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

Young Drivers (24 and Under)

PREPARED FOR

PREPARED BY

REPORT DATE: AUGUST 2021

Purpose & Background

The Massachusetts Department of Transportation (MassDOT) was awarded a grant by the United States Department of Transportation (USDOT) under its Safety Data Initiative (SDI) competition. MassDOT's work under this grant includes the creation of a Safety Analysis Module in their online IMPACT tool. One feature in this module will be a mapping component which will include crash-based and systemic network screening maps. 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 is part of the SDI project and summarizes the risk factor analysis performed for young driver-related (i.e., involving a driver under the age of 25) crashes. It also describes a method to identify risk factors using negative binomial regression, which is one potential method to identify risk factors under the SDI grant. Reports for other emphasis areas describe different methods used to adapt to the needs of those areas.

Focus Crash Types

Over 14,000 fatal (K) and suspected serious injury (A) crashes (1,693 K and 12,528 A crashes, respectively) occurred in Massachusetts between 2013 and 2017. This analysis excluded incomplete crash data from the City of Boston, leaving 1,576 K and 11,997 A crashes during the 5-year study period. Based on discussions with MassDOT, VHB established that any crash that involved a driver aged 24 or younger at the time of the crash is defined as a young driver-related crash (regardless of the party at-fault). Of the 13,573 KA crashes that occurred during the study period, 3,949 involved a young driver (427 K and 3,522 A crashes).

VHB compared the distribution of young driver KA crashes to the distribution of remaining young driver non-incapacitating injury, possible injury, and no injury (B, C, and O) crashes across a series of crash-level characteristics. Where the proportion for a given attribute is statistically larger than the proportion for the comparison group, that attribute is flagged as a potential risk factor. 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 young driver 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.

The following sections document these comparisons and highlight the key takeaways for systemic risk factor analysis.

Manner of Collision and First Harmful Event

Tables 1 and 2 illustrate that young driver related KA crashes are much more likely to be the result of single-vehicle crashes than BCO crashes. Furthermore, head on collisions are a statistically larger proportion of KA crashes than BCO crashes. While this is a common characteristic of all crash types (not just young driver crashes), the combination of head on crashes and single vehicle crashes indicates that lane departure is a key overlapping risk factor with young driver crashes. Speeding may further exacerbate this issue, as younger, inexperienced drivers may lose control at higher speeds. Figures 3 and 4 show that a greater percentage of KA young driver crashes involve a speeding vehicle or reckless driving.

Table 1. Summary of young driver crashes by manner of collision.

Manner of Collision	Young Driver KA Crashes			Young Driver BCO Crashes		
	Total	Percentage	Sampling Error	Total	Percentage	Sampling Error
Single vehicle crash	**1,358**	**34.4%**	**0.8%**	**41,022**	**19.6%**	**0.1%**
Angle	1,145	29.0%	0.7%	61,889	29.6%	0.1%
Rear-end	710	18.0%	0.6%	73,895	35.4%	0.1%
Head-on	**450**	**11.4%**	**0.5%**	**6,147**	**2.9%**	**0.0%**
Sideswipe, same direction	137	3.5%	0.3%	16,263	7.8%	0.1%
Sideswipe, opposite direction	83	2.1%	0.2%	6,008	2.9%	0.0%
Not reported	28	0.7%	0.1%	1,140	0.5%	0.0%
Unknown	26	0.7%	0.1%	1,251	0.6%	0.0%
Rear-to-rear	9	0.2%	0.1%	1,094	0.5%	0.0%
Front to Front	2	0.1%	0.0%	33	0.0%	0.0%
Other	1	0.0%	0.0%	7	0.0%	0.0%
Front to Rear	0	0.0%	0.0%	53	0.0%	0.0%
Rear to Side	0	0.0%	0.0%	16	0.0%	0.0%
Reported but invalid	0	0.0%	0.0%	9	0.0%	0.0%
Blank	0	0.0%	0.0%	2	0.0%	0.0%

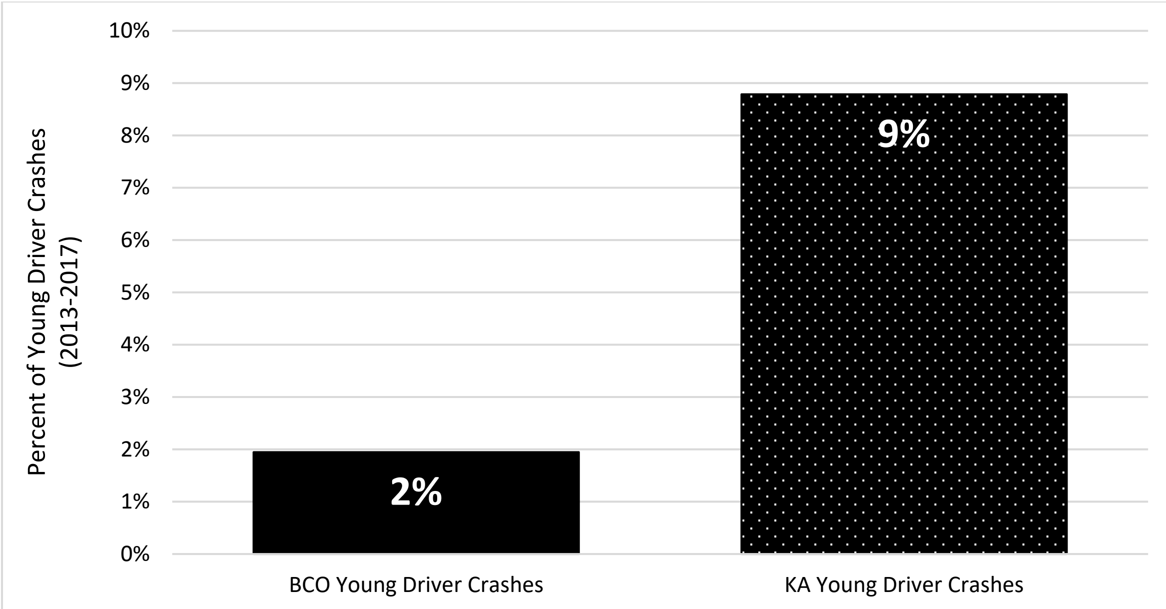

Figure 3. Percentage of young driver crashes involving a driver exceeding the speed limit.[1]

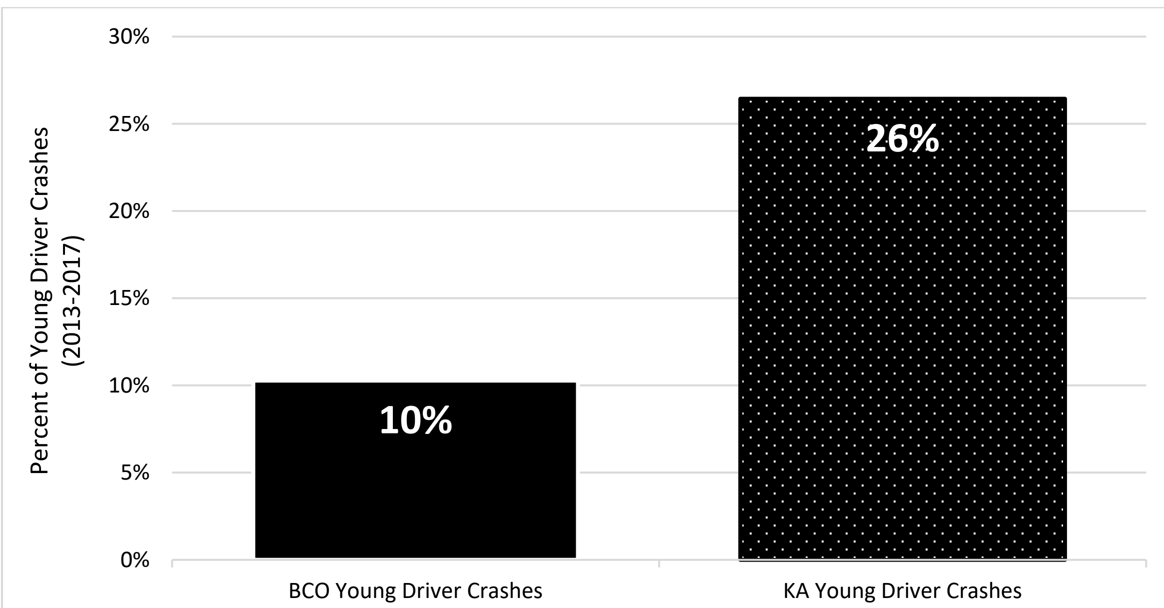

Figure 4. Percentage of young driver crashes involving a driver exceeding the speed limit, driving too fast for conditions, or driving recklessly.[2]

[1] Driver contributing circumstances: Exceeded authorized speed limit.
[2] Driver contributing circumstances: Exceeded authorized speed limit, Operating vehicle in erratic, reckless, careless, negligent, or aggressive manner, or Driving too fast for conditions.

Table 2. Summary of young driver crashes by first harmful event.

First Harmful Event	Young Driver KA Crashes			Young Driver BCO Crashes		
	Total	Percentage	Sampling Error	Total	Percentage	Sampling Error
Collision with motor vehicle in traffic	2,316	58.6%	0.8%	153,673	73.6%	0.1%
Collision with pedestrian	**274**	**6.9%**	**0.4%**	**1,222**	**0.6%**	**0.0%**
Collision with tree	**332**	**8.4%**	**0.4%**	**5,877**	**2.8%**	**0.0%**
Collision with pedalcycle (bicycle, tricycle, unicycle, pedal car)	**68**	**1.7%**	**0.2%**	**681**	**0.3%**	**0.0%**
Collision with utility pole	**226**	**5.7%**	**0.4%**	**7,667**	**3.7%**	**0.0%**
Collision with parked motor vehicle	98	2.5%	0.2%	11,236	5.4%	0.0%
Collision with guardrail	**122**	**3.1%**	**0.3%**	**5,169**	**2.5%**	**0.0%**
Not reported	38	1.0%	0.2%	2,377	1.1%	0.0%
Collision with other	52	1.3%	0.2%	2,313	1.1%	0.0%
Collision with curb	**79**	**2.0%**	**0.2%**	**2,980**	**1.4%**	**0.0%**
Overturn/rollover	**83**	**2.1%**	**0.2%**	**1,009**	**0.5%**	**0.0%**
Other First Harmful Event	261	6.6%	--	14,623	7.0%	--

Intersection Related and Junction Type

Table 3 details the relationship of KA crashes to specific intersection types. Table 3 indicates that young driver-related KA crashes tend to be segment-based and less related to an intersection. This supports the results in Tables 1 and 2 and suggest that more severe young driver crashes tend to be related to lane departure, and potentially at relatively higher speeds, between intersections. Table 4 illustrates that there is no significant difference in the distribution of intersection-related KA or BCO young driver crashes according to traffic control device.

Table 3. Summary of young driver crashes by junction type.

Junction Type	Young Driver KA Crashes			Young Driver BCO Crashes		
	Total	Percentage	Sampling Error	Total	Percentage	Sampling Error
Not at junction	**2,281**	**57.8%**	**0.8%**	**109,688**	**52.5%**	**0.1%**
Four-way intersection	669	16.9%	0.6%	36,727	17.6%	0.1%
T-intersection	650	16.5%	0.6%	38,330	18.4%	0.1%
Driveway	88	2.2%	0.2%	5,495	2.6%	0.0%
Y-intersection	86	2.2%	0.2%	5,082	2.4%	0.0%
Not reported	36	0.9%	0.2%	2,052	1.0%	0.0%
Off-ramp	54	1.4%	0.2%	3,881	1.9%	0.0%
On-ramp	40	1.0%	0.2%	3,438	1.6%	0.0%
Traffic circle	13	0.3%	0.1%	2,340	1.1%	0.0%
Five-point or more	19	0.5%	0.1%	900	0.4%	0.0%
Unknown	11	0.3%	0.1%	778	0.4%	0.0%
Railway grade crossing	2	0.1%	0.0%	92	0.0%	0.0%
Reported but invalid	0	0.0%	0.0%	24	0.0%	0.0%

Table 4. Summary of young driver intersection-related crashes by traffic control device.[3]

Traffic Control	Young Driver KA Crashes			Young Driver BCO Crashes		
	Total	Percentage	Sampling Error	Total	Percentage	Sampling Error
No controls	464	32.3%	1.2%	25,450	30.5%	0.2%
Stop signs	462	32.2%	1.2%	25,583	30.7%	0.2%
Traffic control signal	446	31.0%	1.2%	27,552	33.0%	0.2%
Flashing traffic control signal	34	2.4%	0.4%	1,512	1.8%	0.0%
Yield signs	12	0.8%	0.2%	2,283	2.7%	0.1%
Warning signs	10	0.7%	0.2%	249	0.3%	0.0%
Not reported	7	0.5%	0.2%	534	0.6%	0.0%
Unknown	1	0.1%	0.1%	151	0.2%	0.0%
School zone signs	0	0.0%	0.0%	50	0.1%	0.0%
Railway crossing device	1	0.1%	0.1%	14	0.0%	0.0%

Lighting Condition and Time of Day

Table 5 underscores that young driver-related KA crashes are much more likely to occur during dark lighting conditions than BCO crashes. Related to previous tables, this could be an indication of dark lighting conditions interacting with speeding and lane departure circumstances; darkness may limit visibility, reduce sight distance, and complicate situations for inexperienced drivers. Table 6 reflects this trend, as 33.6 percent of KA young driver crashes occur between 8 pm and 6 am, as opposed to 19.9 percent of BCO young driver crashes.

[3] Includes Five-point or more, Four-way intersection, T-intersection, Traffic circle, and Y-intersection codes in the "rdwy_jnct_type_descr" field.

Table 5. Summary of young driver crashes by lighting condition.

Lighting Condition	Young Driver KA Crashes			Young Driver BCO Crashes		
	Total	Percentage	Sampling Error	Total	Percentage	Sampling Error
Daylight	2,213	56.0%	0.8%	140,526	67.3%	0.1%
Dark - lighted roadway	**1,108**	**28.1%**	**0.7%**	**45,281**	**21.7%**	**0.1%**
Dark - roadway not lighted	**440**	**11.1%**	**0.5%**	**12,118**	**5.8%**	**0.1%**
Dusk	98	2.5%	0.2%	5,784	2.8%	0.0%
Dawn	43	1.1%	0.2%	2,560	1.2%	0.0%
Dark - unknown roadway lighting	34	0.9%	0.1%	1,412	0.7%	0.0%
Not reported	11	0.3%	0.1%	649	0.3%	0.0%
Other	2	0.1%	0.0%	243	0.1%	0.0%
Unknown	0	0.0%	0.0%	256	0.1%	0.0%

Table 6. Summary of young driver crashes by hour of day.

Hour of Day	Young Driver KA Crashes			Young Driver BCO Crashes		
	Total	Percentage	Sampling Error	Total	Percentage	Sampling Error
12 AM - Midnight	**154**	**3.9%**	**0.3%**	**3,953**	**1.9%**	**0.0%**
1 AM	**156**	**4.0%**	**0.3%**	**3,635**	**1.7%**	**0.0%**
2 AM	**132**	**3.3%**	**0.3%**	**3,109**	**1.5%**	**0.0%**
3 AM	**62**	**1.6%**	**0.2%**	**1,972**	**0.9%**	**0.0%**
4 AM	**46**	**1.2%**	**0.2%**	**1,483**	**0.7%**	**0.0%**
5 AM	**63**	**1.6%**	**0.2%**	**1,827**	**0.9%**	**0.0%**
6 AM	63	1.6%	0.2%	4,371	2.1%	0.0%
7 AM	155	3.9%	0.3%	11,248	5.4%	0.0%
8 AM	122	3.1%	0.3%	9,686	4.6%	0.0%
9 AM	123	3.1%	0.3%	7,776	3.7%	0.0%
10 AM	118	3.0%	0.3%	8,022	3.8%	0.0%
11 AM	166	4.2%	0.3%	9,956	4.8%	0.0%
12 PM - Noon	202	5.1%	0.4%	12,299	5.9%	0.1%
1 PM	214	5.4%	0.4%	12,078	5.8%	0.1%
2 PM	232	5.9%	0.4%	16,538	7.9%	0.1%
3 PM	274	6.9%	0.4%	17,499	8.4%	0.1%
4 PM	256	6.5%	0.4%	17,686	8.5%	0.1%
5 PM	270	6.8%	0.4%	18,077	8.7%	0.1%
6 PM	239	6.1%	0.4%	13,105	6.3%	0.1%
7 PM	187	4.7%	0.3%	8,871	4.2%	0.0%
8 PM	**175**	**4.4%**	**0.3%**	**7,488**	**3.6%**	**0.0%**
9 PM	**192**	**4.9%**	**0.3%**	**7,374**	**3.5%**	**0.0%**
10 PM	**178**	**4.5%**	**0.3%**	**5,956**	**2.9%**	**0.0%**
11 PM	**168**	**4.3%**	**0.3%**	**4,800**	**2.3%**	**0.0%**

Seasonality

Table 7 shows that a slightly greater proportion of young driver KA crashes occur during later spring and summer months (April through September); over half of all young driver KA crashes occurred during this 6-month period (56 percent) as opposed to less than half of BCO young driver crashes (48 percent).

Table 7. Summary of young driver crashes by month.

Month	Young Driver KA Crashes			Young Driver BCO Crashes		
	Total	Percentage	Sampling Error	Total	Percentage	Sampling Error
January	286	7.2%	0.4%	19,020	9.1%	0.1%
February	239	6.1%	0.4%	18,309	8.8%	0.1%
March	268	6.8%	0.4%	16,089	7.7%	0.1%
April	**321**	**8.1%**	**0.4%**	**14,802**	**7.1%**	**0.1%**
May	**401**	**10.2%**	**0.5%**	**17,666**	**8.5%**	**0.1%**
June	370	9.4%	0.5%	18,047	8.6%	0.1%
July	**413**	**10.5%**	**0.5%**	**17,171**	**8.2%**	**0.1%**
August	**382**	**9.7%**	**0.5%**	**17,013**	**8.1%**	**0.1%**
September	326	8.3%	0.4%	15,530	7.4%	0.1%
October	345	8.7%	0.4%	17,701	8.5%	0.1%
November	314	8.0%	0.4%	17,486	8.4%	0.1%
December	284	7.2%	0.4%	19,995	9.6%	0.1%

Focus Facility Types

There are not many substantial differences in facility type, traffic volume, or posted speed limit with respect to KA and BCO young driver crashes. The few characteristics that are overrepresented in the young driver KA dataset include:

- Interstate and major collector functional classifications (Table 8).
- Two-way, divided, unprotected median (Table 9).
- Posted speed limits of 45 and 65 miles per hour (Table 10).

Table 8. Summary of young driver crashes by functional classification.

Functional Classification	Young Driver KA Crashes			Young Driver BCO Crashes		
	Total	Percentage	Sampling Error	Total	Percentage	Sampling Error
Minor Arterial	1,240	31.4%	0.7%	62,669	30.0%	0.1%
Principal Arterial - Other	1,089	27.6%	0.7%	65,383	31.3%	0.1%
Major Collector	**448**	**11.3%**	**0.5%**	**20,739**	**9.9%**	**0.1%**
Local	472	12.0%	0.5%	24,748	11.9%	0.1%
Interstate	**336**	**8.5%**	**0.4%**	**15,250**	**7.3%**	**0.1%**
(blank)	182	4.6%	0.3%	11,867	5.7%	0.1%
Principal Arterial - Other Freeways and Expressways	168	4.3%	0.3%	7,580	3.6%	0.0%
Minor Collector	14	0.4%	0.1%	593	0.3%	0.0%

Table 9. Summary of young driver crashes by trafficway description.

Trafficway	Young Driver KA Crashes			Young Driver BCO Crashes		
	Total	Percentage	Sampling Error	Total	Percentage	Sampling Error
Two-way, not divided	2,541	64.3%	0.8%	133,160	63.8%	0.1%
Two-way, divided, unprotected median	**619**	**15.7%**	**0.6%**	**28,555**	**13.7%**	**0.1%**
Two-way, divided, positive median barrier	593	15.0%	0.6%	30,787	14.7%	0.1%
One-way, not divided	131	3.3%	0.3%	11,076	5.3%	0.0%
Not reported	42	1.1%	0.2%	2,365	1.1%	0.0%
Unknown	23	0.6%	0.1%	2,854	1.4%	0.0%

Table 10. Summary of young driver crashes by posted speed limit.

Posted Speed Limit	Young Driver KA Crashes			Young Driver BCO Crashes		
	Total	Percentage	Sampling Error	Total	Percentage	Sampling Error
15	1	0.1%	0.1%	54	0.1%	0.0%
20	29	1.5%	0.3%	2,029	2.0%	0.0%
25	102	5.3%	0.5%	8,201	8.1%	0.1%
30	427	22.0%	0.9%	25,056	24.7%	0.1%
35	350	18.1%	0.9%	19,914	19.6%	0.1%
40	272	14.0%	0.8%	14,453	14.3%	0.1%
45	**193**	**10.0%**	**0.7%**	**7,552**	**7.4%**	**0.1%**
50	99	5.1%	0.5%	5,166	5.1%	0.1%
55	151	7.8%	0.6%	7,982	7.9%	0.1%
60	21	1.1%	0.2%	724	0.7%	0.0%
65	**292**	**15.1%**	**0.8%**	**10,273**	**10.1%**	**0.1%**

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

Based on discussions with MassDOT, VHB used a negative binomial count regression modeling approach to identify community-level characteristics that are associated with higher frequencies of young driver-related KA crashes. Negative binomial regression is a commonly used method in transportation safety as it applies to over-dispersed count data (i.e., the variance exceeds the mean of the observed data). The dependent variable in the model is the number of young driver-related KA crashes, making a count model appropriate for the data. The functional form of the negative binomial regression model is shown in Figure 5.[4]

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

Figure 5. Equation. Negative binomial regression functional form.

Where:

> $e^{\varepsilon i}$ = gamma distributed error term, where $e^{\varepsilon i}$ is gamma-distributed with a mean equal to one and variance equal to α.
>
> λ_i = expected number of young driver-related KA crashes at location i.
>
> β = vector of estimated parameters.
>
> Xi = vector of independent variables that characterize location i and influence young driver-related KA crash frequency.

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 noted this could lead to misunderstanding or outright disregard for potentially noteworthy results.[5]

Data

VHB used ArcGIS to manage and integrate data for this analysis. VHB aggregated data at the city and town level. In Massachusetts, all roads and geographic area are covered by town jurisdictions. Due to limitations with crash data acquisition, VHB excluded the City of Boston from the analysis. MassDOT provided VHB with various sources of data, as described in the following sections.

---

[4] Lord, D., Mannering, F., 2010. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. Transp. Res. Part A Policy Pract. 44 5 , 291–305. doi:10.1016/j.tra.2010.02.001

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

Crash Data

MassDOT provided statewide geolocated young driver-related crash data for the years 2013 through 2017. This included all crashes involving a driver aged 24 or younger at the time of the crash, regardless of the at-fault party.

Roadway Data

VHB downloaded the Massachusetts statewide roadway inventory as of November 2020, available at https://massdot.maps.arcgis.com/home/item.html?id=10a2766a607345928c6a66ffb479c937. 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.

Driver License Data

MassDOT provided driver's license data by age and town for the years 2011 through 2015. Due to a substantial jump in the number of registered drivers in the dataset between 2011 and 2012 (28%), VHB excluded 2011 and used an average number of drivers in each age group for the years 2012 through 2015. VHB then calculated the total, proportion, and density of young drivers (i.e., drivers aged 24 and under) for each city, town, metropolitan planning organization (MPO), and regional planning agency (RPA).

School Location Data

VHB obtained primary and secondary school location data from the Massachusetts Bureau of Geographic Information (MassGIS) open data portal (https://massgis.maps.arcgis.com/home/item.html?id=a7ccf184af704f5fbd17d69f935554d6). VHB only included schools with grades 10 through 12 for the purposes of this analysis.

College and University Data

VHB accessed college and university location data from the U.S. Department of Homeland Security's Homeland Infrastructure Foundation-Level Data (HIFLD) repository (https://hifld-geoplatform.opendata.arcgis.com/datasets/colleges-and-universities-campuses/explore?location=13.091953%2C0.317215%2C2.75). 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.

Citation Data

VHB obtained total traffic citation counts data by town for a four-year period between 2017 and 2020. These data included total citations, as well as subsets of counts for speeding-, seat belt-, impaired driving-, and distraction-related traffic citations.

Results

The following sections describe the results of the negative binomial regression modeling effort.

Variables of Interest

To account for unobserved influences due to road facilities and traffic exposure, VHB established a base model that included the natural log of total centerline mileage and the proportion of centerline mileage classified as arterials (i.e., other principal arterials and minor arterials) and interstates by town. Before including additional variables in the negative binomial, VHB developed a correlation matrix of input variables. Highly correlated variables are indicators of potential complications in the model development process. Although VHB considered all potential variables in this matrix, Table 11 shows the correlation matrix for the following 7 variables included in the final young driver model.

1. Natural log of total length of all centerlines in the town (miles).
2. Proportion of total centerline length that are arterials and interstates in the town (miles).
3. Total number of licensed young drivers in the town (1,000's).
4. Indicator that town has more young drivers per square mile than the governing MPO/RPA as a whole.
5. Density of secondary schools (per square mile).
6. Proportion of traffic citations involving a lack of seat belt use.

Table 11. Correlation matrix of input variables.

Variable	1	2	3	4	5	6
1	**1.00**					
2	**0.19**	**1.00**				
3	**0.68**	**0.20**	**1.00**			
4	**0.30**	**0.19**	**0.49**	**1.00**		
5	**0.27**	**0.18**	**0.54**	**0.46**	**1.00**	
6	**0.33**	**0.16**	**0.41**	**0.24**	**0.26**	**1.00**

Model Results

Table 12 documents the results of the final model. Total centerline length and the proportion of length on arterials and interstates are both positively correlated with an increase in young driver-related KA crashes. This likely reflects overall exposure. Indicator variables associated with the number and density of licensed young drivers were also positively associated with an increase in KA crash frequency. This includes the total number of registered drivers in a town, as well as the comparison of young driver density relative to the specific MPO/RPA in which a town is located. This demonstrates that a relative increase in young drivers in a town compared to the neighboring municipalities can indicate a high-risk community in the State (i.e., even if that town is below the statewide average in density of young drivers).

The density of secondary schools (i.e., those with grades 10 through 12) in a town is also positively correlated with an increase in young driver KA crash frequency; however, indicators of colleges and universities were not statistically significant in the young driver KA crash model. The proportion of all citations involving seat belt use show an intuitive correlation with KA younger driver crashes; a greater proportion of seat belt-related citations correlate with a higher number of younger driver-related KA crashes. These citation data serve as a surrogate measure of risk-taking behaviors by drivers in cities and towns, and they indicate that occupant protection risk factors could be an indicator of young driver risk.

Table 12. Negative binomial count regression model results.

Variable (Number)	Coefficient	Standard Error	z-value	P>|z|	95% Confidence Interval	
Natural log of the total length of all centerlines in the town (miles)	1.12	0.09	12.47	<0.001	0.94	1.29
Proportion of total centerline length that are arterials[6] or interstates in the town (miles)	4.54	0.57	8.03	<0.001	3.44	5.65
Total number of young drivers in the town (1,000's)	0.13	0.04	3.27	0.001	0.05	0.20
Indicator that town has more young drivers per square mile than the governing MPO/RPA as a whole	0.17	0.09	1.97	0.05	0.00	0.34
Density of secondary schools (per square mile)	0.65	0.21	3.01	0.003	0.23	1.07
Proportion of traffic citations involving a lack of seat belt use	7.79	1.62	4.80	<0.001	4.61	10.98
Constant	-4.63	0.41	-11.16	<0.001	-5.44	-3.82
Alpha	0.26	0.03	-	-	0.21	0.34

Note: Number of observations = 348; Log likelihood = -940.67627; Pseudo R2 = 0.2039; LR chi2(7) = 481.75; Prob > chi2 = 0.0000.

Conclusions and Recommendations

The purpose of this analysis is to identify segment-level risk factors for fatal and serious injury risk factors. VHB recommends that MassDOT disregard the coefficients in the negative binomial regression results in Table 12. Instead, MassDOT should assign binary risk factor scores if a characteristic is present on a focus segment (i.e., a 0 if it is not present and a 1 if it is present). For the one unique continuous variable in Table 13, the density of secondary schools, VHB recommends that MassDOT normalize the values for this variable and assign a percentile rank to each town in Massachusetts with the highest value being 1. For instance, if a town had a density of secondary schools greater than 87 percent of all towns in the State, it would be assigned a risk score of 0.87.

Since the crash type assessment underscored the prevalence of lane departure and speeding crash characteristics, VHB recommends that MassDOT use the final normalized values to represent the risk associated with these crash types, rather than the raw score. Table 13 summarizes the risk factors identified in this analysis.

[6] Other principal arterials and minor arterials.

Table 13. Summary of risk factors for young driver KA crashes.

Segment Risk Factors for Young Drivers	Type
Total number of young drivers in the town (1,000’s)	Continuous
Segment is in a town that has more young drivers per square mile than the MPO/RPA as a whole	Binary
Density of secondary schools (Grades 10 through 12; per square mile)	Continuous
Normalized occupant protection systemic risk factor score	Continuous
Normalized speeding systemic risk factor score	Continuous
Normalized roadway departure systemic risk factor score	Continuous

Table 14 provides an example application of the risk factors on a hypothetical segment. To provide context for these risk factor scores in relation to other emphasis areas as part of the SDI grant analysis, MassDOT can normalize the cumulative score of binary risk factors plus the value of the one continuous risk factor. This would generate a risk score of 100 percent if all risk factors for the facility type are present. Under this approach, the risk score for the example segment in Table 14 is 61.1 percent.

Table 14. Example risk score calculations for young driver-related crashes.

Variable	Segment Characteristic	Risk Factor	Risk Score
Population of Younger (Under 25) Driving Residents	1,384 licensed young drivers in the town	Greater than 65% of towns in the State	0.65
Density of Younger (Under 25) Driving Residents	105 licensed young drivers per sq. mi. in the town	Segment is in a town that has more young drivers per square mile than the governing MPO/RPA as a whole (152 licensed young drivers per sq. mi)	0
Density of Secondary Schools (Grades 10 through 12; per square mile)	0.11 secondary schools per square mile in the town	Greater than 70% of towns in the State	0.7
Normalized occupant protection systemic risk factor score (divided by 100)	75.6	Higher risk score associated with occupant protection-related KA crashes	0.76
Normalized risk score for speeding-related KA crashes (Divided by 100)	87.3	Higher risk score associated with speeding-related KA crashes	0.87
Normalized risk score for lane departure-related KA crashes (Divided by 100)	69.4	Higher risk score associated with lane departure-related KA crashes	0.69
		Total Risk Score:	3.67
		Risk Percent Score:	61.1%

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 are 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, segments ranked in the top 5 percentile (95 through 100) were categorized as "Primary Risk Site," segments 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 15 and Table 16 show the distribution of segments with the normalized risk score (presented as percentages) across these categories for Statewide and MPO rankings, respectively.

Table 15. 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	95.34%	100%	21400	6.9%
	Secondary Risk Site	85.10%	93.05%	25933	8.4%

Table 16. Distribution of risk sites by MPO.

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	44.34%	60.85%	1212	9.29%
	Secondary Risk Site	43.15%	44.34%	1228	9.42%
Boston Region MPO	Primary Risk Site	55.79%	72.31%	5304	5.05%
	Secondary Risk Site	52.66%	55.79%	10458	9.95%
Cape Cod Commission	Primary Risk Site	31.68%	43.52%	2349	8.68%
	Secondary Risk Site	28.29%	31.67%	2942	10.87%
Central Massachusetts Regional Planning Commission	Primary Risk Site	53.92%	69.55%	1669	5.66%
	Secondary Risk Site	52.07%	53.85%	4180	14.17%
Franklin Regional Council of Governments	Primary Risk Site	40.25%	54.31%	401	5.34%
	Secondary Risk Site	38.40%	40.18%	806	10.72%
Martha’s Vineyard Commission	Primary Risk Site	33.46%	43.95%	164	5.32%
	Secondary Risk Site	32.27%	32.86%	519	16.85%
Merrimack Valley Planning Commission	Primary Risk Site	52.90%	65.70%	826	5.52%
	Secondary Risk Site	50.46%	52.90%	1433	9.57%
Montachusett Regional Planning Commission	Primary Risk Site	48.86%	67.24%	809	5.04%
	Secondary Risk Site	46.48%	48.80%	1718	10.71%
Nantucket Planning and Economic Development Commission	Primary Risk Site	10.60%	19.91%	114	5.16%
	Secondary Risk Site	9.34%	10.53%	340	15.40%
Northern Middlesex Council of Governments	Primary Risk Site	53.15%	70.52%	760	6.25%
	Secondary Risk Site	51.30%	53.08%	2187	17.97%
Pioneer Valley Planning Commission	Primary Risk Site	56.28%	74.88%	1446	5.00%
	Secondary Risk Site	51.49%	56.26%	4442	15.36%
Old Colony Planning Council	Primary Risk Site	50.42%	67.26%	1967	10.61%
	Secondary Risk Site	49.22%	50.41%	1140	6.15%
	Primary Risk Site	53.36%	72.02%	1500	5.00%

MPO	Risk Category	Minimum Normalized Risk Score Percentage	Maximum Normalized Risk Score Percentage	Number of Segments	Percent of Scored MPO Segments
Southeastern Regional Planning and Economic Development District	Secondary Risk Site	51.00%	53.35%	2998	10.00%