IDENTIFYING KEY FACTORS AFFECTING CRASH SEVERITY IN TORONTO USING AN ORDERED LOGIT MODELLING APPROACH

Lu Li, Md Sami Hasnine, Khandker M Nurul Habib, Bhagwant Persaud, Amer Shalaby

Introduction

Over the years, traffic safety modelling has gained increasing attention among researchers and practitioners. The ability to evaluate roadway safety performance could provide policy makers invaluable understandings of how roadway design and human factors influence traffic safety. Past research has demonstrated the capabilities of evaluating crash frequency and the leading explanatory factors (Abdel-Aty and Radwan, 2000, Persaud and Lyon, 2007, Sayed and De Leur, 2008). However, an aspect of equal importance in the field of traffic safety is the ability to evaluate injury severity as a result of a collision.

In the past decades, Canada has seen a nationwide decreasing trend in the number and percentage composition of fatality and serious injuries as a result of traffic crashes. In 2012, there were 12,817 reported fatalities and serious injuries (6.47% of all reported injuries) nationwide. This was a significant reduction relative to 20 years ago, when there was 27,517 reported fatalities and serious injuries (9.65% of all reported injuries (Transport Canada, 2012)). Such noteworthy progress clearly suggests that countermeasures can be taken to reduce crash severities in addition to reducing crash frequency.

This study focuses on the traffic collisions that occurred between 2006 and 2010 in the city of Toronto. The study area has the largest road network amongst all Canadian metropolitan areas.

The objective of this paper is to explore a modelling tool to study crash severity that can present to policy makers the key contributing factors to more severe crashes in a highly urbanized city such as Toronto. The rich collision database recorded by the City of Toronto's Traffic Safety Unit allowed for a comprehensive analysis of various contributing factors. The factors being considered include (1) behavioural attributes (2) roadway characteristics, (3) environmental characteristics, and (4) crash characteristics. Before presenting the modelling structure and the results, past literature in the field of modelling crash severity is reviewed.

Literature Review

Under the motivation of improving social welfare, a significant amount of effort has been devoted to the field of safety modelling around the world. Crash severity has been modelled by past researchers in many ways and at various levels of complexity.

The more conventionally used approach involves applying regression modelling to study the effects of explanatory variables on the dependent variable (i.e. crash severity). For example, Bedard et al. (2001) applied multivariate logistic regression to reveal that higher crash severity is strongly associated with increasing age, female gender, alcohol involvement, excessive speed, driver-side impact, and unbelted behaviour. They were able to simultaneously evaluate the correlations amongst a large array of explanatory variables and each variable's marginal effect on crash severity.

Abdel-Aty (2003) applied an ordered probit modelling structure to study the effects of roadway and individual-specific characteristics on crash severity. In this model the author concluded that more severe injuries are associated with increasing age, male gender, speeding, unbelted behaviour, roadway curvature, and dark lighting condition. Another interesting finding was that the driver at fault is surprisingly the party having less probability of injury. Conversely this indicates that the more innocent party involved in the crash is more prone to severe injury.

More applications of ordered logit and probit modelling can be seen in the literature examining various different datasets around the globe. The explanatory variables being investigated in each study depended heavily on the availability of the data and the credibility of the data collection process. Some of the commonly seen variables being tested included age, gender, crash impact type, lighting condition, road surface condition, alcohol involvement, vehicle type, and speed. Some less common attributes investigated include seating position, vehicle weight, time of collisions, day of collisions, etc. (Donnell and Connor, 1996. Kockelman and Kweon 2001. Kockelman and Wang, 2005. Garrido et al., 2014. Rifaat and Chin, 2004). As a general consensus, past research has demonstrated that severe injuries are closely associated with alcohol, poor lighting condition, and speed. Age and gender are two demographic variables that have reached mixed conclusions amongst different studies. Other variables that were introduced earlier could not reach a common consensus either because the variables were insignificant in some of the studies or were not tested due to unavailability at the data source.

The ordered regression model has also been used for examining pedestrian and bicyclist injury crashes which, though less frequent than vehicle-vehicle crashes, can be more severe because of the vulnerability of these road users. For example, Eluru et al. (2007) used a more sophisticated mixed generalized ordered response model to study the effects of contributing factors on crashes involving pedestrians and bicyclists. The study concluded that higher age, speed limit, signalized intersections, and darker lighting condition are strongly associated with more severe injury (Eluru et al., 2007).

Modelling Structure

In this study, an ordered logit modelling structure is applied. Ordered Logit Models (OLM) belongs to the family of discrete choice models. In discrete choice modelling, the formulation conditions on a set of utility functions such that each individual ultimately “selects” the choice rationally and in accordance with the most favourable utility. Associated with the utility function is also an unobservable error term distribution. In the discrete choice literature, the error term is commonly assumed to be following the logistic or normal distribution, resulting, respectively, in the formulation of logit and probit models. The results generated from logit and probit models in modelling crash severity were found to be qualitatively similar (Donnell and Connnor, 1996).

Ordered logit models account for the ordinal nature of injury severity, such that each layer of injury severity is progressively more damaging than its preceding layer. In addition, ordered models overcomes a crucial undesirable property, known as the Independence of Irrelevant Alternatives (IIA), which exists in the more conventionally used multinomial choice models (Ben-Akiva and Lerman, 1985).

In the formulation of ordered logit models, instead of having a set of utility functions for each alternative (as is the case for multinomial logit models), only one utility function is present. In terms of crash severity modelling, the latent utility can be analogized with the risks associated with the crash. Eluru et al. (2007) termed this risk as the "injury risk propensity". Conventionally, and in much of the literature discussed earlier, the latent utility measured the risks associated with each victim involved in this crash (some studies were unclear in this regards). However in this paper, the risks are defined to be associated with the overall risks of the crash. This is to correspond with our definition of overall crash severity, which measures the most severe injury experienced amongst all victims within the same crash. The determination of the overall crash severity is introduced in further detail in the next section.

It is also worth noting that unlike traditional discrete choice modelling, individuals cannot actively select an alternative in a traffic crash. The probability of each crash severity level is ultimately determined as a resulting combination of behaviour attributes, roadway characteristics, environmental characteristics and crash characteristics (i.e. explanatory variables of the utility function) associated with the crash. The utility function for this study is represented by:

$$Y_i^* = \beta' X_i + \varepsilon_i$$

where Y_i^* is the continuous latent utility associated with each crash occurrence i, X_i is the [N by 1] vector including all independent variables to be tested in the model, $\beta'$is the corresponding [N by 1] vector containing the parameters to be estimated, and ε_i is the error term distribution that follows logistic distribution for logit models.

The utility Y_i^* describes the latent, continuous and unobservable degree of risks associated with the crash severity level. Each crash does not provide this information directly, but rather as a set of discrete levels of crash severity, denoted here as Y_i. The discrete utility variable Y_i can be determined from the model by the following translation:

$$Y_i = \{ \begin{array}{lll} 0 & if -\infty \le Y_i^* \le \lambda_0 & [no\ injury], \\ 1 & if\ \lambda_0 < Y_i^* \le \lambda_1 & [minimal\ injury], \\ 2 & if\ \lambda_1 < Y_i^* \le \lambda_2 & [minor\ injury], \\ 3 & if\ \lambda_2 < Y_i^* \le \lambda_3 & [major\ injury], \\ 4 & if\ \lambda_3 < Y_i^* \le \infty & [fatality], \end{array}$$

Where the λ_j illustrates the threshold parameters to be estimated from the model. The resulting probability of each level j of the crash severity is then determined in accordance with the cumulative density function of the error term $F(\varepsilon_i)$ which again, for logistic regressions, follows a logistic distribution (Train, 2009):

$$\begin{aligned} Pr(Y_i = j) &= \Pr\big(\lambda_{j-1} < Y_i^* \le \lambda_j\big) \\ &= \Pr\big(\lambda_{j-1} < \beta' X_i + \varepsilon_i \le \lambda_j\big) \\ &= \Pr\big(\lambda_{j-1} - \beta' X_i < \varepsilon_i \le \lambda_j - \beta' X_i\big) \\ &= \Pr\big(\varepsilon_i \le \lambda_j - \beta' X_i\big) - \Pr\big(\varepsilon_i \le \lambda_{j-1} - \beta' X_i\big) \\ &= \mathrm{F}\big(\lambda_j - \beta' X_i\big) - \mathrm{F}\big(\lambda_{j-1} - \beta' X_i\big) \\ &= \frac{e^{\lambda_j - \beta' X_i}}{1 + e^{\lambda_j - \beta' X_i}} - \frac{e^{\lambda_{j-1} - \beta' X_i}}{1 + e^{\lambda_{j-1} - \beta' X_i}} \end{aligned}$$

Note that for $Y_i = 0$, $\mathrm{F}\big(\lambda_{j-1} - \beta' X_i\big) = F(-\infty) = 0$ and for $Y_i = 4$, $\mathrm{F}\big(\lambda_j - \beta' X_i\big) = F(\infty) = 1$.

The parameters of the ordered logit model described above are estimated using the method of Maximum Likelihood Estimation (MLE). MLE is a widely used parameter estimation technique in statistics and seeks the global maximum likelihood of observing the data being studied. The formulation of the MLE for the ordered choice models is such that (Greene and Hensher 2009):

$$LogL = \sum_{i=1}^{N}\sum_{j=1}^{K} \delta_{ij}\log[F(\lambda_j - \beta' X_i) - F(\lambda_{j-1} - \beta' X_i)]$$

where $\delta_{ij}$is a dummy variable such that $\delta_{ij} = 1$ for $Y_i = j$ and $\delta_{ij} = 0$ otherwise, K is the total number of crash severity level, and N is the total number of crash occurrences. Thus, for Ordered Logit Models, the MLE follows that:

$$LogL = \sum_{i=1}^{N}\sum_{j=1}^{K} \delta_{ij}\log[\frac{e^{\lambda_j - \beta' X_i}}{1 + e^{\lambda_j - \beta' X_i}} - \frac{e^{\lambda_{j-1} - \beta' X_i}}{1 + e^{\lambda_{j-1} - \beta' X_i}}]$$

In this paper the MLE is estimated using R such that the global maxima is reached (R Development Core Team, 2008).

Data Specification

This study examines the crash severity levels of all crashes that occurred between 2006 and 2010 (inclusive) within the city of Toronto. The data are administered by the City of Toronto and are a complete record of all the reported crashes that occurred within the city. Toronto is the largest city in Canada and unsurprisingly has a very rich dataset of crash records. Within the study period, there were 116,663 recorded crashes. However, one shortcoming with the dataset is that it does not differentiate which vehicles, when there are more than two involved, are the primary parties for causing the crash. For this reason, in order to more appropriately study the influencing factors, crashes that involve more than 2 vehicles are excluded from this study. This reduces the sample size to 106,324 crashes, of which 92,204 are two-vehicle crashes and 14,120 are single vehicle crashes.

The data were by default recorded in person-based format, such that each observation represents one victim involved in a crash. A total of 266,875 victims were recorded. The dataset recorded every victim involved in the crashes, regardless of the victim's final injury level or whether he or she is in the driver's seat. Drivers, passengers and, on some occasions, bystanders could not be systematically differentiated within the dataset to separately study each category. Thus, several victims (with a mean of 2.51 involved victims per crash) were needed altogether to paint one complete picture of one crash record.

As a result of multiple observations representing one crash incident, any correlation amongst the involved victims needed to be carefully handled before reaching a conclusion. To account for this, we defined an overall crash severity level for each crash that is represented by the worst injury severity level suffered amongst all victims of the crash. Originally the analysis was also considered to be done on individual-based disaggregate level. However, due to the inability to differentiate passengers from drivers, who are presumably held more legally responsible towards causing the crashes, correlation and biasness within the dataset could not be properly handled.

Of the 106,323 sample size of crashes, 82,509 (77.60%) were classified as no injury ($Y_i = 0$), 14,339 (13.49%) were classified as minimal injury ($Y_i = 1$), 8,615 (8.10%) were classified as minor injury ($Y_i = 2$), 782 (0.74%) were classified as major injury ($Y_i = 3$), and 79 (0.07%) were classified as fatality ($Y_i = 4$). The dataset presented a skewedness toward less severe crashes, which is a natural phenomenon in traffic safety. However, the underrepresentation of the more severe crashes was internally handled by the ordered logit model and did not deteriorate the statistical validity of the model results.

Associated with each observation was a set of attributes that include behavioural attributes, roadway characteristics, environmental attributes, and crash characteristics. Clearly the roadway, environmental, and crash characteristics (such as lighting condition and road surface condition) were shared for every victim involved in the same crash. This phenomenon is another reason for aggregating the victims into crash-based analysis to avoid unneeded duplications of the same explanatory variables.

Not all characteristics in the data were fully recorded. Characteristics that had excessive missing data (>50% missing) were excluded from the model. In addition, all characteristics were introduced into the model as dummy variables (such that 1 represents yes and 0 represents no) as shown below:

Table 1 Data Set Description

Explanatory Variables	Mean	S.D.
Day of the Week		
Monday	0.1399	0.3469
Tuesday	0.1544	0.3613
Wednesday	0.1585	0.3652
Thursday	0.1565	0.3633
Friday	0.1661	0.3722
Saturday	0.1284	0.3346
Sunday	0.0961	0.2948
Accident Time		
AM Peak (6AM – 9AM)	0.1147	0.3186
PM Peak (3PM – 6PM)	0.2480	0.4318
Midday (9AM – 3PM)	0.3461	0.4757
Evening (6PM – 12AM)	0.1597	0.3663
Night (12AM – 6AM)	0.1315	0.3380
Traffic Control Scheme		
Traffic Signal	0.6742	0.4687
Stop Sign	0.0233	0.1509
Yield Sign	0.0038	0.0617
Pedestrian Crossover	0.0103	0.1010
No Control	0.2796	0.4488
Impact Type		
Approach	0.0290	0.1677
Angle	0.1617	0.3682
Rear End	0.3544	0.4783
Side Swipe	0.1336	0.3402
Turning Movement	0.1833	0.3869
Single Vehicle	0.0646	0.2458
Visibility Condition		
Clear	0.8428	0.3640
Rain	0.1065	0.3084
Snow	0.0455	0.2083
Lighting Condition		
Daylight	0.2302	0.4210
Dawn or Dusk	0.0056	0.0748
Dark	0.0875	0.2826
Road Surface Condition		
Dry	0.7898	0.4074
Wet	0.1624	0.3688
Snow	0.0351	0.1841
Ice	0.0102	0.1005

Other Condition		
Involves Speeding	0.0068	0.0820
Involves Illegal Driving (Violating Traffic Laws)	0.1216	0.3268
Involves Driving under Alcohol Influence	0.0140	0.1177
Involves Inattentive Driving	0.0694	0.2541
Involve Heavy Vehicle	0.0233	0.1508
Involve Pedestrian	0.0399	0.1957
Involve Cyclist	0.0190	0.1367

Note that the last set of dummy variables ("Other Condition") is individual-specific as was originally recorded in the dataset. However to study the crash-based effect, these dummy variables were defined to be true (i.e. =1) for as long as one of the victims involved in that crash satisfies the criteria of that variable.

Model Results

In this study, crash-based data have been used to estimate the ordered logit model. The model was estimated for three types of dataset: (1) city-wide crashes, (2) intersection-based crashes and (3) mid-block-based (non-intersection) crashes. Table 2, Table 3 and Table 4, respectively, reveal the results of these three models. Most of the non-significant variables were excluded from the models; those non-significant variables remaining in the models were intentionally kept for comparison purposes. Some variables had a significant effect only in certain models (but not in others) and thus were kept. Coefficients of determination Rho-Squared (ρ^2) and Log likelihood values were also provided with each model result.

Table 2 Model Result for All Types of Crashes

Explanatory Variables	Coefficient	P-value.
Friday	0.0391	0.0673
AM Peak	-0.0679	0.0078
Single Moving Vehicle	-0.0631	0.0587
Dark	0.0616	0.0268
IF Involve Speeding	1.5566	<0.0001
IF Involves Illegal Driving	1.0098	<0.0001
IF Involves Alcohol Driving	1.0587	<0.0001
IF Involves Inattentive Driving	1.5912	<0.0001
IF Involve Heavy Vehicle	0.1222	0.0151
Involve Pedestrian	2.7307	<0.0001

Involve Cyclist	2.5236	<0.0001
Thresholds		
λ_0	1.8156	<0.0001
λ_1	3.2887	<0.0001
λ_2	6.3002	<0.0001
λ_3	8.7639	<0.0001
Number of Observation	106,324	
Log likelihood	-63139.16	
Coefficient of determination (ρ^2)	0.277	

Table 3 Model Result for Intersection-Based Crashes

Explanatory Variables	Coefficient	P-value.
Friday	0.0487	0.0600
AM Peak	-0.0749	0.0163
Evening	-0.0317	0.2444
Single Moving Vehicle	-0.0825	0.1553
Dark	0.0570	0.0631
Snow	-0.0788	0.1509
IF Involve Speeding	1.6037	<0.0001
IF Involves Illegal Driving	1.0234	<0.0001
IF Involves Alcohol Driving	0.9847	<0.0001
IF Involves Inattentive Driving	1.5621	<0.0001
IF Involve Heavy Vehicle	0.1280	0.0278
Involve Pedestrian	2.6990	<0.0001
Involve Cyclist	2.5107	<0.0001
Thresholds		
λ_0	1.7993	<0.0001
λ_1	3.2703	<0.0001
λ_2	6.2682	<0.0001
λ_3	8.7100	<0.0001
Number of Observation	71,685	
Log likelihood	-42879.78	
Coefficient of determination (ρ^2)	0.277	

Table 4 Model Result for Mid-block-Based crashes

Explanatory Variables	Coefficient	P-value
Tuesday	0.0634	0.1274

Evening	0.0803	0.0552
Turning Movement	-0.0651	0.1854
Dawn or Dusk	0.9135	0.0099
Dark	0.1330	0.1779
IF Involve Speeding	1.4064	<0.0001
IF Involves Illegal Driving	0.9456	<0.0001
IF Involves Alcohol Driving	1.1606	<0.0001
IF Involves Inattentive Driving	1.6727	<0.0001
Involve Pedestrian	2.8255	<0.0001
Involve Cyclist	2.5562	<0.0001
Thresholds		
λ_0	1.8562	<0.0001
λ_1	3.326	<0.0001
λ_2	6.405	<0.0001
λ_3	8.8627	<0.0001
Number of Observation	29,723	
Log Likelihood	-17349.44	
Coefficient of determination (ρ^2)	0.276	

Model Results for All Types of Crashes

Table 2 reveals the ordered logit model estimation results for all types of crashes. A wide array of variables was tested including: the day of the week, accident time, traffic control scheme, impact type, visibility condition, lighting condition, road surface condition, behavioural attributes, and crash characteristics. Here, the dependent variable represents the level of severity (0 = no injury, 1 = minimal injury, 2 = minor injury, 3 = major injury, and 4 = fatality). Therefore, positive coefficients of explanatory variables indicate an increased probability of higher order severity. Conversely negative coefficients indicate an increased probability of lower order severity. All explanatory variables for days of the week are tested in the model. However only "Friday" is statistically significant (at 90% confidence) and shows positive coefficient, which indicates that on Friday, there exists a higher probability of more severe crashes. Various times of the day segments are tested here also. Interestingly only "AM peak" is found to be significant. "AM peak" is showing a negative coefficient, depicting a higher probability of less severe crashes during the morning peak hour, perhaps suggesting a mitigating influence of congestion. As expected, crash type of single-moving vehicle shows a negative coefficient, representing a lower probability of severe crashes if the crash involves single vehicle. Unexpectedly, none of the other impact type variables, such as rear-end, approach, and side-swipe, shows statistical significance. The coefficient is positive for dark condition, indicating more severe accident is associated with darker environments. Finally, several behavioural actions are tested in the model such as if any victim involved in the crash was speeding, making illegal movement, driving under alcohol influence, driving inattentively, and if any heavy vehicle was involved in the crash. All five of these behavioural variables are showing positive coefficients, indicating positive association with more serious accident. Finally, the model shows involvement of pedestrian and cyclist is associated with higher crash severity, confirming the vulnerability of pedestrians and cyclists in traffic crashes.

Model Results for Intersection Based Crashes

Intersection based modelling estimation results in general follow a similar pattern to those from the city-wide model (which included crashes at both intersections and mid-blocks). “Friday” and “AM peak” are found significant again and they are also showing same signs as before. Single moving vehicle accident is no longer significant in the model. However, it is showing a negative coefficient in accord with the model for city-wide crashes. Dark lightning condition is again associated with more severe crashes whereas snow is not significantly associated with severity. Similar to the city-wide model, the five behavioural variables (if any victim was involved in the crash, speeding, making illegal movement, driving under alcohol influence, driving inattentively, and if any heavy vehicle was involved) are found to be statistically significant. Their positive coefficient indicates that they contribute to a higher probability of causing more severe crashes. Pedestrians and cyclists are again found to be associated with more severe crashes.

Model Result for Mid-Block Based Crashes

The mid-block model results are somewhat different from the other models presented above. “Friday” is determined to be no longer significant, whereas “Tuesday” becomes the only significant variable amongst the variable set containing days of the weeks. “Evening” is significant and shows a positive sign indicating a higher probability of more severe crashes during the evening. Surprisingly, dawn and dusk lighting condition is positively significant here whereas dark lighting condition is no longer significant. Involvement of heavy vehicle is also no longer statistically significant in this model. The fact that the heavy vehicle factor was significant in the intersection-based model but not here suggests that heavy vehicles may be more dangerous in situations when they are more aggressively accelerating/braking (i.e. at intersections). The other four behavioural variables (if any victim was speeding, making an illegal movement, driving under alcohol influence, driving inattentively) are again found statistically significant. Again, the results confirm that pedestrians and cyclists are more prone to severe injuries.

Conclusion

In this study, crash data and ordered logistic regression are used to investigate how to understand factors contributing to the severity of all crashes, intersection-based crashes and mid-block-based crashes. The results indicate differences between the contributing factors for the intersection-based crashes and mid-block-based crashes. It is found that impact type (approach, angle, rear end, side swipe, turning movement, and single vehicle) and traffic control scheme (traffic signal, stop sign, yield sign, pedestrian crossover, and no control) have insignificant effect in resulting crash severity. AM peak is negatively associated with severe crashes in all three models; Friday is positively associated with severe crashes at intersections and Tuesday is negatively associated with severe crashes at mid-block. In addition, darker environment is associated with more severe crashes for all models. Various behavioural characteristics are found to be highly significant in all three models. It is clear that speeding, driving illegally, under alcohol influence, or driving inattentively are associated with more severe crashes. Also it is confirmed that pedestrians and cyclists are more prone to severe crashes.

Most of the results are in accord with intuition and support the validity of the data and the methods used. At the same time, this also suggests that the usefulness of the results is limited. It is hoped, however, that this initial enquiry, using data from a large city, will prompt further research that can be more helpful for planners in such cities to effectively investigate and manage the triggering conditions associated with crash severity. To this end, for example, more disaggregate models can be estimated in the future by assembling available data to explicitly distinguish drivers, passengers, and even bystanders. Additionally, to capture the latent segmentation in the model at disaggregate level, when multiple non-drivers are involved in same accident, multivariate ordered probability model formulation can be tested in future.

Acknowledgement

The authors acknowledge the City of Toronto for providing the dataset used for this study. The first and second authors would also like to thank Mohamed Mahmoud for his valuable comments. Funding of the student authors from Natural Sciences and Engineering Discovery grants is appreciated.

References

Abdel-Aty, M. (2003). Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models. Journal of Safety Research, 34(5), 597–603. doi:10.1016/j.jsr.2003.05.009

Abdel-aty, M. A., & Radwan, A. E. (2000). Modeling Traffic Accident Occurrence and Involvement. Accident Analysis and Prevention 32, 633–642.

Bedard, 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, 717–727.

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand (Vol. 9). MIT press.

Donnell, C. J. O., & Connor, D. H. (1996). Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice, Accident Analysis and Prevention 28(6), 739–753.

Eluru, N., Bhat, C. R., & Hensher, D. a. (2008). A Mixed Generalized Ordered Response Model for Examining Pedestrian and

Bicyclist Injury Severity Level in Traffic Crashes. Accident; Analysis and Prevention, 40(3), 1033–54. doi:10.1016/j.aap.2007.11.010

Garrido, R., Bastos, A., de Almeida, A., & Elvas, J. P. (2014). Prediction of Road Accident Severity Using the Ordered Probit Model. Transportation Research Procedia, 3(July), 214–223. doi:10.1016/j.trpro.2014.10.107

Greene, W. H., & Hensher, D. A. (2009). Greene, W. H., & Hensher, D. A. (2010). Modeling ordered choices: a primer. Cambridge University Press.

Huelke, D. (1995). The Effects of Seat Belts on Injury Severity of Front and Rear Seat Occupants in the Same Frontal Crash. Analysis and Prevention, 27(6), 835–838.

Kockelman, K. M., & Kweon, Y. (2002). Driver Injury Severity : An Application of Ordered Probit Models. Analysis and Prevention, 34, 313–321.

Persaud, B., & Lyon, C. (2007). Empirical Bayes Before-After Safety Studies: Lessons Learned from two Decades of Experience and Future Directions. Accident Analysis and Prevention, 39(3), 546–55. doi:10.1016/j.aap.2006.09.009

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rifaat, S. M., & Chin, H. C. (2005). Accident Severity Analysis Using Ordered Probit Model. UAP Journal of Civil and Environmental Engineering, *1*(1), 45–54.

Sayed, T., & De Leur, P. (2008). Collision prediction models for British Columbia. Ministry of Transportation and Infrastructure.

Train, K. E. (2009). Discrete Choice Methods with Simulation. Cambridge University Press.

Transport Canada. (2012). Canadian Motor Vehicle Traffic Collision Statistics

Wang, X., & Kockelman, K. M. (2005). Use of Heteroscedastic Ordered Logit Model to Study Severity of Occupant Injury Distinguishing Effects of Vehicle Weight and Type, Transportation Research Record (1908), 195–204.