrently.

Table 9. Binary Logistic Regression Results for Rural Segments with Unknown AADT.

Variable (Number)	Odds Ratio	Standard Error	z-value	P>|z|	95% Confidence Interval	
Natural log of segment mile-years. (1)	2.858	0.642	4.68	<0.001	1.840	4.437
MPO is Nantucket (2)	3.874	4.023	1.30	0.192	0.506	29.658
Mountainous (3)	2.441	1.805	1.21	0.228	0.573	10.402
Constant (4)	0.003	.0007	-24.63	<0.001	0.002	0.005

Note: Number of observations = 7,278; Log likelihood = -199.23505; Pseudo R2 = 0.0557; LR chi2(3) = 23.52; Prob > chi2 = <0.0001.

Table 10 shows there is almost no correlation between the two non-exposure variables included in the model.

Table 10. Correlation Matrix for Binary Logistic Regression Model of Rural Segments with Unknown AADT

	2	3
2	1.000	
3	-0.026	1.000

Geometric Risk Factors

As mentioned previously, MassDOT provided VHB with horizontal and vertical alignment data. VHB used an Overrepresentation approach to identify potential alignment risk factors. With this approach, VHB compared the proportion of focus crashes with certain alignment characteristics to the proportion of mileage with the same characteristics. VHB used curve and grade classifications which were assigned in the MassDOT data.

Figure 4 shows the overrepresentation results for horizontal curvature. Note that curves within the 3.5-5.4, 5.5-8.4, and 8.5-13.9 degrees of curvature classifications were all overrepresented compared to mileage, accounting for 29.7 percent of focus crashes compared to just 24.8 percent of focus facility mileage on curves. As such, VHB selected segments including a curve in these classes as an additional risk factor.

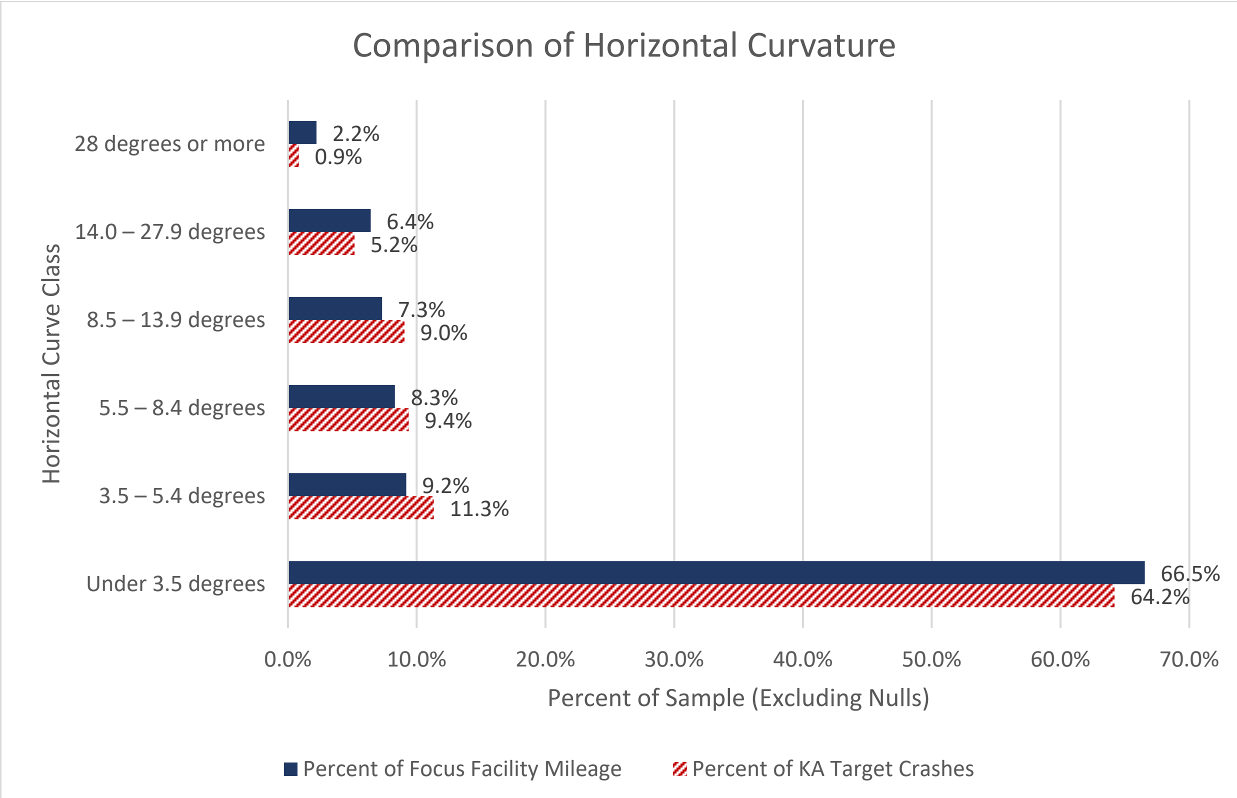

Figure 4. Overrepresentation Analysis of Horizontal Curves.

Similarly, Figure 5 shows the distribution of target crashes and focus facility mileage by vertical grade. Note that flat segments turned out to be overrepresented, with grades from 0 percent to 2.4 percent accounting for 69.8 percent of target crashes compared to 65.5 percent of mileage. However, VHB and MassDOT elected not to include a grade-based risk factor given the small total overrepresentation and little documented relationship between grade and lane departure crash frequency and severity.

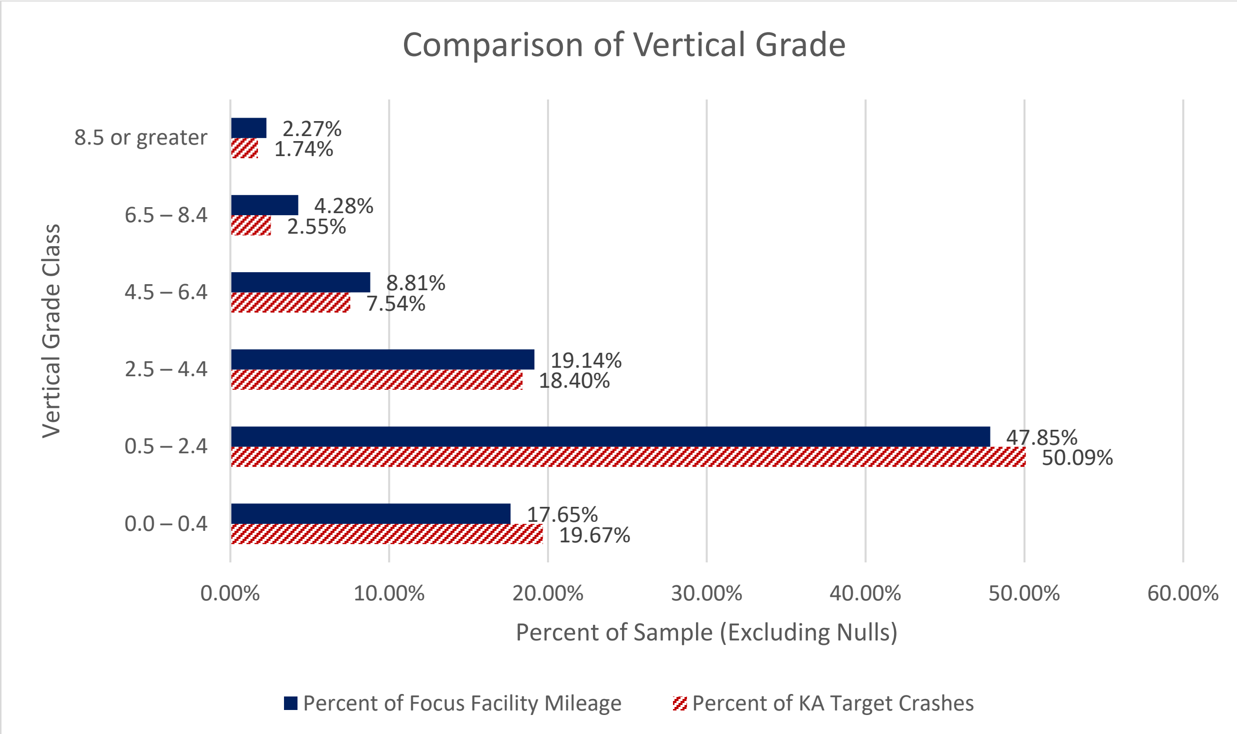

Figure 5. Overrepresentation Analysis of Vertical Grades

Conclusions and Recommendations

The purpose of this analysis is to identify segment-level risk factors for severe lane departure crashes. As discussed previously, there were four sets of risk factors identified based on area type and data availability. The following tables describe the suggested risk factor scoring for each set of risk factors:

- Table 11 summarizes risk factors for urban two-lane undivided roads with known AADT.
- Table 12 summarizes risk factors for urban two-lane undivided roads with unknown AADT.
- Table 13 summarizes risk factors for rural two-lane undivided roads with known AADT.
- Table 14 summarizes risk factors for rural two-lane undivided roads with unknown AADT.

Generally, VHB recommends using risk scores between 0 and 1 so as not to have on risk factor outweighing others. In the event a risk factor was not statistically significant at the 95 percent confidence level, VHB recommended maximum risk scores less than 1.0.

Table 11. Risk Scoring for Urban, Two-Lane, Undivided Roads with Known AADT.

Risk Factor for Lane Departure Crashes	Suggested Scoring
Natural log of AADT	Linear From 0 to 1 for the range of values
Federal functional class is principal arterial – other or minor arterial	1 if true; else
Federal functional class is major collector or minor collector	0.75 if true; 0 otherwise
Average shoulder width is 4 feet or less	Linear from 0 to 1 for 4 feet to 0 feet
Surface width is 23 feet or wider	1 if true; 0 otherwise
MPO is Middlesex or Pioneer	0.75 if true; else
MPO is Old Colony or Southeast	1 if true; else
MPO is Cape Cod or Montachusett	0.5 if true; 0 otherwise
1 or fewer ABCC Licenses within a quarter mile of the segment	1 if true; 0 otherwise
Horizontal curvature present in the segment from 3.5 degrees of curvature to 13.9 degrees of curvature.	1 if true; 0 otherwise
Maximum potential score for a segment	**9**

Table 12. Risk Scoring for Urban, Two-Lane, Undivided Roads with Unknown AADT.

Risk Factor for Lane Departure Crashes	Suggested Scoring
No curb present on roadway	0.5 if true; else
Curb present on one side of roadway	1 if true; 0 otherwise
At least one ABCC license within a quarter mile of the segment	1 if true; 0 otherwise
Surface width is 22 feet or less	Linear from 0 at 22 feet to 1 at 16 feet or narrower.
MPO is not Cape Cod	0.25 if true; else
MPO is Southeast	1 if true; 0 otherwise
Horizontal curvature present in the segment from 3.5 degrees of curvature to 13.9 degrees of curvature.	1 if true; 0 otherwise
Maximum potential score for a segment	**5.75**

Table 13. Risk Scoring for Rural, Two-Lane, Undivided Roads with Known AADT.

Risk Factor for Lane Departure Crashes	Suggested Scoring
Natural log of AADT	Linear from 0 to 1 for the range of volumes
Terrain is Rolling or Mountainous	1 if true; 0 otherwise
MPO is Old Colony	1 if true; 0 otherwise
Right Shoulder Type is Stable - Unruttable compacted subgrade	1 if true; 0 otherwise
AADT is greater than 4,000 vpd	1 if true; 0 otherwise
Posted Speed Limit is 50 MPH	1 if true; 0 otherwise
Horizontal curvature present in the segment from 3.5 degrees of curvature to 13.9 degrees of curvature.	1 if true; 0 otherwise
Maximum potential score for a segment	**7**

Table 14. Risk Scoring for Rural, Two-Lane, Undivided Roads with Unknown AADT.

Risk Factor for Lane Departure Crashes	Suggested Scoring
MPO is Nantucket	1 if true; 0 otherwise
Mountainous	1 if true; 0 otherwise
Horizontal curvature present in the segment from 3.5 degrees of curvature to 13.9 degrees of curvature.	1 if true; 0 otherwise
Maximum potential score for a segment	**3**

Table 15 provides an example application of the risk factors on a hypothetical rural segment with known AADT. To balance prioritization across the different risk scoring schemes, VHB recommends normalizing the segment risk scores against the total possible score for each schema – producing a normalized risk score for each segment ranging from 0 to 100. Table 14 summarizes an example risk scoring calculation for a rural segment with known AADT. This segment has a total risk score of 3.71 out of 7, resulting in a normalized risk score of 53 percent.

Table 15. Example Risk Score Calculation for Lane Departure Crashes on for Rural, Two-Lane, Undivided Roads with Known AADT

Variable	Segment Characteristic	Risk Factor	Risk Score
Natural log of AADT	7.31 (AADT is 1,500)	Linear from 0 to 1 for the range of volumes	0.71
Terrain is Rolling or Mountainous	Rolling	1 if true; 0 otherwise	1
MPO is Old Colony	Cape Cod Commission	1 if true; 0 otherwise	0
Right Shoulder Type is Stable - Unruttable compacted subgrade	True	1 if true; 0 otherwise	1
AADT is greater than 4,000 vpd	False	1 if true; 0 otherwise	0
Posted Speed Limit is 50 MPH	45 MPH	1 if true; 0 otherwise	0
Horizontal curvature present in the segment from 3.5 degrees of curvature to 13.9 degrees of curvature.	Curve of 5.4 degrees present	1 if true; 0 otherwise	1
		Total Risk Score:	3.71
		Risk Percent Score (Out of 7):	53%

MassDOT ranked the segments at both the Statewide and MPO levels using the normalized risk score and the percentile score of 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 percent. For lane departure crashes at segments, normalized risk scores range from 0 to 0.77. The maximum value (0.77) received a percentile rank of 100 and other values received a percentile rank accordingly. For example, a segment with a normalized risk score of 0.60, the calculated state percentile rank was 92.60, and fell in the secondary risk category. MassDOT then assigned risk categories using the computed ranks. For example, segments ranked in the top 5 percentile (95 through 100) were categorized as “Primary Risk Sites” and segments ranked in the next 10 percentile (85 through 95) were categorized as “Secondary Risk Sites”; the remaining segments 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 percent or next 10 percent may not be equal to 5 or 10 percent. This is a byproduct of the weak ranking approach. Table 16 and Table 17 show the distribution of segments and crashes with the normalized risk score (presented as percentages) across these categories for Statewide and MPO rankings, respectively. Note the goal was to see a higher proportion of target crashes for primary and secondary risk sites than proportion of segments. Similarly, Figure 6 is a map of the risk segments ranked statewide, while Figure 7 is a map of the risk segments ranked by MPO. The data elements included in the maps are those necessary to identify and provide context for the roadway segment as well as the risk factors present along the segment. VHB excluded features considered in the analysis but not ultimately included as risk factors.

There are a total of 7,429 segments in the primary risk category (top 5 percent), that captured 16.4 percent of the severe lane departure crashes. Similarly, there are 14,856 segments in the secondary risk category (next top 10 percent), which captured an additional 25.4 percent of the severe pedestrian crashes. The highest number of primary risk category segments were in Boston Region MPO (2,754 segments), followed by Pioneer Valley Planning Commission (816 segments) and Central Massachusetts Regional Planning Commission (739 segments).

Table 16. Statewide Risk Categories.

State	Risk Category	Minimum Normalized Risk Score Percentage	Maximum Normalized Risk Score Percentage	Number of Segments	Percent of Scored State Segments	Percent of Target Crashes
MA	Primary Risk Site	62.3%	77.8%	7,429	5.0%	16.4%
	Secondary Risk Site	52.6%	62.3%	14,856	10.0%	25.4%

Table 17. Distribution of Risk Segments my MPO.

MPO	Risk Category	Minimum Normalized Risk Score Percentage	Maximum Normalized Risk Score Percentage	Number of Segments	Percent of Scored MPO Segments	Percent of Target Crashes in MPO
Berkshire Regional Planning Commission	Primary Risk Site	51.4%	70.6%	200	5.0%	26.7%
	Secondary Risk Site	41.7%	51.2%	400	10.0%	13.3%
Boston Region MPO	Primary Risk Site	58.6%	70.4%	2754	5.0%	15.8%
	Secondary Risk Site	52.0%	58.6%	5949	10.8%	23.7%
Cape Cod Commission	Primary Risk Site	58.4%	72.1%	575	5.0%	14.4%
	Secondary Risk Site	51.2%	58.4%	1149	10.0%	26.1%
Central Massachusetts Regional Planning Commission	Primary Risk Site	58.5%	71.0%	739	5.0%	11.8%
	Secondary Risk Site	51.0%	58.5%	147	10.0%	21.9%
Franklin Regional Council of Governments	Primary Risk Site	54.1%	71.3%	162	5.0%	25.8%
	Secondary Risk Site	43.2%	54.1%	322	10.0%	16.1%
Martha's Vineyard Commission	Primary Risk Site	50.7%	64.5%	31	5.3%	0.0%

MPO	Risk Category	Minimum Normalized Risk Score Percentage	Maximum Normalized Risk Score Percentage	Number of Segments	Percent of Scored MPO Segments	Percent of Target Crashes in MPO
	Secondary Risk Site	39.2%	50.4%	58	9.8%	0.0%
Merrimack Valley Planning Commission	Primary Risk Site	58.3%	66.5%	384	5.0%	19.5%
	Secondary Risk Site	52.0%	58.3%	769	10.1%	11.7%
Montachusett Regional Planning Commission	Primary Risk Site	65.5%	72.2%	369	5.0%	14.0%
	Secondary Risk Site	56.2%	65.5%	735	10.0%	25.2%
Nantucket Planning and Economic Development Commission	Primary Risk Site	51.7%	66.4%	18	5.3%	0.0%
	Secondary Risk Site	41.9%	51.7%	34	10.0%	50.0%
Northern Middlesex Council of Governments	Primary Risk Site	67.7%	75.0%	343	5.0%	10.4%
	Secondary Risk Site	60.3%	67.7%	880	12.9%	20.8%
Pioneer Valley Planning Commission	Primary Risk Site	62.9%	75.0%	816	5.5%	21.9%
	Secondary Risk Site	54.2%	62.9%	1872	12.7%	20.4%
Old Colony Planning Council	Primary Risk Site	66.1%	77.8%	441	5.0%	15.8%
	Secondary Risk Site	61.7%	66.1%	890	10.1%	17.7%
Southeastern Regional Planning and Economic Development District	Primary Risk Site	65.3%	77.7%	694	5.0%	20.3%
	Secondary Risk Site	60.8%	65.3%	1420	10.3%	16.4%

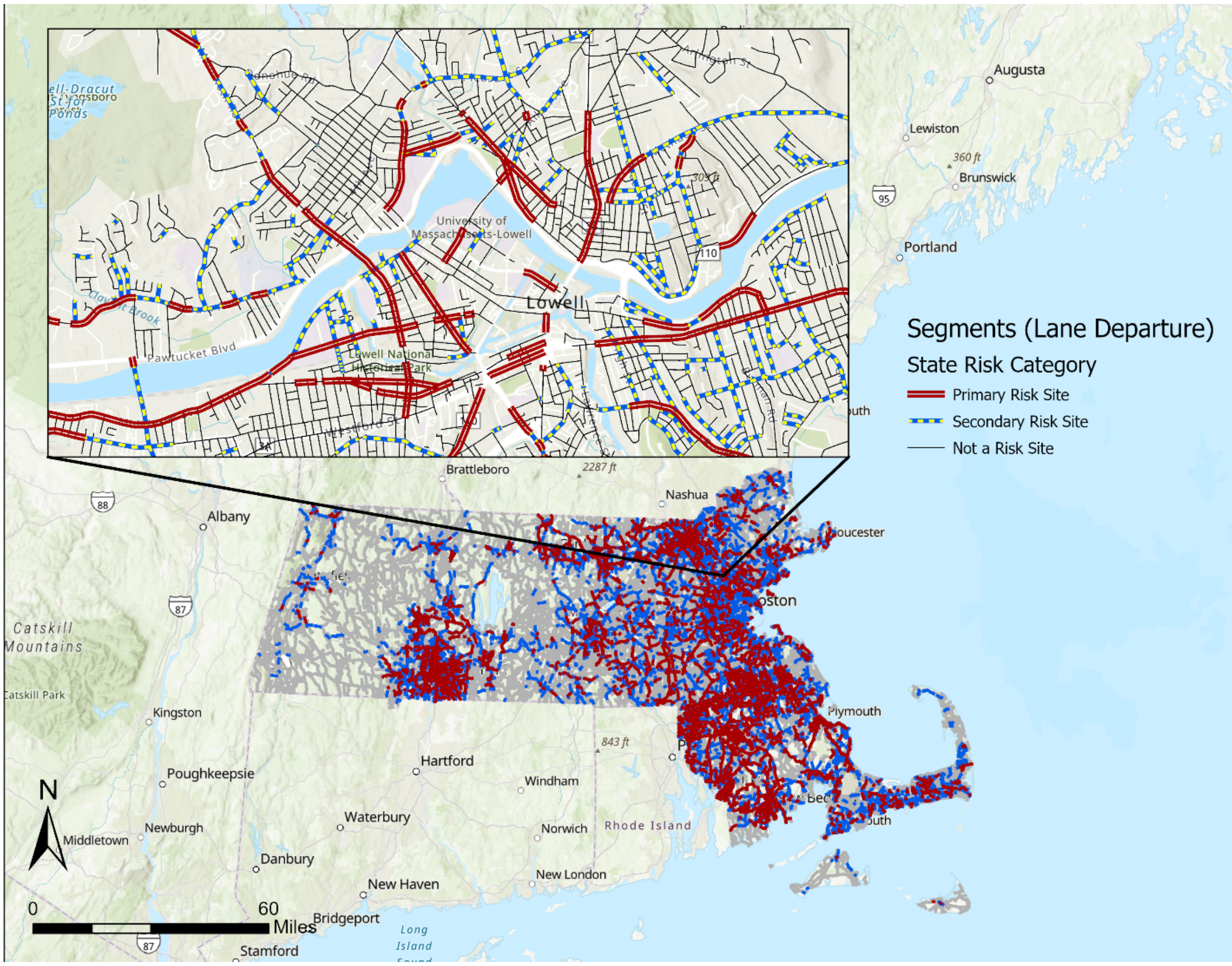

Figure 6. Map depicting the primary and secondary risk segments for severe lane departure crashes, ranked statewide.

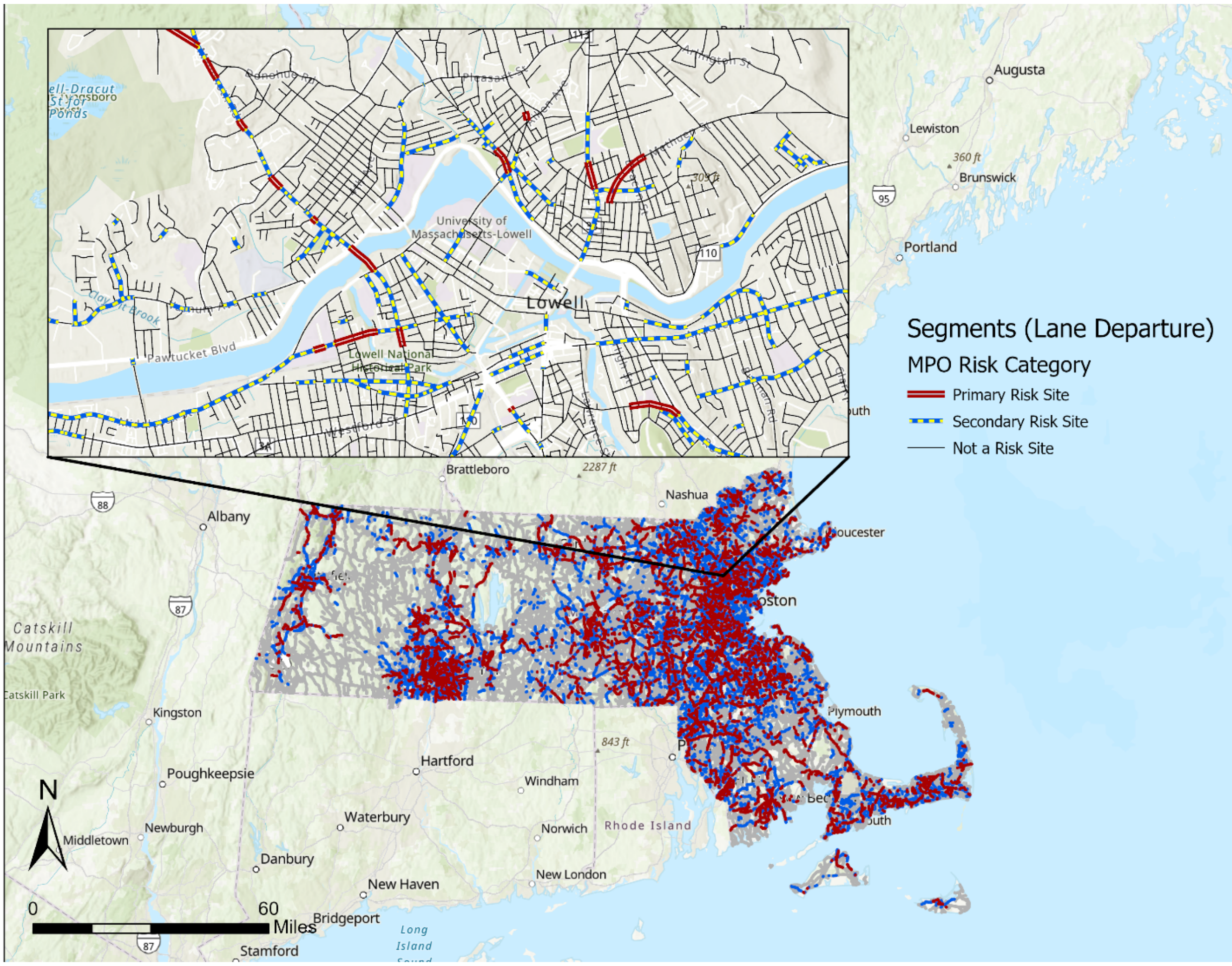

Figure 7. Map depicting the primary and secondary risk segments for severe lane departure crashes, ranked by MPO.