Macroscopic Evaluation of Active Transportation Safety; City of Vancouver as a Case Study

Ahmed Osama

Civil Engineering Department, University of British Columbia

Part of Ph.D. research under the supervision of Prof. Tarek Sayed

ahmed.osama@eng.asu.edu.eg

ITE District 7

Abstract

This study investigates the factors that impact active commuters' safety. A large GIS dataset was extracted for the City of Vancouver 134 traffic analysis zones (TAZs) including traffic exposure, socioeconomics, land use, built environment, and road facility. In addition, bike and sidewalk network indicators such as connectivity, continuity, slope, and length, were quantified using graph theory. Negative binomial generalized linear modeling is used to develop macro-level collision prediction models (CPMs) for pedestrian crashes as well as for cyclist crashes. The CPMs are used to capture the associations between the aforementioned TAZs' attributes and cyclists' and pedestrians' safety.

The results revealed the significant, and almost similar, impact of zone characteristics on cyclist crashes and pedestrian crashes. Recreational area density, residential area density, proportion of local roads, off-street bike paths, network continuity, network length, and network slope were all found negatively associated with active commuters' crashes; on the contrary of employment density, household density, signal density, bus stop density, arterial-collector roads proportion, and network's connectivity variables that were found positively associated with active commuters' crashes. This indicates that any suggested policies for increasing the number of active commuters should be accompanied by risk actions. It also implies the important role that the quality of active transportation network, the built environment, the street network, and the land use can play in improving active commuters' safety.

Keywords

Active Transportation Safety, Macro-Level CPMs, Network Indicators.

1. Introduction

Many cities worldwide are recognizing the important role that active transportation can play in creating safer, sustainable, and livable communities. Active transportation has been known for its various advantages such as reducing the congestion and the associated emissions as well as increasing the physical activity levels which would result in various heath benefits (De Hartog et al., 2010). This motivates the cities' authorities to apply various policies that would promote such trend of transportation. However, active commuters are vulnerable road users and their safety is always at risk. They are usually subjected to an elevated level of injury risk and discomfort, which may discourage them from using these active modes. Locally, the cyclists and pedestrians in the City of Vancouver accounted for approximately 3% of the reported crashes between years 2007 and 2012; nevertheless, they represented a large portion (approximately 50%) of the fatalities over the same period according to the insurance company of British Columbia (ICBC) statistics.

According to the aforementioned, there is a growing need for advocating proactive planning strategies and viable decision support tools that are capable of assessing active transportation safety planning policies. Macro-level CPMs are statistical models that can play such role suitably. In macro-level CPMs, crashes (in this study cyclist/pedestrian crashes) are modeled as a function of wide area (e.g. neighborhood, traffic analysis zone, etc.) characteristics.

This study investigates five years of pedestrian-motorist crashes and cyclist-motorist crashes at 134 traffic analysis zones (TAZs) in the City of Vancouver. Generalized linear modeling (GLM) is used to develop the macro-level CPMs that are capable of evaluating the pedestrian and cyclist safety. Traffic exposure data including vehicle kilometers travelled, bike kilometers travelled, and walking trips are collected for each TAZ. In addition, a large dataset of geographical information system (GIS) data regarding socioeconomics, land use, built environment, road facility is compiled. Also, graph theory is used to quantify several bike and sidewalk networks' indicators such as connectivity, continuity, slope, and length. The CPMs are developed using the aforementioned data in order to find associations between those various zone characteristics and active transportation safety.

2. Literature Review

2.1. Safety Models for Cyclists

In the last few years, several studies had discussed the relationship between cyclist crashes and various explanatory variables. In terms of socio-demographic variables, Chen (2015) showed that the increased employment density and household density were associated with a decline in collisions' frequency. Siddiqui et al. (2012) showed that the numbers of population and employment, along with the median household income, were positively related to the cyclist crash frequency.

Among traffic control variables, low-speed streets were found negatively associated with the number of cyclist crashes, while on contrary, high-speed streets had an increased number of cyclist crashes (Siddiqui et al., 2012). Additionally, high traffic signal density was found positively associated with the cyclist crashes in a TAZ (Wei and Lovegrove, 2013); (Chen, 2015).

As for travel demand variables, Jacobsen (2003) examined the relationship between the number of vulnerable road users (pedestrians and cyclists) and their collisions with motor vehicles based on five data sets from different locations worldwide. Results showed that, at the population level, the number of motorists colliding with vulnerable road users would increase at approximately 0.4 power of the number of people cycling or walking. Taking into account the amount of walking and cycling, the collision probability would decline with about -0.6 power of that number. Robinson (2005) studied three datasets from Australia, and concluded similar results to Jacobsen's study. Other studies found out that the increase in bike volumes (Miranda-Moreno et al., 2011); (Strauss et al., 2013) and vehicle volumes (Hamann and Peek-Asa, 2013) had positive associations with the cyclist crash frequency. Prato et al. (2015) showed that the association between the bike or vehicle volumes and the cyclist-motorist collisions was positive but non-linear.

Regarding land use, the increased commercial land use was positively associated with the cyclists' crash frequency and injury (Narayanamoorthy et al., 2013); (Vandenbulcke et al., 2014). Also, Amoh-Gyimah (2016) found out that the increase in the percentage of the residential area, the percentage of the industrial area, and the land use balance mix was positively associated with the cyclist-motorist collisions.

As for networks' features, increased intersection density was found to be positively associated with cyclist crashes (Siddiqui et al., 2012); (Strauss et al., 2013); (Wei and Lovegrove, 2013). Moreover, bike collisions frequency was found to increase as the intersection's complexity got higher (Vandenbulcke et al., 2014). For different types of bike facilities, the off-road bike lanes were found safer than the on-road ones (Hamann and Peek-Asa, 2013); (Reynolds et al., 2009); (Teschke et al., 2012). Chen et al. (2012) also showed that the installation of bike lanes did not lead to additional crashes, but a possible increase in the number of cyclists instead. On the contrary, more vehicles' lanes were found positively associated with the cyclist crashes. In terms of street elements, a higher bus stop density was associated with an increased cyclist-motorist crash frequency (Strauss et al., 2013); (Wei and Lovegrove, 2013).

2.2. Safety Models for Pedestrians

Few studies have been published on the relationship between pedestrian crashes and different correlates. Lee and Abdel-Aty (2015) studied the pedestrian crashes over a 4 year period at a set of intersections in Florida. Using log-linear models, they found out that the road geometry, traffic and environmental conditions, and demographic factors are associated with the frequency and severity of pedestrian crashes.

Wang and Kockelman (2013) used a modeled value for the walk miles travelled as an exposure variable to prove the positive association between the pedestrian crash risk and a higher mixing of commercial and residential land uses. They also proved a negative association between the sidewalk provision and lower crash risk.

Siddiqui et al. (2012) developed macro-level CPM for pedestrian crashes and found out significant association between pedestrian crashes and the number of intersections, length of roads with posted speed limit of 35 kph, population, employment, the number of dwelling units, the number of hotel units, and long-term parking cost.

Miranda-Moreno et al. (2011) studied the link between built environment, including land use types, road network connectivity, transit supply, and demographic characteristics, on pedestrian activity and pedestrian-motorist crashes. They found out that strategies that increase densification, mix of land use, and transit supply will increase pedestrian activity and may indirectly, increase the total number of injured pedestrians.

Lastly, Kim et al. (2010) investigated the association between different types of collisions (total injury/injury/fatal and pedestrian/cyclist) and demographic, land use, and roadway accessibility in Honolulu, USA. They found out that the number of bus stops, number of intersections, number of people living below poverty level, and number of jobs are all associated with pedestrian crashes.

2.3. Graph theory

Graph theory concepts originated from the solution of the "Seven Bridges of Konigsberg" problem, which was done by Euler in the 18^{th} century. Graph theory can provide techniques for evaluating network quality and measuring its impact on travel behavior. Garrison and Marble (1962) were the first to introduce graph theory principles to transportation geography. Kansky (1963) presented indices that characterized network connectivity and complexity.

More recently, Gattuso and Mirello (2005) were able to evaluate the topology and geography of metro networks in some European cities and New York City based on graph indicators. Quintero et al. (2014) introduced a novel approach to redraw transit networks as graphs, and hence they were able to include new connectivity indicators. Those indicators were used in developing macro-level CPMs to assess the safety of Metro Vancouver transit network (Quintero et al., 2013). In addition, graph theory measures had been applied to the field of transportation planning in several other studies (Xie and Levinson 2007; Derrible and Kennedy 2009; Rodrigue et al. 2009; Berrigan et al., 2010).

2.4. Contribution to the Literature

This paper presents several contributions to the literature as follow:

- The GIS data collected for this study and the suggested safety correlates are far more comprehensive than the previous studies that attempted to study cyclist or pedesterian safety.
- Real active commuting exposure measures, i.e. bike kilometers travelled and walk trips, are incorporated in the cyclist and pedesterian CPMs for the first time.
- This paper is the first to investigate the association between bike and sidewalk network indicators, quantified using graph theory, and active commuters' safety.

3. Data Collection

3.1 Data Sources

Zone-level CPMs are developed in this study based on 134 TAZs in the city of Vancouver. Explanatory variables that are related to network and zone characteristics are included in the CPMs. Walk trips, bike kilometer travelled, and vehicle kilometers travelled are incorporated in the models as exposure variables. The data needed for the explanatory variables is compiled using ArcGIS software, for processing and visual representation, after being extracted from five main sources:

1. Insurance Corporation of British Columbia, a public automobile insurance company, provided the crash data for a 5 years period (2009-2013). Only pedestrian-motorist and cyclist-motorist crashes are included in the analysis, as shown in Figure1. A 5 years period is selected to collect an adequate sample size. The sample included 3 severity levels, i.e. fatality, injury, and property damage only. However, the total number of crashes is included in the analysis in order not to disperse the sample size. The reported crash data may have some limitations such as unreported crashes due to low severity and absence of records for cyclist-pedestrian crashes.

2. Translink, the Metro Vancouver transportation authority, provided the geocoded files for the city of Vancouver road network, sidewalk network, bike network, land use, and TAZ boundaries. Moreover, Translink provided the output of an Emme2 transportation planning model for the travel demand in Metro Vancouver in year 2011. Translink used the 2011 household travel survey to calibrate the model, and the 2011 cordon counts to validate the model assignments.

3. Acuere Analytics provided the Vancouver Cycling Data Model (VCDM 2011). The VCDM used the bike count occurred between years 2005 and 2011 to estimate the annual average daily bike traffic (AADB) over the city of Vancouver bike network (El Esawey et al., 2015). The available data covered more than 810,000 hourly volumes over seven years. The model was efficient in estimating the AADB on most of the bike network links (3180 links, or more than 70% of the network).

4. The open data catalogue of the City of Vancouver (http://vancouver.ca/your-government/open-data-catalogue.aspx) provided the city built environment data (i.e. transit stops, traffic signals, and light poles).

5. Census Canada provided the socio-economic data (i.e. employment and household data) of the City of Vancouver according to 2011 census.

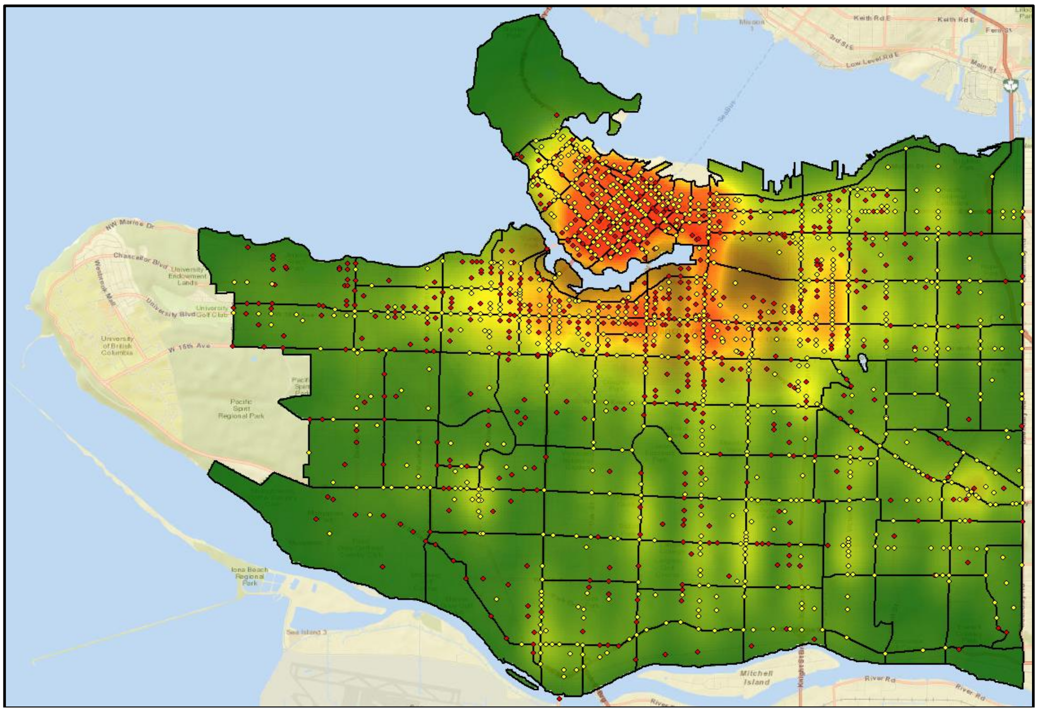


Figure 1 Heat Map of Cyclist Crashes (Red Points) and Pedestrian Crashes (Yellow Points) at City of Vancouver TAZs

3.2 Data Variables

The variables included in the analysis are divided into eight main categories; crashes, exposure, socio-economic, land use, built environment, road facility, cycling network, and pedestrian network. Table 1 provides the definitions and the descriptive statistics of the variables. The aggregation process of the different variables, at the zone level, is done using the ArcGIS software as discussed below.

Cyclist-motorist and pedestrian-motorist crashes are aggregated at the different TAZs according to their geospatial locations. Boundary crashes are distributed between the adjacent TAZs according to the relative proportion of BKT or W (depending on the type on the model, cyclist collisions model or pedestrian collisions model) at these zones. This way of distribution is selected due to the direct association between traffic exposure and both and crash risk, as revealed by the models developed in this study and several previous studies (Prato et al., 2016), (Strauss et al., 2013), (Miranda-Mreno et al., 2011).

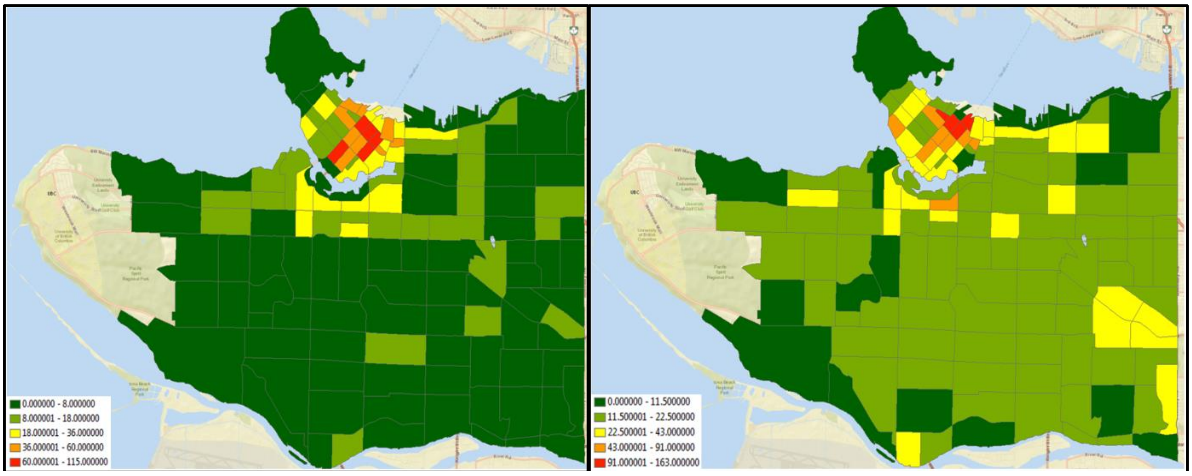
Three exposure measures are used in this study, i.e. BKT and VKT for cyclist CPMs, and W and VKT variables for pedestrian CPMs. Vancouver cycling data model provided the bike trip counts on the city of Vancouver road segments (El Esawey et al., 2015). A link based method is then used to calculate the zonal BKT. The road segments are represented by links, and the trip count at each link is multiplied by the link length to obtain the link BKT that is then aggregated to obtain the total BKT at each TAZ. As for the VKT and W variables, the data were readily provided by the Emme2 model on the TAZ level.

The socio-economic variables (i.e. employment, and household) were already provided by the Emme2 model in a TAZ aggregated form, and are then divided by the corresponding TAZ area to compute their densities.

The freeway, arterial, collector, and local roads are represented as links, and their link lengths are aggregated at each TAZ. The aggregation of each road class is then divided by the road network total length to determine the proportion of each class of the total road network. As well, the total length of the off-street bike links is aggregated at each TAZ, and then divided by the total road network length to determine its proportion.

For the land use variables, the areas of the commercial, residential, and recreational zonings are aggregated at the different TAZs, and then divided by the corresponding TAZ area to obtain the density of each zoning type.

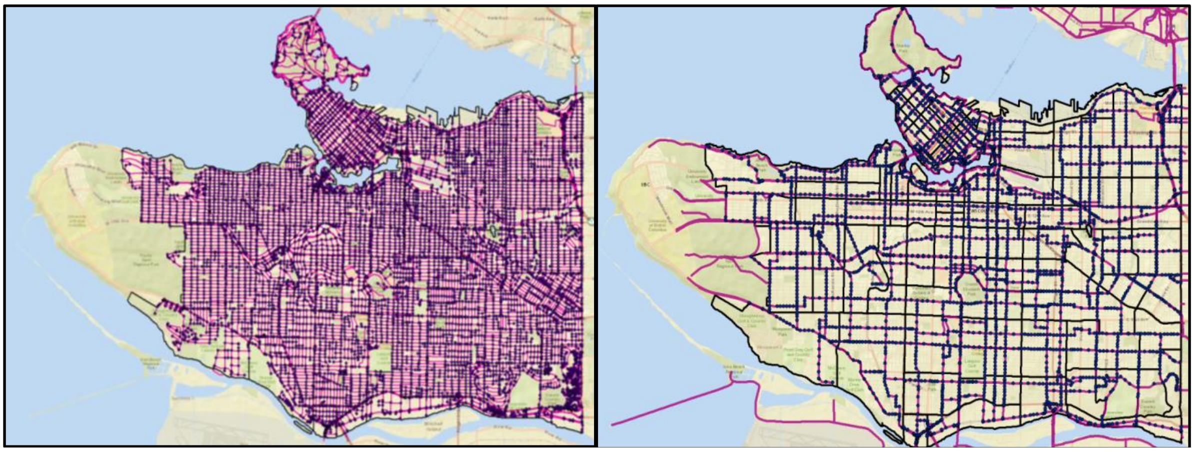
For built environment indicators, the number of traffic signals and bus stops are also aggregated at each TAZ, and then divided by the corresponding TAZ area to obtain their densities. Figure 2 shows the distribution of the traffic signal and bus stop densities among the city of Vancouver TAZs.



(a) **(b)**

Figure 2 (a) Traffic Signal Density and (b) Bus Stop Density at City of Vancouver TAZs

In addition to the aforementioned variables, more variables can be obtained using graph theory for both the bike and the sidewalk networks. The first step is to characterize the bike and sidewalk networks as graphs (i.e. sets of links and nodes) in order to collect the basic measurements of the graph characterization (e.g. number of nodes, number of links, length of links, etc.). The links represent the network infrastructure (i.e. bike lanes, sidewalks, etc.), and the nodes represent the intersections between the links. Since a zonal level of aggregation is selected, a technique is developed for splitting the entire network among the smaller zones. The links and nodes are distributed between the different TAZs according to their geospatial location. However, if a link is found to pass through two zones, then it is divided between the two adjacent zones using a weight that is relative to the link's length within each zone.



(a) **(b)**

Figure 3 Sidewalk (a) and Bike (b) Networks Characterizations at the different TAZs

Using graph representation, several indicators can be estimated as follow:

Degree of connectivity represents the ratio between the actual number of network (bike or sidewalk) links in the TAZ and the maximum possible number of links in the TAZ. Assuming a planar graph, the maximum possible number of links within a graph is calculated according to equation 1 (Garrison and Marble, 1962), where *n* is the number of nodes within a graph.

$$l_{max} = 3\ (n-2) \qquad (1)$$

The value of the degree of connectivity is bounded between 0 and 1. A completely connected network will have a degree of connectivity equal to 1, while a completely disconnected network will have a degree of connectivity equal to 0. The degree of connectivity indicator has been used in previous studies for evaluating transit networks (Derrible and Kennedy, 2011); (Quintero et al., 2013).

As for network continuity, Scheltema (2012) previously proposed manual methodologies by counting every crossing along the key bike routes to characterize continuity. Nevertheless, such methodology is inconvenient for macro level studies as it consumes considerable time and effort. Therefore, a new way to calculate continuity needs to be applied in the context of macro-level studies. Average edge length can characterize the continuity of the transportation network since it explicitly represents the average length of the network link that is free from interruptions. Average edge length is calculated as the ratio between the total length of the zonal network and the total number of links in the corresponding TAZ (Kansky, 1963).

Lastly, the total length of the network (bike/sidewalk) links and the average weighted slope of the network links are evaluated. The total length of the zonal network represents the explicit size of the network infrastructure (sidewalks, bike lanes, etc.) within a TAZ. The average weighted slope of the zonal network gives an indication of the average steepness of the bike/sidewalk network within each zone. The total length of the bike/sidewalk network is calculated by aggregating all the network links within each TAZ. The zonal average weighted slope of the network is calculated according to the following steps. First, the links' slopes are computed (assuming absolute values) using the contour map of the city of Vancouver and then the slope at each link is given a weight relative to its length. Afterward, the average weighted slope of the links is calculated for each TAZ using equation 2; where *l* represents the link length and s represent the link's slope.

$$Average\ Weighted\ Slope\ in\ TAZ = \frac{\sum_1^n l_1 * s_1 + l_2 * s_2 + \ldots + l_n * s_n}{\sum_1^n l_1 + l_2 + .. + l_n} \qquad (2)$$

Table 1 Variables Definition and Data Summary (n=134 TAZs)

Variable	Description	Mean	SD	Min	Max
Crashes					
CColl	Cyclist-Motorist Crashes over 5 Years	12.71	13.48	0	78
PColl	Pedestrian-Motorist Crashes over 5 Years	15.45	11.45	0	54
Exposure					
VKT	Vehicle Kilometer Travelled	4290.43	3315.10	189.46	22288.79
BKT	Bike Kilometer Travelled	1047.78	2102.07	0	21462.77
W	Walk Trips	3971.64	2677.49	247.11	13906.56
Socio-Economic					
EmpD	Employment Density (Employment/Zone Area)	12236.26	26399.07	84.54	170910
HhsD	Household Density (Households/Zone Area)	416.52	436.35	0	2141.88
Land Use					
ResD	Residential Density (Residential Areas/Zone Area)	0.34	0.20	0	0.67
CommD	Commercial Density (Commercial Areas/Zone Area)	0.08	0.11	0	0.58
RecD	Recreational Density (Recreational Areas/Zone Area)	0.10	0.13	0	0.91
Built Environment					
SigD	Signal Density (Number of Signals/Zone Area)	14.26	18.43	0	110.55
StopD	Transit Stops Density (Number of Stops/Zone Area)	24.28	23.62	0	162.24
Road Facility					
ArtColl_Prop	Arterial-Collector Roads Proportion (Arterial + Collector Roads Length/ Road network Length)	0.35	0.21	0.12	1
Loc_Prop	Local Roads Proportion (Local Roads Length/Road Network Length)	0.64	0.21	0	0.87
OffSt_Prop	Proportion of Off-Street Bike Links (Total Length of Off-Street Bike Links/ Road Network Length)	0.00014	$7.51x10^{-5}$	0	0.0009
Bike Network					
CConn	Degree of Bike Network Connectivity	0.38	0.11	0	1
CAvgEdLen	Bike Network Average Edge Length	0.13	0.05	0	0.57
CSlope	Avg. Weighted Slope for Bike Network	2.52	0.90	0.63	6.65
CLen	Total Length of Bike Network Links	3.37	2.52	0	17.40
Pedestrian Network					
PConn	Sidewalk Network Connectivity	0.47	0.058	0.32	0.70
PAvgEdLen	Sidewalk Network Avg. Edge Length	109.18	23.45	57.90	242.43
PSlope	Avg. Weighted Slope of the Sidewalk Network	3.01	1.77	0.53	14.76
PLen	Total Length of Sidewalk Network	12	8.78	0.95	54.30

4. Methodology

The macro-level CPMs developed in this study use the generalized linear modeling (GLM) approach to investigate the impact the zones' attributes on crash frequency. GLM approach, which assumes a non-normal distribution error structure, is widely used as a state of practice for the development of CPMs since conventional linear regression models lack the distributional property to adequately describe collisions. This inadequacy is due to the random, discrete, nonnegative, and typically sporadic nature that characterizes the occurrence of collisions (Miaou and Lum, 1993; Sawalha and Sayed, 2006 and 2001). A Negative Binomial error distribution assumption has become the standard for CPMs developed using the GLM approach (Hauer et al., 1988); (Sawalha and Sayed, 2001). The model form used for CPMs should generally satisfy two conditions (Sawalah and Sayed, 2006). First it should not yield negative results in terms of predicting negative collisions and also should predict zero collisions for zero traffic exposure (when there are no vehicles or bikes on the road). As well, there need to be a link function to transform the model into a linear form. Based on empirical studies (Miaou and Lum, 1993); (Sawalha and Sayed, 2001), a commonly used model form includes an exposure measure (e.g. vehicle kilometers traveled) raised to some power and multiplied by an exponential function including the other non-exposure explanatory variables. The model can be expressed mathematically as shown in equation 3.

$$E(Y) = a_0 V^{a1} V^{a2} exp(\Sigma b_j x_j) \qquad (3)$$

Where $E(Y)$ is the predicted collision frequency, V is the measure of the traffic volume (VKT with BKT for cyclist CPMs and VKT with W for pedestrian CPMs), x_j represents any other explanatory variables, and a_0, a_1, a_2, and b_j are the model parameters. The recommended statistical methodology to add explanatory variables into a CPM is a forward stepwise procedure (Sawalha and Sayed, 2001). Variables are added one by one, and their significance is tested. Variables representing exposure must be included first.

Two statistical measures are used to assess the goodness of fit of the GLM models, including Pearson chi-square (χ^2) and scaled deviance (SD) statistics. For a well-fitted model and a relatively large number of observations, the expected value of Pearson χ^2 and SD will be approximately equal to the number of degrees of freedom (Sawalha and Sayed, 2001). Both Pearson χ^2 and SD are shown in equations 4 and 4 respectively, where y_i is the frequency of collisions, Var(y_i) is the variance of the frequency of collisions, E(Λ_i) is the expected frequency of collisions, and κ is the dispersion parameter. The SAS software is used develop the CPMs, undergo the significance tests for the explanatory variables, and assess the developed models' goodness of fit.

$$Pearson\ \chi^2 = \sum_{ni} [yi - E(\Lambda_i)]^2/(Var(y_i)) \qquad (4)$$

$$SD = 2\sum_{ni} [y_i ln(yi/E(\Lambda_i)) - (y_i+\kappa) ln((y_i+\kappa)/(E(\Lambda_i)+\kappa))] \qquad (5)$$

5. Analysis and Results

The developed CPMs use both VKT and BKT as main exposure variables for cyclist CPMs and VKT and W as main exposure variables for pedestrian CPMs. Tables 2 and 3 show the developed models for the active transportation safety correlates as discussed below.

Cyclist/pedestrian crash frequency is found non-linearly positively associated with the respective types of traffic exposure. The exponents of the traffic exposure variables are less than one, which support the "safety in numbers" hypothesis (Jacobsen, 2003). These results are intuitive and consistent with several previous studies (Prato et al., 2016), (Amoh-Gyimah et al., 2016).

5.1 Socio-Economic Models

Socio-economic CPMs are primarily based on the explanatory variables extracted from census data. The models reveal positive associations between the cyclist/pedestrian crashes on the one hand and the employment density and household density on the other hand. The results are reasonable since the aforementioned variables can be considered surrogate measures for traffic exposure, thereby explaining their positive associations with active commuters' crashes. These results are in agreement with previous studies by Siddiqui et al. (2012), (Cai et al., 2016) and Prato et al. (2016).

5.2 Land Use Models

The models in this category incorporate explanatory variables that refer to land zonings within the TAZs. The results show that the increase in residential and recreational area densities is associated with the decline in the number of cyclist/pedestrian crashes. The result for the recreational areas is logical because these areas usually provide off-street and continuous paths for active transportation commuters reducing the conflict risk between these vulnerable commuters and vehicles. This result is consistent with a study by Ukkusuri et al. (2011), who found a negative association between parks total area and pedestrian crashes. The negative association between residential area density and active commuters' crashes can be justified by the ongoing traffic calming measures applied by city of Vancouver to promote active transportation and limit motorized traffic at the residential neighborhoods (http://vancouver.ca/streets-transportation/traffic-calming-and-safety.aspx). This is done using speed humps, diverters, traffic circles, etc. to reduce the speed of the traffic and control its navigation within residential areas.

On the other hand, the increase in commercial area density is found associated with an increase in the cyclist/pedestrian crashes. This can be attributed to the side street activities that raise the potential risk of a cyclist/pedestrian going into conflict with motorized traffic. The association between commercial areas and active commuters' safety agrees with few previous studies (Narayanamoorthy et al., 2013), (Ukkusuri et al., 2011), (Vandenbulcke et al., 2014).

5.3 Built Environment Models

Built environment variables refer to the elements that are physically present on the road networks. The models show that cyclist/pedestrian crashes are positively associated with transit stop density and with traffic signal density, which agrees with previous studies by (Lee et al., 2015), Strauss et al. (2013) Siddiqui et al. (2012), Cai et al. (2016), and Wei and Lovegrove (2013). More traffic signals imply the presence of more wide intersections that usually include complex vehicle and active commuter maneuvers elevating the probability of crash occurrence. On the other hand, the presence of bus stops indicates the occurrence of interactions between buses, vehicles, and active commuters, which is also speculated to increase active commuters' crash risk.

5.4 Road Facility Models

For this category, a higher proportion of arterial plus collector roads is found positively associated with cyclist crashes as well as pedestrian crashes. This can be explained by the higher speeds and the heavier traffic on these types of roads, which would increase the risk of conflict occurrence between active commuters and vehicles. On the other hand, a decline in cyclist-motorist and pedestrian-motorist crash frequency is found associated with higher proportion of local roads. A likely reason for such negative association is the relatively low speeds on the local roads, which would result in an increase in the drivers' attentiveness and, therefore, reduce conflicts' potential. The former results agree with previous studies conducted by Chen (2015), Wang and Kockelman, (2013) and Siddiqui et al. (2012). The proportion of the off-street bike links is found negatively associated with cyclist crashes. This result is reasonable and agrees with several previous studies suggesting that separating bike traffic from motorized traffic would likely improve cyclist safety (Reynolds et al., 2009).

5.5 Network Models

The CPMs show that there is a positive association between the active commuters' crashes and the connectivity measure (i.e. Conn). The degree of connectivity is related to the number of network links and network configuration. The bike network connectivity is found to be positively associated with cyclist crashes, while sidewalk connectivity is positively associated with pedestrian crashes. This is likely due to the fact that more links between the nodes would lead to higher exposure to cyclist-motorist and pedestrian-motorist conflicts, and consequently higher collision potential. On the contrary, the average edge length variable is found to have a negative association with active commuters' crashes, while the average length per vertex is found to be statistically non-significant. These results imply that a higher average edge length, which indicates longer links without hindrances or discontinuities, is presumably more convenient and safer to the active commuters. Similar results were found for transit networks, where the degree of connectivity was found positively associated with crash frequency, on the contrary of the average edge length that was negatively associated (Quintero et al., 2013).

The zonal length of the bike network and sidewalk network is found negatively associated with the cyclist-motorist collisions and pedestrian-motorist collisions respectively. This agrees with recent studies that concluded that more bike infrastructure would increase cyclist safety (de Rome et al., 2014) (Prato et al., 2015), and that more pedestrian infrastructure would increase pedestrian safety (Yu, 2015), on the contrary of a study by Cai et al. (2016), who used sidewalk length as a traffic exposure for pedestrians. Also, a negative association is found between the weighted slope of the zonal bike network and the weighted slope of the zonal sidewalk network and cyclist crashes and pedestrian crashes respectively. This may be explained as cyclists and vehicles usually reduce their speeds at climbing slopes, which would lower crash risk. This result is consistent with a study by Chen (2015), who used a zonal mean slope variable to represent the average absolute slopes of the TAZs and found a non-significant negative association with cyclist crashes; as well as a study by Chen and Zhou (2016), in which they found out that the proportion of steep areas within zones was negatively associated with pedestrian crashes.

Table 2 GLM Results for Cyclist Crashes

Model Cyclist Collisions =	*K*	*df*	*SD*	X^2
Socio-Economic Models				
$0.004BKT^{0.58}VKT^{0.48}exp(8.09x10^{-6}EmpD+5.52x10^{-5}HhsD)$	2.85	128	120.05	141.53
Land Use Models				
$0.016BKT^{0.57}VKT^{0.33}exp(1.83\ CommD)$	2.64	129	111.57	142.25
$0.051BKT^{0.61}VKT^{0.23}exp(-0.56ResD^{*}-2.38RecD)$	3.01	128	125.79	144.20
Built Environment Models				
$0.010BKT^{0.54}VKT^{0.38}exp(0.015SigD)$	3.03	129	120.79	142.44
$0.012BKT^{0.58}VKT^{0.33}exp(0.010StopD)$	2.77	129	115.04	141.49
Road Facility Models				
$0.037\ BKT^{0.57}VKT^{0.31}exp\ (-0.74Loc\_Prop)$	2.56	129	117.31	142.61
$0.023\ BKT^{0.64}VKT^{0.21}exp\ (1.17ArtColl\_Prop-1884OffSt\_Prop)$	2.94	128	133.09	144.56
Bike Network Models				
$0.020\ BKT^{0.61}VKT^{0.31}exp(-0.063CLen)$	2.50	129	121.32	143.07
$0.032BKT^{0.60}\ VKT^{0.26}exp(1.64CConn-3.58\ CAvgEdLen-0.17CSlope)$	2.70	127	117.27	140.37

*Significant at the 10% level.
All other parameters are significant at the 5% level or higher

K: Over-dispersion parameter *df*: degrees of freedom X^2: Pearson chi square
SD: Scaled Deviance

Table 3 GLM Results for Pedestrian Crashes

Model Pedestrian Collisions=	*K*	*df*	*SD*	X^2
Socio-Economic Models				
$0.0008W^{0.61}VKT^{0.54}exp(2.8x10^{-5}HhsD^{*}+0.56x10^{-5}EmpD)$	5.88	129	138.40	143.79
Land Use Models				
$0.0007W^{0.69}VKT^{0.47}exp(1.40CommD)$	5.88	130	139.13	140.29
$0.0014W^{0.70}VKT^{0.45}exp\ (-1.07RecD-0.74ResD)$	6.13	129	138.59	142.86
Built Environment Models				
$0.001W^{0.64}VKT^{0.49}exp\ (0.011SigD)$	6.36	130	138.48	143.51
$0.0008W^{0.68}VKT^{0.47}exp\ (0.0079StopD)$	5.95	130	137.95	140.54
Road Facility Models				
$0.0005W^{0.72}VKT^{0.47}exp\ (0.81Art\_Coll)$	6.09	130	138.76	144.98
$0.0012W^{0.72}VKT^{0.47}exp\ (-0.79Loc\_Prop)$	6.02	130	138.66	144.59
Pedestrian Network Models				
$0.31VKT^{0.32}exp(2.55PConn)$	2.38	131	143.63	153.48
$0.0014W^{0.76}VKT^{0.47}exp\ (-0.0091PAvgEdLen)$	5.88	130	136.68	138.98
$0.0003W^{0.73}VKT^{0.65}exp(-0.035PLen-0.087PSlope)$	7.46	129	135.55	138.69

*Significant at the 10% level.
All other parameters are significant at the 5% level or higher

K: Overdispersion parameter *df*: degrees of freedom X^2: Pearson chi-square *SD*: Scaled Deviance

6. Conclusions

This paper provides transportation engineers, planners, and policy makers with tools (CPMs) that can be used to improve active commuters' safety. The study used large GIS data from city of Vancouver 134 traffic analysis zones, to develop empirical macro-level CPMs incorporating variables related to traffic exposure, socioeconomics, land use, built environment, and road facility. In addition, bike and sidewalk network indicators were developed using graph theory. A GLM technique was used to build cyclist and pedestrian CPMs incorporating the various explanatory variables aforementioned. The cyclist CPMs were developed using VKT and BKT, while the pedestrian CPMs were developed using VKT and W. BKT and W represent the actual cyclist and pedestrian traffic exposure respectively unlike previous safety studies that used proxies for cycling and pedestrian exposure. Some findings in this study agreed with the results from former studies in the literature, though better exposure indicators and more comprehensive set of cyclist and pedestrian safety correlates were used in the current study.

The CPMs' results showed that the cyclist crashes and pedestrian crashes had almost similar associations with the studied zone characteristics. The cyclist/pedestrian crashes were non-linearly and positively associated with the traffic exposure variables. The exponents of the exposure measures were less than one supporting the "safety in numbers" hypothesis. The results also showed that the increase in the cyclist/pedestrian crashes was associated with the increase in the socio-economic attributes such as employment and household densities, and the built environment attributes such as transit stop and traffic signal densities. Regarding, land use, a positive association was found between cyclist/pedestrian crash frequency and commercial area density, while both residential and recreational areas' densities had negative associations with the active commuters' crashes. For road network facilities, higher cyclist/pedestrian crash frequency was found associated with more arterial and collector roads proportion, while a decline in those crashes was found associated with the increase in local roads proportion. Cyclist crashes were negatively associated with the off-street bike links proportion. Bike and sidewalk networks' connectivity was found positively associated with cyclist and pedestrian crashes respectively, on the contrary of the networks' continuity, slope, and length.

Several recommendations can be suggested to improve active commuters' safety based on this study as follow. It is obvious that it is not enough to just provide bike and sidewalk networks; rather, the quality of the network is an important factor. The length, connectivity, continuity, and the slope of the bike and pedestrian networks need to be prudently studied and addressed. In addition, the signalized and non-signalized intersections need to be treated according to complete streets acts (such as NACTO guides) to be friendlier and safer to the active commuters. Separated bus lanes, bike lanes, and walkways can also be good solutions for cyclist and pedestrian interactions with transit and vehicles. Moreover, mixed land use, that provides shorter commute distance for active commuters, can reduce the crash risks for those users. Lastly, a better hierarchy of the streets, that can reduce the dependence on high speed and high traffic volume roads, shall make active commuters more comfortable and safer.

Some areas of further research can be also investigated. First, the transferability of the models needs to be validated by applying the developed CPMs to various cycling and walking environments. Also, the models can be used to detect hot zones for active transportation, and consequently provide remedies to elevate the safety of active commuters at these areas. Lastly, the association between the bikeability and walkability and the cyclist and pedestrian crashes can be an interesting topic for future research.

7. References

Amoh-Gyimah, Richard, Majid Sarvi, and Meead Saberi. "Investigating the Effects of Traffic, Socioeconomic, and Land Use Characteristics on Pedestrian and Bicycle Crashes: A Case Study of Melbourne, Australia." *Transportation Research Board 95th Annual Meeting*. No. 16-1931. 2016.

Berrigan, D., Pickle, L.W., Dill, J.: Associations between street connectivity and active transportation. Int. J. Health Geogr. 9 (20) (2010). doi: 10.1186/1476-072X-9-20.

Cai, Qing, et al. "Macro-level pedestrian and bicycle crash analysis: incorporating spatial spillover effects in dual state count models." *Accident Analysis & Prevention* 93 (2016): 14-22.

Chen, Peng, and Jiangping Zhou. "Effects of the built environment on automobile-involved pedestrian crash frequency and risk." *Journal of Transport & Health* (2016).

Chen, Li, et al. "Evaluating the safety effects of bicycle lanes in New York City." *American journal of public health* 102.6 (2012): 1120-1127.

Chen, Peng. "Built environment factors in explaining the automobile-involved bicycle crash frequencies: a spatial statistic approach." *Safety science* 79 (2015): 336-343.

De Hartog, J. J., Boogaard, H., Nijland, H., & Hoek, G. (2010). Do the health benefits of cycling outweigh the risks?. *Environmental health perspectives*, 1109-1116.

de Rome, Liz, et al. "Bicycle crashes in different riding environments in the Australian capital territory." *Traffic injury prevention* 15.1 (2014): 81-88.

Derrible, Sybil, and Christopher Kennedy. "Characterizing metro networks: state, form, and structure." *Transportation* 37.2 (2010): 275-297.

El Esawey, Mohamed, Clark Lim, and Tarek Sayed. Development of a cycling data model: City of Vancouver case study. *Canadian Journal of Civil Engineering*, Vol. 42, No. 12, 2015, pp. 1000-1010.

Garrison, William L., and Duane F. Marble. *The structure of transportation networks*. No. TR62 11. NORTHWESTERN UNIV EVANSTON ILL, 1962.

Gattuso, Domenico, and Ernesto Miriello. "Compared analysis of metro networks supported by graph theory." *Networks and Spatial Economics* 5.4 (2005): 395-414.

Hamann, Cara, and Corinne Peek-Asa. "On-road bicycle facilities and bicycle crashes in Iowa, 2007–2010." *Accident Analysis & Prevention* 56 (2013): 103-109.

Jacobsen, Peter L. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury prevention*, Vol. 9, No. 3, 2003, pp. 205-209.

Kansky, Karel Joseph. "Structure of Transportation Networks: Relationships Between Network Geometry and Regional Characteristics." University of Chicago Press. Chicago, IL. (1963).

Kim, Joon-Ki, et al. "Bicyclist injury severities in bicycle–motor vehicle accidents." *Accident Analysis & Prevention* 39.2 (2007): 238-251.

Kim, Joon-Ki, et al. "A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model." *Accident Analysis & Prevention* 42.6 (2010): 1751-1758.

Kim, Karl, Pradip Pant, and Eric Yamashita. "Accidents and accessibility: Measuring influences of demographic and land use variables in Honolulu, Hawaii." *Transportation Research Record: Journal of the Transportation Research Board* 2147 (2010): 9-17.

Lee, Jaeyoung, Mohamed Abdel-Aty, and Ximiao Jiang. "Multivariate crash modeling for motor vehicle and non-motorized modesat the macroscopic level." *Accident Analysis & Prevention* 78 (2015): 146-154.

Miaou, Shaw-Pin, and Harry Lum. Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention*, Vol. 25, No. 6, 1993, pp. 689-709.

Miranda-Moreno, Luis, Jillian Strauss, and Patrick Morency. "Disaggregate exposure measures and injury frequency models of cyclist safety at signalized intersections." *Transportation Research Record: Journal of the Transportation Research Board* 2236 (2011): 74-82.

Narayanamoorthy, Sriram, Rajesh Paleti, and Chandra R. Bhat. "On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level." *Transportation research part B: methodological* 55 (2013): 245-264.

Prato, Carlo Giacomo, et al. Infrastructure and spatial effects on the frequency of cyclist-motorist crashes in the Copenhagen Region. *Journal of Transportation Safety & Security*, 2016, pp. 1-15.

Quintero, Liliana, Tarek Sayed, and Mohamed M. Wahba. "Safety models incorporating graph theory based transit indicators." *Accident Analysis & Prevention* 50 (2013): 635-644.

Quintero-Cano, L., Wahba, M. and Sayed, T. (2014) "Bus Networks As Graphs: New Connectivity Indicators with Operational Characteristics", Canadian Journal of Civil Engineering, Vol. 41, pp. 788–799.

Reynolds, C. C., et al. "The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature." *Environmental Health* 8.1 (2009): 47.

Rodrigue, J.P., Comtois, C., Slack, B.: The Geography of Transport Systems, 2nd edn. Routledge, New York (2009).

Robinson, Dorothy L. "Safety in numbers in Australia: more walkers and bicyclists, safer walking and bicycling." *Health promotion journal of Australia* 16.1 (2005): 47-51.

Sawalha, Ziad, and Tarek Sayed. "Evaluating safety of urban arterial roadways." *Journal of Transportation Engineering* 127.2 (2001): 151-158.

Sawalha, Ziad, and Tarek Sayed. Traffic accident modeling: some statistical issues. *Canadian Journal of Civil Engineering*, Vol. 33, No. 9, 2006, pp. 1115-1124.

Scheltema, E. B. *ReCYCLE City: Strengthening the bikeability from home to the Dutch railway station*. Diss. TU Delft, Delft University of Technology, 2012.

Siddiqui, Chowdhury, Mohamed Abdel-Aty, and Keechoo Choi. "Macroscopic spatial analysis of pedestrian and bicycle crashes." *Accident Analysis & Prevention* 45 (2012): 382-391.

Strauss, Jillian, Luis F. Miranda-Moreno, and Patrick Morency. "Cyclist activity and injury risk analysis at signalized intersections: A Bayesian modelling approach." *Accident Analysis & Prevention* 59 (2013): 9-17.

Teschke, Kay, et al. "Route infrastructure and the risk of injuries to bicyclists: a case-crossover study." *American journal of public health* 102.12 (2012): 2336-2343.

Toroyan, Tami. "Global status report on road safety 2015." *Supporting a decade of action. Geneva: World Health Organization, Department of Violence and Injury Prevention and Disability* (2013).

Ukkusuri, Satish, Samiul Hasan, and H. Aziz. "Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency." *Transportation Research Record: Journal of the Transportation Research Board* 2237 (2011): 98-106.

Vandenbulcke, Gregory, Isabelle Thomas, and Luc Int Panis. "Predicting cycling accident risk in Brussels: A spatial case–control approach." *Accident Analysis & Prevention* 62 (2014): 341-357.

Wang, Yiyi, and Kara M. Kockelman. "A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods." *Accident Analysis & Prevention* 60 (2013): 71-84.

Wei, Feng, and Gordon Lovegrove. "An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression." *Accident Analysis & Prevention* 61 (2013): 129-137.

Xie, F., Levinson, D.: Measuring the structure of road networks. Geogr. Anal. 39 (3), 336–356 (2007). doi: 10.1111/j.1538-4632.2007.00707.x

Yu, Chia-Yuan. "Built Environmental Designs in Promoting Pedestrian Safety." *Sustainability* 7.7 (2015): 9444-9460.