Collision Prediction Models for Calgary Arterial Roads and Intersections

Ahmed Abdelnaby, Bashar Dhahir, and Alf Guebert

Abstract

Roadway safety is a global goal most cities aim to reach. There are different significant factors that control safety; however, most cities and municipalities around the globe use empirical or experimental approaches to identify significant factors that affect roadway safety. Most of these factors depend on subjective observation, experience, and direct comparison between collisions before and after implementing a specific safety treatment. Therefore, roadway safety after changing any controllable factor, as a safety treatment, is difficult to measure. To overcome this issue, modelling techniques such as Negative Binomial Collision Prediction Models (CPMs) have been used by researchers and few cities and jurisdictions to describe the distributional qualities of rare events such as collisions. This approach was proven to identify significant factors that affect collisions and also model collisions more accurately than many other regression approaches. This study aims at using statistical means to determine significant factors affecting roadway safety, represented as vehicular collisions, and develop Collision Prediction Models for Calgary arterial roads segments and intersections. Collected data included collision records, Average Annual Weekday Traffic (AWDT), heavy vehicular traffic, operational speeds and speed limits, number travel lanes, intersections basic geometry and controls, road lengths, and intersection enforcement camera locations. Significant covariates will be used in the development of CPMs for Property Damage Only (PDO), Injury, and Fatal collisions. Using the developed CPMs will help the city to predict collisions on Calgary arterial roadways and intersections, identify hot spot locations, evaluate roadway safety along arterials, and assist on the quantification of safety at a high-level planning stage.
INTRODUCTION

Collisions are usually associated with different costs to the society such as: property damage, emergency response, legal costs, loss of productivity, environmental costs, etc.... Therefore, collisions are considered in the literature as a significant loss to the economy. 2008 collisions’ cost research published by the International Road Assessment Programme in Delhi, estimated that collision losses reach up to 3% of the global GDP [1]. The National Highway Traffic Safety Administration estimated that the cost associated with motor vehicle crashes in the United States during 2010 totaled approximately $242 billion [2]. Another collision costing study based on collision data from the province of Alberta for the capital region (Devon, Edmonton, Ft. Saskatchewan, Leduc, Sherwood Park, Spurce Grove, St. Albert, and Stony Plain) showed that the total direct cost in 2007 of 43 fatal collisions, 8,517 injury collisions, and 51,822 Property Damage Only (PDO) collisions were approximately $7.8, $336.6, and $565 million dollars, respectively [3]. In Calgary alone, the 2006-2010 average collisions cost was estimated to be between $871 million and $1.68 billion [4].

Globally, the World Health Organization (WHO) estimated that road traffic accident deaths are an average of 1.24 million per year [5]. The same report shows that about 20 to 50 million road users sustain nonfatal injuries due to collisions. Traffic collisions are the ninth leading cause of deaths around the world and are expected to be the fifth leading cause of deaths by 2030 [5] [6]. WHO data bank supports evidence that highest death rates per 100,000 population are in developing countries [7].

In Canada, the number of reported fatal and injury collisions in 2013 were 1,741 and 120,660, respectively [8]. Based on the statistics published by Transport Canada, the province of Alberta is in the highest 5 provinces and territories in fatality and injury rates (8.9 fatalities and 465.4 injuries per 100,000 population) [8]. In urbanized areas such as the City of Edmonton, the number of reported fatal, injury, and PDO collisions is 22, 2,912, and 21,693, respectively [9]. This translates into 2.5 fatal, 331.7 injury, and 2,470 PDO collisions per 100,000 population. The City identified in its annual report that following traffic too closely is the leading cause of intersection and midblock collisions. Furthermore, opening door into traffic followed by failure to observe traffic signals are the leading causes of fatal collisions. The City of Calgary’s Safety Mobility Plan indicates that the City’s annual average total collisions between 2009 and 2011 was 34,244 of which 2,279 collisions were injury and 22 were fatal [4].

Although the collision frequency has a decreasing trend over the past years [4] [8] [9] [10], it should not be concluded that safety improves over time. One of the early studies of crash data showed that collision data gathered from the United States, Canada, Sweden, Israel, and the United Kingdom support evidence of regression to the mean [11]. This means that the number of collisions at a specific location or studied intersections, for example, is expected to have a downward (decreasing) trend over time. Therefore, a direct comparison of the safety over time could be a biased measure of safety improvement or deterioration.

Three main practices could lead to biased outcomes when it comes to safety evaluation and studies. First, collisions are usually reported as rates per population, licensed drivers, registered vehicles, or vehicle kilometers travelled, which does not reflect the randomness nature of collisions. According to Sayed and De Leur, the reasons behind this could be that rates reflect an exposure variable that safety could be referenced to. Furthermore, collision frequencies do not reflect safety, when presented as a stand-alone measure [12]. Second, modelling collisions using
linear regression does not accurately estimate collisions. Linear regression does not support the properties of collision data, which are random, discrete, and nonnegative count data. In other words, the relationship between the mean and variance, non-negativity of the dependent variable (collision frequency), and the non-normal error structure cannot be modelled using multiple linear regression [13]. Third, comparison of collision frequencies between different years or before and after a safety treatment without a meaningful statistical correction may lead to biased results [11].

To study roadway safety without falling into bias of the aforementioned three practices, different statistical regression modelling approaches could be used such as Poisson, Negative Binomial, and their respective Zero Inflated model structures [13], [14], [15]. This study focuses on identifying significant covariates that affect collisions along arterial roads in Calgary. Since, most of the reported collisions in the collected database and in the City of Calgary are generally intersection related [4], separate models for intersections and segments were developed to identify significant covariates affecting collisions. A 5 year collisions history of 15 roads, broken down into 63 segments, and 78 intersections were used in this study in addition to other variables believed to have association with collisions.

**SCOPE AND STUDY OBJECTIVE**

The scope of this study is limited to the City of Calgary’s arterial road network within the ring road limit and excluded the downtown area. This research is limited to available data limitations and is intended as an attempt in establishing a relationship between safety and covariates believed to have a relationship with safety. It should be noted that this paper and developed models do not focus and may not be applicable for other road classes such as local or skeletal roadways, which serve different flow volumes and have different design and operational characteristics.

**DATA COLLECTION & PROCESSING**

*Data Collection*

Collision data request was tailored to gather representative collision records that would help in assessing and studying variations between arterial road segments with varying characteristics. Out of 108 initially candidate segments, 63 segments were selected based on having the following mixture; studied segments can be found in Appendix I:

1. **Average Annual Weekday Traffic (AWDT):** Studied segments had to ensure sufficient Average Annual Weekday Traffic heterogeneity. AWDTs were classified into seven (7) groups as shown in Table 1. Initially, a minimum of 6 segments for each AWDT group exceeding 40,000 veh/day was targeted (due to their infrequency compared to arterial segments with lower volumes). For the rest of the study segments a minimum of 10 segments was set a target. The set targets could not be achieved for segments containing an AWDT of 40,000 – 47,000 veh/day (4 selected segments) and segments exceeding 55,000 veh/day (3 selected segments).

2. **Length:** The studied segment were grouped at into ≤ 1 km segments and > 1 km segments. A split of 50% was targeted in the collected segment data.
3. Number of lanes: All studied segments were grouped into two (2) groups, \( \leq 4 \) lanes for both directions and > 4 lanes. The research team aimed at a 50-50 split between the two groups.

4. Speed limit: At the outset of this study, operational speed data was not available. Therefore, speed limits were used to subgroup the selected arterial road segments into \( \leq 60 \) km/h groups and > 60 km/h. Similarly, a 50% split was targeted in the candidate segments selection.

It should be noted that none of the procedures 1-4 were interdependent; data was entered into Excel spreadsheets, then a four-layer filtering process was applied. For example, the first filter was an AWDT \( \leq 17,500 \) combined with a segment length \( \leq 1 \) km, \( \leq 4 \) lanes filter and finally a speed limit \( \leq 60 \) km/h. This process was repeated over 56 times to get to the best level of control possible over the requested data. For all selected segments, 57% roadway segments were less than 1 km long, 68% were 4 lanes or less, 52% had a speed limit of 60 km/hr or less. The studied AWDT groups are summarized in Table 1.

### Table 1: Studied AWDT Segment Groups

<table>
<thead>
<tr>
<th>Group No.</th>
<th>AWDT Group (veh/day)</th>
<th>Number of Segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \leq 17,500 )</td>
<td>10 (15.9%)</td>
</tr>
<tr>
<td>2</td>
<td>17,500 – 25,500</td>
<td>14 (22.2%)</td>
</tr>
<tr>
<td>3</td>
<td>25,000 – 32,500</td>
<td>10 (15.9%)</td>
</tr>
<tr>
<td>4</td>
<td>32,500 – 40,000</td>
<td>16 (25.4%)</td>
</tr>
<tr>
<td>5</td>
<td>40,000 – 47,500</td>
<td>4 (6.3%)</td>
</tr>
<tr>
<td>6</td>
<td>47,500 – 55,000</td>
<td>6 (9.5%)</td>
</tr>
<tr>
<td>7</td>
<td>&gt; 55,000</td>
<td>3 (4.7%)</td>
</tr>
</tbody>
</table>

All selected 63 segments were merged into 15 longer roads to make the collision data extraction practical to the City staff. The 15 studied road sections are shown in Figure 1.
Collision records were obtained for the above selected roads from the City of Calgary’s Transportation Planning Division staff who provided the research team with continuous and prompt support during the course of this study. Received data included 5 years worth of collision records, spot speed and vehicle classification studies, and truck flow maps for 2011 and 2013. ADWT, speed limits, and Intersection Safety Cameras locations were collected from the City’s website. Other data were obtained through open data sources such as number of lanes at roads and length of roadway segments, which were measured and observed using Google Earth software package.

**Data Processing**

The City of Calgary’s police collision records are accurate in identifying intersection related versus non-related intersection collisions. However, there is no information regarding segment collisions accurate location details along a segment. Therefore, data was processed to identify and model collisions related to:

FIGURE 1: Studied Arterial Roads.
1. **Main/Major Intersections**: These are intersections where the AWDT changes between the road segment before and after (based on the City of Calgary flow maps). In other words, every studied roadway segments is started at a Main Intersection and ended at another one.

2. **Non-Major Intersections**: These are intersections where the AWDT does not change between the road segment before and after (based on the City of Calgary flow maps). In other words, every studied roadway segments could have one or more Non-Major Intersection. These intersections can be Right-In-Right-Outs (RIRO), un-signalized or signalized intersections.

3. **Segments**: Segments are roadway sections, which start and end at Major Intersections and have a measured length between the start and end points. Segments can have a number of Non-Major Intersections or may have none.

Collisions were extracted using Excel Pivot Tables by filtering each studied intersection separately. Then, V-look-up functions were used to read the filtered collision details such as fatality, notes, and location using the collision ID from the original dataset. Intersection collisions were easy to identify; however, segment collisions required going through all collision notes to precisely identify the segments of collision occurrence. A statistical summary of the study segments and intersection is shown in **Table 2**.

**TABLE 2: Collision Data Summary**

<table>
<thead>
<tr>
<th>Description</th>
<th>Selected Arterial Roadways</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. Segment’s length (km)</td>
<td>0.36</td>
</tr>
<tr>
<td>Max. Segment’s Length (km)</td>
<td>2.45</td>
</tr>
<tr>
<td>Average Segments Length (km)</td>
<td>1.01</td>
</tr>
<tr>
<td>Min. Number of Lanes</td>
<td>2</td>
</tr>
<tr>
<td>Max. Number of Lanes</td>
<td>6</td>
</tr>
<tr>
<td>Average Number of Lanes</td>
<td>4</td>
</tr>
<tr>
<td>Min. AWDT veh/day</td>
<td>8,000</td>
</tr>
<tr>
<td>Max. AWDT veh/day</td>
<td>66,000</td>
</tr>
<tr>
<td>Average AWDT veh/day</td>
<td>30,214</td>
</tr>
<tr>
<td>Max. Yearly Major Intersections collisions (PDO [yr], Injury [yr], Fatal [yr]) col/yr</td>
<td>15.49 [2010], 1.69 [2014], 0.10 [2012]</td>
</tr>
<tr>
<td>Average Major Intersections collisions (PDO, Injury, Fatal) col/yr</td>
<td>12.6, 1.39, 0.03</td>
</tr>
<tr>
<td>Max. Yearly Non-Major Intersections collisions (PDO [yr], Injury [yr], Fatal [yr]) col/yr</td>
<td>2.57 [2010], 0.4 [2013], 0.016 [2013, 2014]</td>
</tr>
<tr>
<td>Average Non-Major Intersections collisions (PDO, Injury, Fatal) col/yr</td>
<td>2.32, 0.33, 0.006</td>
</tr>
<tr>
<td>Max. Yearly Segment collisions (PDO [yr], Injury [yr], Fatal [yr]) col/yr</td>
<td>6.43 [2014], 0.63 [2011 &amp; 2013], 0.016 [2012, 2014]</td>
</tr>
<tr>
<td>Average Segment Intersections collisions (PDO, Injury, Fatal) col/yr</td>
<td>6.00, 0.56, 0.006</td>
</tr>
</tbody>
</table>
Collision frequency over the studied years for Major Intersections, Non-Major Intersections, and Segment Collisions are shown in Figure 2. A decreasing trend was observed for PDO collisions at Major Intersections. However, this plot is not inclusive evidence that there is a safety improvement or that this decreasing collision trend was not influenced or biased by other unknown factors. No other trends in Figure 2 were observed. Furthermore, Collision frequency versus total AWDT passing through major Intersections, Non-Major Intersections, and arterial road segments are shown in Figure 3. Length of study segments versus reported collision frequencies are plotted in Figure 4.

FIGURE 2: Collision Frequencies Versus Time.
Figure 3 indicates the possibility of a positive association between Major Intersections PDO collision frequency and AWDT. A similar association could be indicated between segments PDO collision frequency and AWDT. It should be noted that the coefficient of determination ($R^2$)

**FIGURE 3: Collision Frequency versus AWDT.**
values displayed are not a linear modelling attempt. Instead, they are only indicators of a possible association.

![Segment Length versus Collision Frequency](image)

**FIGURE 4: Collision Frequency versus Segment Length.**

Similarly, no relationship was indicated between length and collision frequencies in the studied segments.

**METHODOLOGY**

Collisions were analyzed using two modelling approaches: Negative Binomial and Zero-Inflated Negative Binomial. As mentioned earlier, Negative Binomial is used to overcome the distributional qualities of collision data. Zero-Inflated Negative Binomial models are used to cover for the excess zeros or Bernoulli trials with unequal chances of success [16], which highly describe fatal collision events.

**Negative Binomial Models**

The developed Collision Prediction Models (CPMs) for PDO and Injury collisions were based on the Negative-Binomial model structure [17] [18]. The expected number of collisions at a given intersection or segment is given by the following equation:

\[ E[y] = \exp(B^T X) \]  

Where,

- \( B \) = the vector of estimated model parameters, and
- \( X \) = the vector of model covariates.

The variance of the expected number of collision is given by:

\[ \text{Var}(\mu_i) = \mu_i + \frac{\mu_i^2}{\kappa} \]  

Where,

- \( \kappa \) = the inverse of the over-dispersion parameter which controls the extra-Poisson variance term.
The Generalized Linear Model package of the R statistical language was used to independently develop CPMs, for PDO and Injury Collisions.

**Zero-Inflated Negative Binomial Models**

Due rareness of fatal collisions at all segments and intersections and injury collisions at the segments and none-major intersections, the standard Negative Binomial regression could be insufficient. Since this collision data would include excess zeros, Zero-Inflated models were used in the modelling attempts. The model can be described by [14] [19] [16]:

\[
P(y_i) = \begin{cases} 
  p + (1 - p) g(0 | X_i) & y = 0, \\
  (1 - p) g(y_i | X_i) & y > 0 
\end{cases}
\]

(3)

Where,

\( g(y_i | X_i) \) = the probability mass function of a NB distribution, and

\( p = \) logistic function describing the response becoming a zero count.

In other words, the model is characterized by excess zeros, excess large response, or both. The final model form can be written as [16] [20]:

\[
\log(\lambda_i) = x_i \beta \text{ and } \text{logit}(p_i) = z_i \gamma, (i = 1, \cdots, n)
\]

(4)

Where \( x_i \) and \( z_i \) are the covariate matrix of the log and logit functions.

The expected number of collisions could be described by

\[
E(y_i|x, z) = \frac{\exp(x_i \beta)}{1 + \exp(x_i \gamma)}
\]

(5)

Akaike’s Information Criterion (AIC) [21] was used in model selection. All parameters were estimated using the Maximum Likelihood estimates method; the coefficients are estimated by equating the first order conditions to zero [22].

**Studied Covariates**

The main challenge in modelling collision data in this study is setting up the study covariates. Some of the data challenges faced are:

- Limited operational speeds data: The data point studies provided were limited to few locations and could not be generalized to the studied road segments and intersections. Therefore, posted speed limit was used as one of the roads’ operational indicators.

- Trucks flow data: There are no trucks flow maps for the years 2010, 2012, and 2014. Truck data are not necessarily collected on yearly basis in Cities. Therefore, 2010, 2012, and 2014 trucks percentages were extrapolated for every segment or intersection based on weighted AWDT extrapolation. Segments or intersections that had unreasonable extrapolation (i.e. lower than 0 in 2010) were corrected by equating the percentage of 2010 to 2011’s and the percentage of 2014 to 2013’s.

For every model type, the following covariates were selected:
• **Major Intersection Collisions Models Covariates:**

1. **AWDT\textsubscript{NS}:** AWDT was extracted for the segments just north and south of each of the studied intersections. The average was used to present north-south traffic using the intersection. If the intersection has only one leg, north or south, the average was not calculated.

2. **AWDT\textsubscript{EW}:** AWDT was extracted for the segments just east and west of each of the studied intersections. The average was used to present east-west traffic using the intersection. Similarly, if the intersection has only one leg, the average was not calculated.

3. **Trucks\textsubscript{NS}:** Percentage of trucks was extracted for the segments just north and south of each of the studied intersections. The average weight based on AWDT was used to present north-south truck traffic using the intersection. Similarly to the methodology followed in the average AWDT calculations, if the intersection has only one leg, the average was not calculated.

4. **Trucks\textsubscript{EW}:** Percentage of trucks was extracted for the segments just east and west of each of the studied intersections. The average weight based on AWDT was used to present east-west truck traffic using the intersection. Similarly, the average was not calculated, if the intersection has only one leg.

5. **Enforcement:** Intersection Safety Cameras were set as a dummy 0 or 1 variable. If the studied intersection had an Intersection Safety Camera this variable would be 1, otherwise 0.

6. **Intersection Weight:** this covariate was developed to calculate the change that drivers would experience once they approach the intersection. Calculation of this variable depends mainly on two parts: intersection control (un-signalized, signalized, Right-In-Right-Out, grade separated intersections) and the rate of change drivers experience before and after getting into the intersection. The following formula was used to determine this covariate:

\[
Int.Wt = C + \frac{A}{B}
\]  

(6)

Where,

- **C** = control type (0 for un-signalized, 1 for signalized, 2 for free through traffic).
- **A** = Total number of through lanes at all legs of the intersecting road before and after the intersection.
- **B** = Total number of approach and departure lanes at the intersection.

An example of a 4-legged signalized intersection weight calculation is shown in **Figure 5**.

\[
Int.Wt = 1 + \frac{16}{20} = 1.8
\]
FIGURE 5: Intersection Weight Calculation Example

In the above example, the Intersection weight is a measure of how far is the intersection’s configuration from a conventional 4-legged signalized intersection with no auxiliary lanes.

7. Year Variable: A dummy variable was assigned to study the relationship between collision frequency and time. This variable would present possible time series correlation.

Correlation between the studied covariates is shown in Table 3:

TABLE 3: Correlation between Major Intersections Studied Covariates.

<table>
<thead>
<tr>
<th></th>
<th>AWDT&lt;sub&gt;NS&lt;/sub&gt;</th>
<th>AWDT&lt;sub&gt;EW&lt;/sub&gt;</th>
<th>Trucks&lt;sub&gt;NS&lt;/sub&gt;</th>
<th>Trucks&lt;sub&gt;EW&lt;/sub&gt;</th>
<th>Enforcement</th>
<th>Intersection Weight</th>
<th>Year Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWDT&lt;sub&gt;NS&lt;/sub&gt;</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWDT&lt;sub&gt;EW&lt;/sub&gt;</td>
<td>0.249</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trucks&lt;sub&gt;NS&lt;/sub&gt;</td>
<td>0.127</td>
<td>-0.096</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trucks&lt;sub&gt;EW&lt;/sub&gt;</td>
<td>0.042</td>
<td>-0.069</td>
<td>0.614</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enforcement</td>
<td>-0.066</td>
<td>0.015</td>
<td>0.018</td>
<td>0.035</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intersection Weight</td>
<td><strong>0.553</strong></td>
<td><strong>0.527</strong></td>
<td>-0.014</td>
<td>-0.046</td>
<td>-0.112</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>0.016</td>
<td>0.018</td>
<td>-0.023</td>
<td>0.018</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

As shown in the correlation matrix, there is a positive correlation between AWDT and Intersection Weight which is expected; one of the intersection upgrades warrants is traffic
volumes. Furthermore, trucks percentage in an approach is positively correlated, as shown in the bold font format, with other approaches, which could be expected from an operational perspective.

- **Non-Major Intersection Collisions Models Covariates:**
  1. AWDT: AWDT numbers passing through the Non-Major intersections and travelling on arterial roads were extracted from Calgary’s traffic volumes flow maps. Taking into account volumes traveling on the intersecting roads was not possible since they were not reported on the flow maps.
  2. Trucks: Percentages of trucks passing through the Non-Major intersections and travelling along arterial roads were extracted from Truck flow maps.
  3. Non-Major Signalized Intersections Weight: this weight was calculated using Equation 5. The purpose of this weight is to describe the effect of signalized intersections geometry along a segment on collisions. If two or more Non-Major Signalized intersections were at the same segment, the individual weight for each was calculated first. Then, the weighted average based on distance to the closest intersection was used to represent this covariate at each segment.
  4. Non-Major Un-Signalized Intersections Weight: this weight was calculated using Equation 5. The purpose of this weight is to describe the effect of un-signalized intersections geometry along a segment on collisions. A similar approach, weighted average based on distance from the nearest intersection, was used when more than one Non-Major Un-Signalized intersection were observed at a segment.
  5. RIRO Weight: Equation 5 was used to calculate this covariate. The purpose of this weight is to describe the effect of RIROs’ geometry along a segment on collisions, Figure 6.

\[
Int.Wt = 2 + \frac{6}{8} = 2.75
\]

![FIGURE 6: Intersection Weight Calculation Example](image-url)

Likewise, average weight based on distance from the nearest intersection, was used when more than one RI, RO, or RIRO were observed in a segment.
6. Year Variable: A dummy variable was assigned to study the relationship between collision frequency and time. This variable would present time series possible correlation.

- **Segment Collisions Models Covariates:**
  1. AWDT: AWDT travelling on arterial roads were extracted from Calgary’s traffic volumes flow maps.
  2. Length: Length of each segment (between every two Major Intersections) was measured using Google Earth and was rounded to the nearest 5 meters.
  3. Trucks: Percentages of trucks travelling on arterial roads were extracted from Truck flow maps.
  4. Number of Lanes: the number of lanes of each segments were obtained using Google Earth.
  5. Speed Limit: Speed limit on each arterial segment was obtained from Posted Speed Limit maps available at the City’s website.
  6. Year Variable: A dummy variable was assigned to study the relationship between collision frequency and time. This variable would present possible time series correlation.

| TABLE 4: Correlation between Non-Major Intersections and Segments Studied Covariates |
|----------------------------------------|--------|--------|--------|--------|-------|-----|-----|-------------------|
|                                       | AWDT   | Length | Sig_Wt | Unsig_Wt | RIRO_Wt | Speed | Lanes | Trucks | Year Variable |
| AWDT                                   | 1.000  |        |        |          |          |       |       |        |                |
| Length                                 | -0.140 | 1.000  |        |          |          |       |       |        |                |
| Sig_Wt                                 | 0.042  | 0.336  | 1.000  |          |          |       |       |        |                |
| Unsig_Wt                               | -0.240 | 0.231  | 0.087  | 1.000    |          |       |       |        |                |
| RIRO_Wt                                | -0.082 | 0.254  | -0.082 | 0.230    | 1.000    |       |       |        |                |
| Speed                                  | 0.369  | 0.369  | 0.159  | -0.147   | -0.163   | 1.000 |       |        |                |
| Lanes                                  | 0.614  | -0.276 | 0.103  | -0.176   | -0.137   | 0.392 | 1.000 |        |                |
| Trucks                                 | 0.033  | 0.087  | 0.255  | 0.319    | -0.063   | 0.121 | 0.110 | 1.000  |                |
| Year_V                                 | 0.026  | 0.000  | 0.000  | 0.000    | 0.000    | 0.000 | 0.000 | 0.035  | 1.000          |

As shown in the correlation matrix, there is a positive correlation between **AWDT** and the **number of lanes** (bold font style), which is expected because the number of lanes is determined by the traffic volumes at the design stage of any road. Furthermore, it is highly likely that higher number of lanes (higher capacity) attracts more traffic (demand).
RESULTS

As stated earlier in the Methodology, the negative binomial regression approach was used to model the distributional qualities of collision data. Extremely rare collision events, such as fatal collisions, were modelled using both standard and zero inflated-negative binomial regression. CPMs were modelled using step wise regression (both directions, forward and backward) in order to find the best fitted model with all significant variables. Identifying significant covariates was done at a level of significance of 5%, which is frequently used in research (SE <0.05). All modelling output can be found in Appendix B.

**Major Intersections CPMs**

<table>
<thead>
<tr>
<th>Collision Type</th>
<th>Intercept (SE)</th>
<th>AWDT_{NS} (SE)</th>
<th>AWDT_{EW} (SE)</th>
<th>Trucks_{NS} (SE)</th>
<th>Trucks_{E} (SE)</th>
<th>Enforcement (SE)</th>
<th>Intersection Wt (SE)</th>
<th>Year Variable (SE)</th>
<th>Dispersion</th>
<th>2xlog-likelihood</th>
<th>Residual Deviance</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDO 1 (NB)</td>
<td>2.436 (&lt;0.001)</td>
<td>9.172x10^{-6}</td>
<td>2.158x10^{-5}</td>
<td>--</td>
<td>--</td>
<td>0.268</td>
<td>0.312</td>
<td>0.085</td>
<td>1.686</td>
<td>-2691.898</td>
<td>467.66</td>
<td>2705.9</td>
</tr>
<tr>
<td>PDO 2 (NB)</td>
<td>2.066 (&lt;0.001)</td>
<td>9.259x10^{-6}</td>
<td>1.947x10^{-5}</td>
<td>--</td>
<td>--</td>
<td>0.271</td>
<td>0.252</td>
<td>0.081</td>
<td>1.636</td>
<td>-2127.34</td>
<td>373.98</td>
<td>2139.3</td>
</tr>
<tr>
<td>Injury (NB)</td>
<td>0.597 (&lt;0.022)</td>
<td>9.659x10^{-6}</td>
<td>1.366x10^{-5}</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>-0.496</td>
<td>0.004</td>
<td>2.159</td>
<td>-1227.95</td>
<td>424.44</td>
<td>1237.9</td>
</tr>
<tr>
<td>Fatal (NB, Log ( \theta = 0.917 ))</td>
<td>0.759 (0.511)</td>
<td>--</td>
<td>--</td>
<td>88.321</td>
<td>60.785</td>
<td>--</td>
<td>0.076, 0.0172</td>
<td>(0.043)</td>
<td>(0.115)</td>
<td>2.501</td>
<td>--</td>
<td>98.97</td>
</tr>
<tr>
<td>Fatal (ZINB)</td>
<td>4.308 (&lt;0.001)</td>
<td>--</td>
<td>--</td>
<td>-99.756</td>
<td>62.802</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>2.501</td>
<td>--</td>
<td>--</td>
<td>98.97</td>
</tr>
</tbody>
</table>

SE = Standard Error of Estimate.

PDO collisions were modelled using negative binomial regression, which is justified due to the over dispersion of the data. It was found that the year variable is significant (Model: PDO 1), which indicates time series correlation. This could be related to the change of PDO collisions reporting threshold from $1,000 to $2,000 on January 1st, 2011 [10]. Therefore, 2010 PDO collisions were removed from the data and the CPMs were redeveloped (PDO 2). It was found that the year variable was insignificant and the model had a better goodness of fit compared to PDO 1.

15
Injury collisions were modelled using both negative binomial and zero-inflated negative binomial regression. It was concluded that the negative binomial model was more informative and better fitted the data. Similarly, fatal collisions were modelled using both negative binomial and zero-inflated negative binomial regression. It was found that the zero-inflated approach yielded better model fitting.

**Non-Major Intersections CPMs**

<table>
<thead>
<tr>
<th>Collision Type</th>
<th>Intercept (SE)</th>
<th>AWDT (SE)</th>
<th>Trucks (SE)</th>
<th>Sig_Wt (SE)</th>
<th>Unsig_Wt (SE)</th>
<th>RIRO_Wt (SE)</th>
<th>Year Variable (SE)</th>
<th>Dispersion</th>
<th>2xlog-likelihood</th>
<th>Residual Deviance</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDO (NB)</td>
<td>-2.322</td>
<td>2.468x10^{-7}</td>
<td>--</td>
<td>1.317</td>
<td>1.755</td>
<td>0.28</td>
<td>--</td>
<td>0.501</td>
<td>-1016.702</td>
<td>249.46</td>
<td>1030.7</td>
</tr>
<tr>
<td>Injury (NB)</td>
<td>-2.777</td>
<td>--</td>
<td>6.916</td>
<td>(0.007)</td>
<td>&lt;0.001</td>
<td>0.294</td>
<td>--</td>
<td>0.845</td>
<td>-401.88</td>
<td>181.73</td>
<td>411.88</td>
</tr>
<tr>
<td>Fatal (ZINB)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

*SE= Standard Error of Estimate.*

Using negative binomial regression, trucks and year variable were not found significant for PDO collisions at Non-Major Intersections. Injury collisions were modelled using both negative binomial and zero-inflated negative binomial regression. It was concluded that the negative binomial model was more informative and better fitted the data. Only percentage of trucks, signalized intersection and RIRO weights were found significant. Fatal collisions were modelled using both negative binomial and zero-inflated negative binomial regression. However, no statistical significant relationship was found using both modelling approaches, which could be highly related to the extreme rareness of fatal collisions on Non-Major fatal collisions (2 fatal collisions over 315 entities).

**Segment CPMs**

<table>
<thead>
<tr>
<th>Collision Type</th>
<th>Intercept (SE)</th>
<th>AWDT (SE)</th>
<th>Length (SE)</th>
<th>Trucks (SE)</th>
<th>Speed Limit (SE)</th>
<th>No. Lanes (SE)</th>
<th>Year Variable (SE)</th>
<th>Dispersion</th>
<th>2xlog-likelihood</th>
<th>Residual Deviance</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDO (NB)</td>
<td>1.275</td>
<td>4.602x10^{-5}</td>
<td>6.827x10^{-4}</td>
<td>--</td>
<td>-0.027</td>
<td>--</td>
<td>--</td>
<td>4.548</td>
<td>-1620.759</td>
<td>377.78</td>
<td>1630.8</td>
</tr>
<tr>
<td>Injury (NB)</td>
<td>-0.756</td>
<td>4.06x10^{-5}</td>
<td>--</td>
<td>--</td>
<td>-0.264</td>
<td>(0.003)</td>
<td>--</td>
<td>2.882</td>
<td>-620.337</td>
<td>290.81</td>
<td>618.34</td>
</tr>
<tr>
<td>Fatal (ZINB)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

*SE= Standard Error of Estimate.*
Trucks, number of lanes and year variable were not significantly affecting PDO collisions at the studied segments, using negative binomial regression. For injury collisions models, it was concluded that the negative binomial model was more informative and better fitted the data; only AWDT and number of lanes were significant covariates. Fatal collisions were modelled using both regression methods. However, no significant relationship was found using both modelling approaches, which could be highly related to the extreme rareness of fatal collisions on segment fatal collisions (2 fatal collisions over 315 entities).

**Total Segment & Non-Major Intersections CPMs**

<table>
<thead>
<tr>
<th>Collision Type</th>
<th>Intercept (SE)</th>
<th>AWDT (SE)</th>
<th>Length (SE)</th>
<th>Trucks (SE)</th>
<th>Speed Limit (SE)</th>
<th>No. Lanes (SE)</th>
<th>Sig_Wt (SE)</th>
<th>Unsig_Wt (SE)</th>
<th>RIRO_Wt (SE)</th>
<th>Year Variable</th>
<th>Log (theta)</th>
<th>Dispersion</th>
<th>2xlog-likelihood</th>
<th>Residual Deviance</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDO (NB)</td>
<td>0.899</td>
<td>4.381x10^-3</td>
<td>4.185x10^-4</td>
<td>--</td>
<td>-0.019</td>
<td>--</td>
<td>0.251</td>
<td>0.396</td>
<td>0.134</td>
<td>--</td>
<td>--</td>
<td>3.506</td>
<td>-1796.73</td>
<td>369.71</td>
<td>1812.7</td>
</tr>
<tr>
<td>Injury (NB)</td>
<td>-2.089</td>
<td>2.3x10^-3</td>
<td>4.738x10^-4</td>
<td>3.418</td>
<td>--</td>
<td>--</td>
<td>0.417</td>
<td>--</td>
<td>0.129</td>
<td>--</td>
<td>--</td>
<td>3.197</td>
<td>-745.777</td>
<td>314.77</td>
<td>759.78</td>
</tr>
<tr>
<td>Fatal (ZINB)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

SE= Standard Error of Estimate.

Combined collisions at segments and Non-Major Intersections were modelled using negative binomial regression, which is justified due to the over dispersion of the data. It was found that trucks, number of lanes and year variable were not significant.

Injury collisions were modelled using both negative binomial and zero-inflated negative binomial regression. It was concluded that the negative binomial model was more informative and better fitted the data. Similarly, fatal collisions were modelled using both regression methods but no significant relationship was found, which could be again highly related to the extreme rareness of fatal collisions on Non-Major fatal collisions (4 fatal collisions over 315 entities).
DISCUSSION AND CONCLUSION

Collision data collection took into account variation between different arterials across the city. Studied variations between different arterial segments were: AWDT, length, number of lanes, and, speed limit. Four models were developed to quantify collisions along arterials. Time series correlation was tested and it was found that Major Intersections collision data showed such correlation. 2010 PDO collisions at Major intersections were removed from the modelling dataset as reporting limit was increased in the province of Alberta at the beginning of 2011. Collision data was remodeled and showed no time dependence or correlation. Since no time correlation was observed on segment collisions or Non-Major intersections collisions, no collisions from 2010 studied year were removed during the modelling process.

Major Intersections CPMs

As per the City’s statistics and the studied collisions database, most collisions occur at intersections. The developed PDO CPMs models show that AWDTs crossing intersections are positively correlated with PDO collisions frequency. Furthermore, installed Safety Cameras were positively associated with PDO collision frequency. It is unclear if Intersection Safety Cameras are a cause of PDO collisions or are just associated with being installed at hot spot locations. More research studying this aspect is suggested.

Developed Intersection Weights were negatively correlated with PDO collisions frequency, which has two interpretations. First, un-signalized intersections are expected to have more PDO collisions compared to signalized and free flow controlled intersections. Similarly, signalized intersections are expected to have more PDO collisions as compared to free flow controlled intersections. Second, conventional intersections with no auxiliary lanes are expected to have less collision frequency as compared to intersections with right and left turn lanes. In other words, as the driver experiences less changes in the number of approach and departure lanes, less PDO collisions are expected.

Modelled Injury collisions showed a positive relationship with AWDT and a negative relationship with Intersection weight. Other studied covariates showed no significant relationship with injury collisions frequency. Truck traffic was positively associated with fatal collisions. A simulation run was performed assuming that north-south truck percentage is equal to east-west truck traffic. It was found that fatal collisions have a positive relationship with truck traffic up to 9.9%. The relationship turns negative at 10% truck traffic as fatal collisions approach zero, Appendix II. This relationship made sense to as the speed differentials in traffic could be high up to a point where more heavy traffic would no longer affect speed differential within traffic, rather it would affect the whole speed of the platoon. Further, at increased speed differentials as the kinetic impact between a heavy vehicle and any other vehicle class would likely cause a more severe damage as compared to collisions between passenger cars or lighter vehicle classes. However, upon careful investigation of the traffic police notes, this association could not be confirmed. Police fatal collision data notes indicated alcohol impairers, pedestrian involvement or no information.
**Non-Major Intersections CPMs**

PDO collisions were positively correlated with AWDT and un-signalized, signalized and RIRO intersection weights. The positive correlation with intersection weights could be related to:

1. The lower class intersecting roads’ AWDT, which was not included in the studied covariates as such information was not available.
2. The lower operational speeds at Non-Major Intersections due to possible succession of a number of Non-Major intersections over shorter lengths (more traffic interruption). Operational speeds were not covered in the studied covariates as data was not available for the majority of the studied intersections.

Injury collisions were positively correlated with truck traffic travelling on arterial roads and passing through Non-Major intersections. Similar to PDO collisions, positive relationship was found between injury collisions and signalized and RIRO intersection weights. Fatal collision data points were not enough to develop CPM models.

**Segment CPMs**

PDO collisions were positively correlated with AWDT and segment lengths, which is expected as they are main indicators of exposure. A negative relationship between PDO collisions and speed limits was found. Literature indicates no clear relationship between speed and safety. Some studies support a relationship between speed and collision frequency [23] [24] [25] [26]. On the other hand, different studies indicate that collision frequency is related to the speed differentials between traffic, while severity could be related to operational speeds [27] [28].

Injury collisions had a positive relationship with AWDT and a negative relationship with the number of lanes along the studied segments. Fatal collision data points were not enough to develop CPM models.

**Total Segment & Non-Major Intersections CPMs**

Collision and covariate data points for segments and Non-Major Intersections were combined and it was possible to establish CPMs for any segment including the Non-Major Intersections within that segment. It was found that PDO collisions had a positive relationship with AWDT, length, and un-signalized, signalized and RIRO intersection weights. The relationship between PDOs and speed was negative.

Injury collisions had a positive relationship with AWDT, Length, Trucks, and signalized and RIRO intersection weights. No fatal CPMs were developed due to the lack of sufficient fatal collisions data points.

This study provides statistical evidence of the relationship between collisions and different studied covariates. A summary of significant covariates, at 5 percent level of significance, affecting collisions is provided below:

- **AWDT:** Is the main exposure study covariate in the developed models. A positive relationship between AWDT and all collision models, except for fatal collisions, was found.
• **Length**: This is another exposure covariate in the segment CPMs. A positive relationship was found between segment length and PDO and injury collision frequency.

• **Trucks**: The only interpretable relationship between collisions and trucks percentage was in the developed injury CPMs at Non-Major Intersections and Total Segment & Non-Major Intersections CPMs. It was found that trucks travelling on arterials are positively correlated to collision frequency. Another significant relationship was found between trucks and fatal collisions at Major-Intersections. However, this relationship was not highly supported by the traffic police notes of the analyzed fatal collisions.

• **Enforcement**: Intersection Safety Cameras had a positive relationship with PDO collisions at Major Intersections. As the City spends efforts to install these at collision prone locations. It is not possible to judge if this relationship is due to drivers’ behavior at these locations (braking which could increase rear-ends) or due to just being installed in collision prone locations.

• **Intersection Weight**: Intersection control was found to affect safety. A positive relation was found at Non-Major Intersections while a negative relation was found at Major Intersections. Consistency of Intersection number of lanes as compared to the intersecting road number of lanes, before arriving to the intersection and after leaving the intersection, was found to impact safety.

According a synthesis of highway practice published by NCHRP, based on previous studies, adding turn lanes decreases both PDO and injury related collisions by 32% and 50%, respectively [29]. The developed CPMs show that adding turning lanes has different effect on collisions depending on the context of the intersection, Major or Non-Major. Driver expectancy to the upcoming number of lanes leaving or approaching an intersection could highly affect behavior, which may lead to effects on safety. For example, adding turning lanes was found to negatively impact PDO collision frequency at Major Intersections. On the other hand, the opposite relationship was observed at Non-Major Intersections. Many other variables could affect collisions when studying adding or removing turning lanes such as driver expectations, intersection and road consistency, operational speeds, intersection movements’ volumes, and other possible context parameters.

• **Speed**: A positive relationship was found between segment’s posted speed limit and PDO collision frequency.

• **Number of lanes**: Number of lanes had a negative relationship with injury collisions at the studied segments.

Negative Binomial Collision prediction models are considered as an unbiased explicit safety measure. If properly developed, safety and effectiveness of roadway treatments could be evaluated easily.

In general, developed models could be used to:

• Predict location and category specific accidents, identify and rank them for every arterial segment & intersection.

• Evaluate the safety of the arterial segments using Empirical Bayes as explained by Sayed and De Leur [12].
- Evaluate the effectiveness of before and after studies using Empirical Bayes.
- Assist in the design of new freeways with similar features at a high level.

There is a very good room for improvement in the collision prediction models, which can be done by:

- Adding more collision history could highly increase the goodness of fit of the developed model. Further, it may make it possible to model fatal collisions.
- Studying construction history and including it as a study covariate could provide new findings. Data was requested from the City. Once received, models could be refined.
- Operational speeds may reflect a better operational covariate in the developed CPMs as compared to speed limits; including such data will help in understanding the effect of operational speed and speed differential between drivers on collision frequencies on arterial roads. Such data would be a major refinement for the developed models.
- The developed intersection weights do not take into account the frequency of Non-Major intersections. Including the frequency of these intersections could be a significant refinement to the developed Non-Major intersection models.

ACKNOWLEDGEMENTS

The study team would like to express their appreciation to the City of Calgary Staff in the Transportation Data Division and the Traffic Safety Group for their continued help and prompt support. The City staff accommodated the study data request and answered all the research team questions within very short time frames. Their efforts facilitated the on-time delivery of this research.
REFERENCES


[27] Number of fatalities in accidents caused by cars, by speed: Fatal accidents are caused more frequently in the midspeed speed by drivers being inattentive or distracted, *ITARDA INFORMATION*, no. 90, pp. 1-12, September 2011.


Authors Information:

Ahmed Abdelnaby  
Transportation E.I.T,  
Watt Consulting Group  
3016 5 Ave NE #310, Calgary,  
Alberta, T2A 6K4  
Tel: 403-273-9001 Ext 244  
E-mail: AAbdelnaby@wattconsultinggroup.com

Bashar Dhahir  
PHD Candidate, Carleton University  
1125 Colonel By Drive, Ottawa  
Ontario, K1S 5B6  
E-mail: bashardhahir@cmail.carleton.ca

Alf Guebert  
Manager of Transportation Division,  
Watt Consulting Group  
3016 5 Ave NE #310, Calgary,  
Alberta, T2A 6K4  
Tel: 403-273-9001 Ext 722  
E-mail: AGuebert@wattconsultinggroup.com

Paper submitted for Publication in Canadian Institute of Transportation Engineers Conference  
June 2016