Identifying Risk Factors of Fatal Motor Vehicle Collisions in Toronto, Canada Using Binary Logistic Regression

Abtin K. Shirvani

April 7, 2024

1 Introduction

1.1 Background

Toronto among many urban centres has experienced year-over-year growth of vehicles on the road, with the latest recorded peak in 2022 hitting 1.44 million active registered vehicles in Metro Toronto (Ministry of Transportation, 2022). Naturally, motor vehicle collisions (MVCs) have been a rising cause for concern in recent decades. Although the number of MVCs and deaths from MVCs have been on the decline, 45 people were still killed in a MVC in Toronto in 2023 and it is estimated that 94% of serious crashes are due in part to human error, suggesting there is room for improvement (City of Toronto, 2024).

Current literature on the subject has mostly targeted large infrastructural issues and potential policy changes that may take years or decades to properly implement (Siddiqui, 2022; Thompson & Vaz, 2023). However, some studies have chosen to focus on driver-related interventions and environmental factors that could serve as practical advice for the average person. For instance, the 2012 study by Rzeznikiewiz et al. examining fatal MVCs on Ontario's highways found that old age, time of day, and weather conditions were all associated with risk of a fatal MVC. Other studies found associations between crash characteristics (i.e. Head-on collisions, single-vehicle collisions), driver decision-making, young age, alcohol consumption and the risk of dying in an MVC (Zhang et al., 2000; Bédard et al., 2002).

For the most part, the road safety guidelines agree with past literature findings, however a lot has changed since then. Toronto's population has grown nearly 16% from 2011 to July 2022 (Statistics Canada, 2011; City of Toronto, 2023). Road safety initiatives such as the Vision Zero plan have been enacted, establishing speed limit reductions and senior safety zones (City of Toronto, 2024). Changes to infrastructure throughout the city including new bike lanes, pedestrian crossings, widening highways, and road reconstructions continue to be implemented. Therefore, although risk factors may very well remain the same, we might expect changes in the significance of previously established risk factors in modern-day Toronto.

1.2 Objective

The goal of this analysis is to identify risk factors associated with fatal MVCs in Toronto, Canada. This analysis is intended to inform the average person of the most important factors to take into consideration, to ultimately help guide the individual to take the necessary precautions to reduce their likelihood of dying in a MVC. As such, the focus whilst constructing the model will be on interpretation.

2 Methods

2.1 Dataset Description

The dataset was sourced from the Toronto Open Data Portal and collected by the Toronto Police Services. It contains all motor vehicle collisions where a person was either killed or seriously injured in the City of Toronto from 2006-2021. Exact locations of the crashes are masked for privacy concerns using the nearest intersection and an offset term, entailing distance and direction from collision. To reduce dimensionality of the dataset, subgroupings of several categorical covariates with low counts were grouped. Several covariates were omitted due to redundancy and/or missing data (removed before model fitting to reduce computational cost). The meaning of 'None' data points was found to be inconsistent across covariates, thus meaning was inferred on a case-by-case basis.

2.2 Variable Selection

Speeding, type of impact, age of those involved, pedestrian involvement and whether it was an alcohol-related crash were predictors of interest for this analysis with Accident Class (Fatal or Non-Fatal) as the response. Variables were selected using stepwise selection and LASSO. Bi-directional stepwise selection was run using AIC and BIC to select candidate models. LASSO was run using cross validation to determine the minimum lambda (where cross validation error is minimized), which was then used to calculate the regression coefficients. Those that were not shrunk down to 0 were selected as significant covariates. Given that most of the covariates in this dataset were categorical and multilevel, group LASSO would have been more suitable, however for the purposes of the course, standard LASSO was used (all covariates that had at least one significant subgrouping was deemed significant). In total, 3 logistical models were fit based on the variable selection outputs. Based on the literature, there was no evidence of clustering or nesting in this type of dataset, thus mixed effects were not included.

2.3 Model Diagnostics & Validation

The efficacy of candidate models was then assessed using Area Under the Curve (AUC) on ROC curves and Mean Absolute Error (MAE) calculated from cross validation. Cross validation was run using 10 repetitions and was visualized on a calibration plot along with MAE. Models with a lower MAE, higher AUC, and a better overall fit to the ideal 1:1 predicted to actual probability on the calibration plot were favoured. The final model was then checked for influential points using Cook's Distance with a threshold of 4/n-k-1, where n is the number of observations and k is the number of covariates. Observations with Cook's Distances that were substantially larger than the rest were evaluated on a case-by-case basis. General Variance Inflation Factors (GVIFs) and Adjusted GVIFs were used to check for multicollinearity. A reduced model was fit (omitting the multicollinear covariates) and assessed using the same diagnostics and validation, along with a Likelihood Ratio Test as an additional confirmatory test to compare the full and reduced models.

3 Results

3.1 Descriptive Data

There were 18194 observations of MVCs in the dataset collected – 2573 (14.1%) of which were fatal. Table 1 presents the distribution of predictors of interest in relation to accident class. The most common type of impact for MVCs is pedestrian collisions, making up 40.1% (7295) of all MVCs and 52% of all fatalities. Early adults (20-40 years old) are the age group most involved in MVCs, accounting for 33.2% of all MVCs. Large discrepancies between non-fatal and fatal MVCs among late adults (60 years and over) and pedestrians involved may suggest that the elderly as well as pedestrians are particularly susceptible to dying in an MVC (8 and 14% difference respectively). The same can be seen in speeding-related collisions, with a 9% difference in the proportion of non-fatal speeding-related collisions and fatal speeding-related collisions. All predictors of interest were found to be associated with accident class, as expected based on past literature ($P < 0.001$).

Table 1. Distribution of fatal and non-fatal MVCs by risk factor

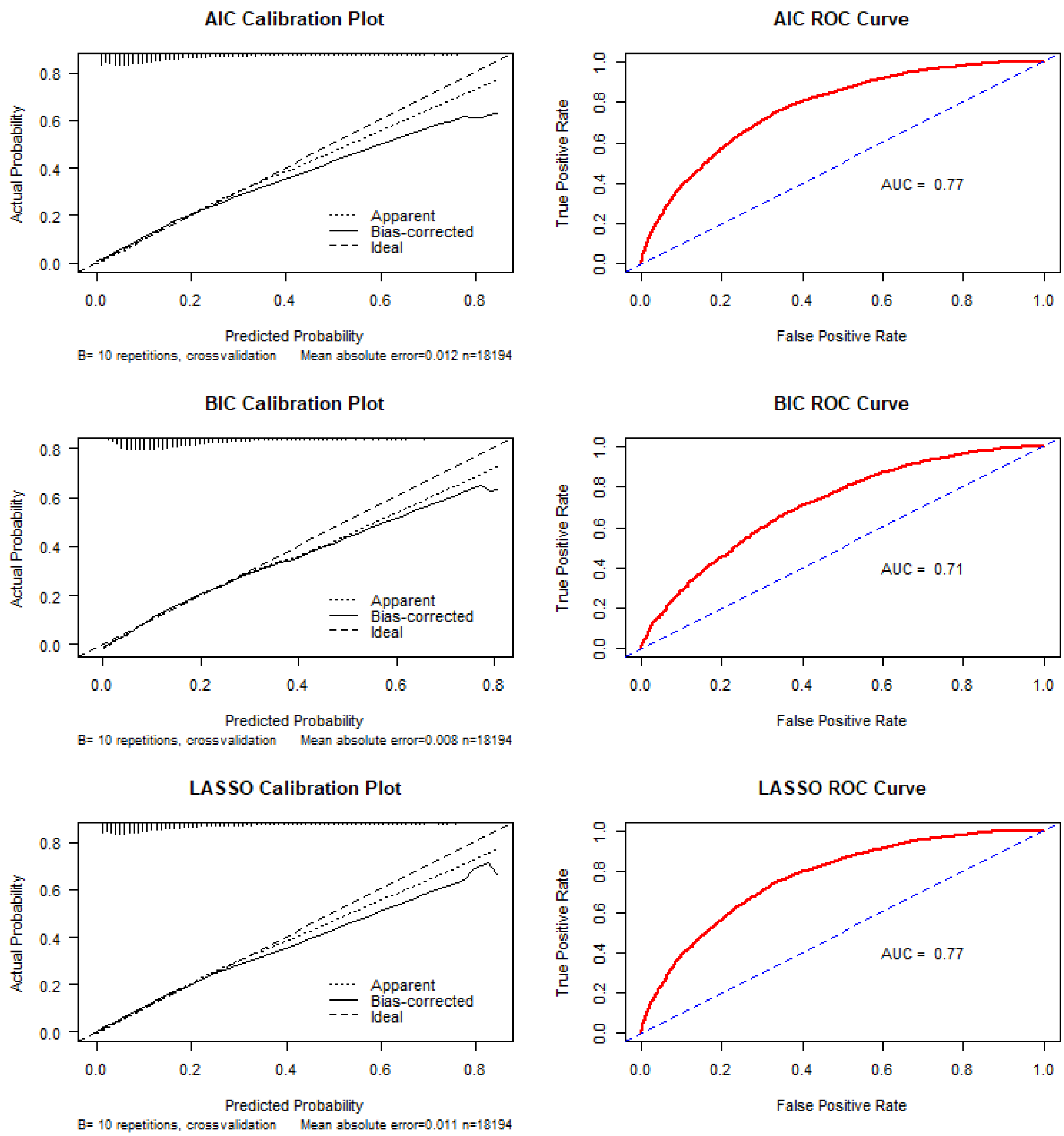
	Accident Class			
Risk Factor	**Non-Fatal**, N = 15,621[1]	**Fatal**, N = 2,573[1]	**Total**[2]	**p-value**[3]
Speeding-related				<0.001
None	13,615 (87%)	2,004 (78%)	15,619	
Yes	2,006 (13%)	569 (22%)	2,575	
Type of Impact				<0.001
Pedestrian Collisions	5,951 (38%)	1,344 (52%)	7,295	
Turning Movement	2,504 (16%)	288 (11%)	2,792	
Cyclist Collisions	1,671 (11%)	124 (4.8%)	1,795	
Rear End	1,620 (10%)	126 (4.9%)	1,746	
SMV	1,348 (8.6%)	302 (12%)	1,650	
Angle	1,121 (7.2%)	162 (6.3%)	1,283	
Approaching	755 (4.8%)	173 (6.7%)	928	
Sideswipe	470 (3.0%)	36 (1.4%)	506	
Other	181 (1.2%)	18 (0.7%)	199	
Age Group of Those Involved				<0.001
Adolescence	1,283 (8.2%)	194 (7.5%)	1,477	
Early Adulthood	5,275 (34%)	773 (30%)	6,048	
Mid Adulthood	4,277 (27%)	636 (25%)	4,913	
Late Adulthood	2,540 (16%)	607 (24%)	3,147	
Unknown	2,246 (14%)	363 (14%)	2,609	
Pedestrian Involved				<0.001
None	9,609 (62%)	1,231 (48%)	10,840	
Yes	6,012 (38%)	1,342 (52%)	7,354	
Alcohol-related				<0.001
None	14,979 (96%)	2,427 (94%)	17,406	
Yes	642 (4.1%)	146 (5.7%)	788	

[1] Count (%)
[2] Count
[3] Pearson's Chi-squared test

3.2 Model Selection

Calibration and ROC curves can be seen in Figure 2. The AIC and LASSO model both yielded AUCs of 0.77 and MAEs of 0.012 and 0.011 respectively. The BIC model yielded a better fitting calibration plot with an MAE of 0.008, but a lower AUC of 0.71 on the ROC curve compared to the other two candidate models. The BIC model was chosen for its more parsimonious variable selection and lower MAE.

Figure 2. ROC curves and Calibration Plots for Candidate Models



Adjusted GVIFs of the predictors were all around 1 except for pedestrian and cyclist, which were 17.4 and 10.4 respectively, indicating a violation in multicollinearity. The likelihood ratio test between the BIC model and the reduced model that was fit without pedestrian and cyclist was found to be statistically significant ($P < 0.05$), suggesting that pedestrian and cyclist are significant. The reduced model yielded virtually the same MAE of 0.008 and AUC of 0.71 as the Full BIC model, thus the reduced model was favoured.

Table 3. Final Logistic regression model results, representing the odds of fatality in a MVC in Toronto from 2006-2021.

Risk Factor	AOR[1]	95% CI[1]	p-value
Year	1.05	1.03, 1.06	**<0.001**
Weather Condition			
Clear	—	—	
Other	3.07	2.18, 4.30	**<0.001**
Rain	0.73	0.63, 0.85	**<0.001**
Snow	0.40	0.27, 0.58	**<0.001**
Level of Light			
Dark	—	—	
Dawn	1.32	0.91, 1.88	0.14
Daylight	0.60	0.54, 0.68	**<0.001**
Dusk	1.06	0.81, 1.38	0.7
Other	3.75	0.57, 31.4	0.2
Type of Impact			
Angle	—	—	
Approaching	1.64	1.25, 2.16	**<0.001**
Cyclist Collisions	0.73	0.54, 0.97	**0.029**
Other	0.37	0.21, 0.63	**<0.001**
Pedestrian Collisions	2.16	1.73, 2.72	**<0.001**
Rear End	0.52	0.39, 0.69	**<0.001**
Sideswipe	0.51	0.33, 0.76	**0.001**
SMV	1.60	1.25, 2.07	**<0.001**
Turning Movement	0.84	0.66, 1.07	0.15
Age of Those Involved			
Adolescence	—	—	
Early Adulthood	0.95	0.80, 1.14	0.6
Mid Adulthood	1.07	0.89, 1.28	0.5
Late Adulthood	1.81	1.50, 2.18	**<0.001**
Unknown	1.11	0.91, 1.36	0.3
Motorcycle Involved	1.55	1.30, 1.85	**<0.001**
Truck Involved	3.67	3.17, 4.25	**<0.001**
City Transit Vehicle Involved	2.06	1.74, 2.42	**<0.001**
Passenger involved	1.27	1.14, 1.41	**<0.001**
Speeding-related	2.92	2.54, 3.36	**<0.001**
Aggressive Driving-related	0.62	0.56, 0.69	**<0.001**
Redlight-related	1.54	1.26, 1.87	**<0.001**
Artificial light	0.60	0.52, 0.69	**<0.001**

[1] AOR = Adjusted Odds Ratio, CI = Confidence Interval

3.3 Final Model

Year was found to have a positive association with the odds of fatality in a MVC (AOR = 1.05). Poor weather conditions such as rain and snow yielded relatively low odds of fatality of dying in a MVC compared to clear conditions (AOR = 0.40, 0.73 respectively). When compared to the dark, daylight was the only category that showed significant differences. Collisions occurring during the daylight showed much lower odds of fatality (AOR = 0.6). When the type of collision falls under approaching, pedestrian collision, or Slow-Moving Vehicles, the odds of dying in a MVC increase, given angle collisions as the reference group (AOR = 1.64, 2.16, 1.6 respectively). As age increases, the odds of dying in a MVC generally increased, except for a slight dip in early adulthood. When compared to the adolescent group, only late adults were found to have a significant association with death in a collision (AOR = 1.81). Having a motorcycle, truck, or transit city vehicle involved in the collision was found to have significant positive associations with accident class – the most notable being trucks (AOR = 3.67). The odds of fatality increase even more when there is a passenger in the vehicle (AOR = 1.28). Driver-related error such as speeding, aggressive driving, and passing redlights all raised your odds of dying in a MVC, especially speeding at 2.92 AOR, except for aggressive driving (AOR = 0.62).

4 Discussion

4.1 Year

The results suggest a year-over-year increase in the odds of fatality. Although the total number of MVCs in Canada has been on the decline for the past few decades, the rate at which the non-fatal MVCs are decreasing seems to be outpacing the fatal MVCs, which may explain this association (Transport Canada, 2021). The exact cause is unclear; however, one possible explanation is that the sales of large-mass vehicles such as trucks and SUVs have increased in recent years, leading to more deadly collisions (Ministry of Transport, 2022; McCartt et al., 2004; Khattak, 2001).

4.2 Environmental Conditions

Snowy and rainy weather conditions showed a decreased odds of fatality, which agrees with the findings from Qiu & Nixon (2008) and Rzeznikiewiz et al. (2012). Rain and snow were both found to increase overall crash occurrence, however in terms of fatality, there was a negative association. One possible reason could be that people tend to drive slower and more attentively in adverse weather conditions (Qiu & Nixon, 2008). Collisions during the daylight showed much lower odds of fatality compared to the dark, which could be due to poor visibility, fatigue, and increased drug/alcohol consumption during the nighttime (Rice et al., 2003). The odds decreased even further when the light is artificial. One interpretation of this result is that artificial light improves visibility (i.e. Street lights during the night), which has been shown in Québec to reduce the collision frequency, however there didn't appear to be research on the fatality odds with artificial light (Matout, 2013).

4.3 Collision Type

The type of impact of the collision plays a huge role in fatality odds. Rear ends likely lead to one of the lowest odds of fatality (0.52) compared to angled collisions due to the point of contact being the furthest away from the driver (Khattak, 2001). Similarly, sideswipes are relatively minor collisions that typically don't generate enough force to kill, which explains the low odds (0.51)(Liu & Pressley, 2016). Conversely, approaching collisions generate much more force and thus, a more deadly collision (McCartt et al., 2004). Collisions involving pedestrians raised the odds of dying over 2 times relative to angled collisions, which is likely due to there being no vehicle to bear the impact from the collision (Chakravarthy et al., 2007). One would expect the same to be said for cyclists however the results imply that cyclists have a lower odds of dying compared to angled collisions. There are a couple explanations for this; (a) the result holds less power than the rest, so it may be disregarded ($P = 0.029$); (b) cyclists are still prone to dying in the collision, but angled collisions are simply more deadly; or (c) bikers tend to be younger, thus cyclist collisions is merely acting as a mediator for age (Schepers, 2017).

4.4 Vehicle Type

The results suggest that motor vehicles other than cars (motorcycles, trucks, city transit vehicles, slow-moving vehicles) all increase your odds of death, with the highest being trucks at 3.67, making it the most significant risk factor in determining death in a MVC. The 2008 Swedish study by Björnstig et al. found that large, heavy vehicles such as trucks and buses killed 5 times as many car occupants as cars did, mostly due to the smaller vehicle taking the brunt of the change in velocity during the collision. Studies in the US found similar results, specifically with Light-Truck and Vans (LTVs), where pedestrians were found to have up to 3 times the likelihood of dying given a LTV is involved (Lefler & Gabler, 2004; Tyndall, 2021). On the other hand, motorcyclist deaths have been largely attributed to a lack of protection compared to a car as well as risky behaviors such as speeding (Fagnant & Kockelman, 2015; Nunn, 2011).

4.5 Persons Involved

Having a passenger(s) in the vehicle was found to slightly increase the odds of fatality. The 1988 study by Evans & Frick found no significant differences in the odds of dying between left and right passengers and a significant decreased odds in the rear, suggesting that the passenger(s) has the same if not less of a fatality risk than the driver. However, from a statistical standpoint, the more people that are involved in a collision, the higher the likelihood that at least one of them will die, which could help explain this slight positive association. Although early adults are most likely to get in a MVC, only late adults were found to be significantly associated with dying; being 60 and over raises your odds of dying by 1.81 times compared to adolescents. These rseults agree with the findings by Zhang et al. (2000), which suggest that medical/physical conditions such as decline in cognitive function and poor visual acuity played a role in increasing odds of death for drivers 75 and up, with a similar albeit weaker association amongst drivers aged 65-74.

4.6 Driver Behaviour

Collisions involving running red lights and speeding were positively associated with fatality – speeding raising your odds of dying by nearly 3 times. Red light collisions tend to be right-angle or head-on collisions, which encapsulate more deadly collisions, and speeding has long been established as a major risk factor in collision severity (Erke, 2008; Zhang et al., 2000; Rice et al., 2003). Aggressive driving – usually a marker for car crashes – was negatively associated with fatality in the crash, which goes against previous findings by Paleti et al. (2010) who found that aggressive driving was positively associated with crash severity and acted as a mediator for young age and inexperience. The differing results could be attributable to their dataset including all collisions, while the one used for this analysis only captured severe and fatal crashes. Therefore, if this dataset included mild collisions as well, we would have likely seen similar results. Another hypothesis is that aggressive driving encapsulates a broader range of crashes that inherently include more mild collisions such as not yielding right-of-way, stopping on a pedestrian crosswalk, and tailgating, however additional information on the dataset measures would be needed (Ministry of Transportation, 2024).

4.7 Limitations

While the insights for the most part agree with the literature, the dataset only collected motor vehicle collisions that resulted in a death or a serious injury, which represents a small subset of collisions that happen day-to-day. Thus, it is important to keep the interpretation of the results strictly within the confines of estimating fatality odds amongst severe MVCs and not all MVCs. To be able to generalize the results, a dataset containing all types of accidents, including minor ones, would have to be used. Furthermore, other than year, the date was omitted from the model because it was assumed that most of the variation/potential seasonality from date was encapsulated by other covariates such as visibility (rain, snow, etc.), which may not hold. Additionally, some meanings were inferred, specifically none values, which could have been inaccurate and skew the regression coefficients. There were a few extreme outliers that upon inspection seemed plausible (i.e. Truck and speeding involved, but no death) and not due to some measurement/input error, which were kept in the dataset with the notion that it may introduce some bias.

References

1. Bédard, M., Guyatt, G. H., Stones, M. J., & Hirdes, J. P. (2002). The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. *Accident; analysis and prevention*, *34*(6), 717–727. https://doi.org/10.1016/s0001-4575(01)00072-0

2. Björnstig, U., Björnstig, J., & Eriksson, A. (2008). Passenger car collision fatalities – with special emphasis on collisions with heavy vehicles. *Accident Analysis & Prevention, 40*(1), 158-166. https://doi.org/10.1016/j.aap.2007.05.003

3. Chakravarthy, B., Lotfipour, S., & Vaca, F. E. (2007). Pedestrian injuries: emergency care considerations. *Western Journal of Emergency Medicine, 8*(1), 15-21. PMID: 20440388; PMCID: PMC2859736.

4. City of Toronto. (2023). *Toronto at a Glance*. Retrieved from toronto.ca: https://www.toronto.ca/city-government/data-research-maps/toronto-at-a-glance/

5. City of Toronto. (2024). *Vision Zero Road Safety Plan*. Retrieved from toronto.ca: https://www.toronto.ca/services-payments/streets-parking-transportation/road-safety/vision-zero/

6. Erke, A. (2008). Red light for red-light cameras?: A meta-analysis of the effects of red-light cameras on crashes. *Accident Analysis & Prevention, 41*(5), 897-905. https://doi.org/10.1016/j.aap.2008.08.011

7. Evans, L. & Frick, M. C. (1988). Seating position in cars and fatality risk. *American Journal of Public Health, 78*(11), 1456-1458. https://doi.org/10.2105/AJPH.78.11.1456

8. Fagnant, D. J. & Kockelman, K. M. (2015). Motorcycle Use in the United States: Crash Experiences, Safety Perspectives, and Countermeasures. *Journal of Transportation Safety & Security*, *7*(1), 20–39. https://doi.org/10.1080/19439962.2014.894164

9. Khattak, A. J. (2001). Injury Severity in Multivehicle Rear-End Crashes. *Transportation Research Record, 1746*(1), 59-68. https://doi.org/10.3141/1746-08

10. Lefler, D. E. & Gabler, H. C. (2004). The fatality and injury risk of light truck impacts with pedestrians in the United States. *Accident Analysis & Prevention, 36*(2), 295-304. https://doi.org/10.1016/S0001-4575(03)00007-1

11. Liu, C. & Pressley, J.C. (2016). Side impact motor vehicle crashes: driver, passenger, vehicle, and crash characteristics for fatally and nonfatally-injured rear-seated adults. *Injury Epidemiology, 3*(23). https://doi.org/10.1186/s40621-016-0088-1

12. Matout, N. (2013). Estimation of the Influence of Artificial Roadway Lighting on Road Collision Frequency. *Masters thesis, Concordia University.*

13. McCartt, A. T., Northrup, V. S., & Retting, R. A. (2004). Types and characteristics of ramp-related motor vehicle crashes on urban interstate roadways in Northern Virginia. *Journal of safety research*, *35*(1), 107–114. https://doi.org/10.1016/j.jsr.2003.09.019

14. Ministry of Transportation (2022). *Vehicle population data.* Retrieved from data.ontario.ca: https://data.ontario.ca/dataset/vehicle-population-data/resource/7ff95539-e7b2-45c7-b2bb-89070ca739cf

15. Ministry of Transportation (2024). *Speeding and aggressive driving*. Retrieved from ontario.ca: https://www.ontario.ca/page/speeding-and-aggressive-driving

16. Nunn, S. (2011). Death by Motorcycle: Background, Behavioral, and Situational Correlates of Fatal Motorcycle Collisions. *Journal of Forensic Sciences, 56(*2), 429-437. https://doi.org/10.1111/j.1556-4029.2010.01657.x

17. Paleti, R., Eluru, N., & Bhat, C. R. (2010). Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. *Accident Analysis & Prevention, 42*(6), 1839-1854. https://doi.org/10.1016/j.aap.2010.05.005

18. Qiu, L., & Nixon, W. A. (2008). Effects of Adverse Weather on Traffic Crashes: Systematic Review and Meta-Analysis. *Transportation Research Record, 2055*(1), 139-146. https://doi.org/10.3141/2055-16

19. Rice, T. M., Peek-Asa, C., & Kraus, J. F. (2003). Nighttime driving, passenger transport, and injury crash rates of young drivers. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention*, *9*(3), 245–250. https://doi.org/10.1136/ip.9.3.245

20. Rzeznikiewiz, D., Tamim, H., & Macpherson, A. K. (2012). Risk of death in crashes on Ontario's highways. *BMC Public Health, 12*(1125). https://doi.org/10.1186/1471-2458-12-1125

21. Schepers, P., Stipdonk, H., Rob Methorst, & Jake Olivier (2017). Bicycle fatalities: Trends in crashes with and without motor vehicles in The Netherlands. *Transportation Research Part F: Traffic Psychology and Behaviour, 46*(B), 491-499. https://doi.org/10.1016/j.trf.2016.05.007

22. Siddiqui, A.B. (2022). Congestion on Canada's Busiest Highway, 401 Problems, Causes, and Mitigation Strategies. In: Akhnoukh, A., *et al.* Advances in Road Infrastructure and Mobility. IRF 2021. Sustainable Civil Infrastructures. Springer, Cham. https://doi.org/10.1007/978-3-030-79801-7_73

23. Statistics Canada. (2011). *Statistics Canada*. Retrieved from statcan.gc.ca: https://www12.statcan.gc.ca/census-recensement/2011/as-sa/fogs-spg/Facts-csd-eng.cfm?Lang=eng&GK=CSD&GC=3520005

24. Thompson, A. & Vaz, E. (2023). Spatial Analysis of Characteristics and Influencing Factors of Killed or Seriously Injured Persons from Motor Vehicle Collisions Within the City of Toronto. In: Regional and Urban Change and Geographical Information Systems and Science. Advances in Geographic Information Science. Springer, Cham. https://doi.org/10.1007/978-3-031-24731-6_4

25. Transport Canada. (2021). *Government of Canada*. Retrieved from https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2021

26. Tyndall, J. (2021). Pedestrian deaths and large vehicles. *Economics of Transportation, 26–27*(100219). https://doi.org/10.1016/j.ecotra.2021.100219

27. Zhang, J., Lindsay, J., Clarke, K., Robbins, G., & Mao, Y. (2000). Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario. *Accident; analysis and prevention*, *32*(1), 117–125. https://doi.org/10.1016/s0001-4575(99)00039-1