Accident Analysis and Prevention 137 (2020) 105439

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

A composite zonal index for biking attractiveness and safety

Mohamed Bayoumi Kamel*, Tarek Sayed, Alexander Bigazzi

Department of Civil Engineering, University of British Columbia, 6250 Applied Science Lane, Vancouver, BC, V6T 1Z4, Canada

ARTICLE INFO

Keywords:
Bikeability
Cyclist safety
Biking attractiveness
Bike index

ABSTRACT

Zonal characteristics (e.g. built environment, network configuration, socio-demographics, and land use) have been shown to affect biking attractiveness and safety. However, previously developed bikeability indices do not account for cyclist-vehicle crash risk. This study aims to develop a comprehensive zone-based index to represent both biking attractiveness and cyclist crash risk. The developed Bike Composite Index (BCI) consists of two sub-indices representing bike attractiveness and bike safety, which are estimated using Bike Kilometers Travelled (BKT) and cyclist-vehicle crash data from 134 traffic analysis zones (TAZ) in the City of Vancouver, Canada. The Bike Attractiveness Index is calculated from five factors: bike network density, centrality, and weighted slope as well as land use mix and recreational density. The Bike Safety Index is calculated from bike network coverage, continuity, and complexity as well as signal density and recreational density. The correlation between the Bike Attractiveness Index and the Bike Safety Index in Vancouver is low ($r = 0.11$), supporting the need to account for both biking attractiveness and safety in the composite index.

1. Introduction

Many cities are promoting active transportation (e.g. walking and biking) to develop more sustainable and livable communities and improve public health (Banister, 2008; Giles-Corti et al., 2010; Pucher and Buehler, 2010). However, cyclists are vulnerable road users that can have elevated injury and fatality risks compared to vehicle drivers and passengers (Safety of vulnerable road users (DSTI/DOT/RTR/RS7(98) 1/FINAL No. 68074), 1998; World Health Organization, 2009). Therefore, there is a growing interest in evaluating and communicating the various factors (e.g. network characteristics, socio-demographics, and land use) associated with biking activity and safety. Indices are a common tool to summarize the combined effects of various factors that influence active travel (Carr et al., 2010; Harkey et al., 1998; Landis et al., 1997; Larsen et al., 2013; Lin and Wei, 2018; Winters et al., 2013). However, existing “bikeability” indices do not account for objective crash risks for cyclists while accounting for bike exposure.

To develop a comprehensive city biking index, two major aspects should be considered: the attractiveness of cycling and the risk of injury or fatality for cyclists. The attractiveness of cycling is derived from several factors including land use, socio-demographics, trip types and distances, cycling facilities, terrain, weather, and the availability and attractiveness of other modes of travel. Cyclist risk of injury or fatality comes from the likelihood of crashes of various types, either with or without interactions with other road users such as motorists, pedestrians, or other cyclists. The crash likelihood is also derived from several factors including traffic volumes, and many of the factors that influence attractiveness such as cycling facilities.

This study aims to develop a statistically calibrated composite zonal index (Bike Composite Index) that accounts for both attractiveness to biking and cyclist-vehicle crash risk. The proposed Bike Composite Index (BCI) consists of two sub-indices: Bike Attractiveness Index and Bike Safety Index. The indices are developed using Bike Kilometers Travelled (BKT) and cyclist-vehicle crash data from 134 traffic analysis zones (TAZ’s) in the City of Vancouver, Canada.

2. Literature

This section is divided into four parts: factors affecting biking levels, factors affecting biking safety, tools for biking attractiveness assessment, and tools for cyclist crash risk assessment.

2.1. Factors affecting biking levels

Extensive research describes the importance of bike network connectivity for biking levels (Berrigan et al., 2010; Cervero et al., 2019; Marshall and Garrick, 2010; Mekuria et al., 2012; Osama et al., 2017; Schoner and Levinson, 2014; Winters et al., 2016). Berrigan et al. (2010) analyzed showed that network connectivity has a positive Pearson correlation with walking and biking levels by investigating

* Corresponding author.
E-mail addresses: bayoumi1@mail.ubc.ca (M.B. Kamel), tsayed@civil.ubc.ca (T. Sayed), abigazzi@civil.ubc.ca (A. Bigazzi).

https://doi.org/10.1016/j.aap.2020.105439
Received 2 October 2019; Received in revised form 6 January 2020; Accepted 9 January 2020
Available online 28 January 2020

2001 counties' data from Los Angeles and San Diego. Marshall and Garrick (2010) used data from 24 Californian cities to investigate the effect of street connectivity and intersection density on the choice to drive, walk, bike, or transit. They found that intersection density and street connectivity were associated with more walking, biking, and transit use. Network continuities are perceived as important by cyclists (Handy and Xing, 2011; Osama et al., 2017). Bike lanes interruptions (discontinuities) may compel a cyclist to cycle in mixed traffic or take a longer route to the cyclist destination (Schoner and Levinson, 2014). Various studies investigated the impact of land use patterns on biking levels in North America. Mixed land use was found correlated with biking levels in Portland Oregon (Dill and Voros, 2007) and in San Diego (Jones et al., 2010). Offices, fast-food restaurants, hospitals, together with multifamily residential settings, may support biking (Moudon et al., 2005). Additionally, various studies investigated the impact of network facilities on biking levels. Wilkinson (1994) suggested adding bike lanes, separated paths, bike boulevards, and local streets to the street network to calm traffic and encourage biking. Parkin et al. (2008) found that the proportion of the off-street bike routes had a significant positive association with bike ridership, which is consistent with the findings of Caulfield et al. (2012). Marshall and Garrick (2011) showed that arterials and collectors streets were less friendly and perceived as less safe by cyclists. Tilahun et al. (2007) showed that the presence of on-street parking parallel to bike lanes reduced the utility of those bike lanes. Lastly, different studies have discussed the relationship between demographic variables and biking levels. Schneider and Stefanich (2015) showed that household density and population density have a [positive association with bike commuting by using census tract data in Wisconsin, USA.

2.2. Factors affecting biking safety

Traffic exposure is usually included in traffic crash analysis to equalize for differences in intensity of use that make comparisons meaningful (Hauer, 1995). Osama and Sayed (2016) found that bike network length is associated with a decrease in cyclist-vehicle crashes while accounting for bike traffic exposure through bike kilometers traveled. For street network configuration, previous research investigated the effect of connectivity, directness, density, length, and topography. For example, Osama and Sayed (2016) found positive associations between cyclist crashes and the bike network connectivity and linearity (the ratio between curved links and its effective straight link). In contrast, they found negative associations between cyclist crashes and the bike network continuity (bike lane continuation without intersecting with the street element) and slope. Regarding land use, the increase in the residential area, industrial area, commercial area, and land use mix was positively associated with cyclist crashes, while accounting for bike traffic exposure through population density (Amoh-Gyimah et al., 2016; Narayanamoorthy et al., 2013; Vandenbulcke et al., 2014). Similarly, Osama and Sayed (2016) used bike kilometers traveled to account for bike exposure, and they concluded that residential area and commercial area were positively associated with cyclist crashes, while the recreational area was negatively associated with cyclist crashes. In terms of demographic variables, Siddiqui et al. (2012) showed that population, employment, and median household income were positively associated with cyclist crash frequency. As for travel demand variables, bike traffic (Miranda-Moreno et al., 2011; Strauss et al., 2013), as well as vehicle traffic (Hamann and Peek-Asa, 2013), were found positively associated with cyclist-vehicle crash frequency. Traffic signal density was found positively associated with cyclist crashes while accounting for bike traffic exposure (Chen, 2015; Wei and Lovegrove, 2013). The off-street bike lanes were found safer than the on-street ones while controlling for bike traffic exposure (Hamann and Peek-Asa, 2013; Reynolds et al., 2009; Teschke et al., 2012). Recently, Kamel et al. (2019) investigated the mediated effects that some zonal characteristics variables have on cyclist crashes through their effects on bike exposure (by setting bike exposure as a mediator). They found that bike network coverage and recreational density have a negative direct association with cyclist-vehicle crashes, and the positive indirect association leading to a positive total effect on cyclist-vehicle crashes.

2.3. Tools for biking attractiveness assessment

Mainly there are two units of analysis for developing bike attractiveness tools: roadway level and zone level. Assessment tools that are suitable for roadway segments were previously developed (Harkey et al., 1998; Landis et al., 1997). Landis et al. (1997) and Harkey et al. (1998) employed linear regression modeling to calibrate index weightings. Landis et al. (1997) developed the Bike Level of Service (BLOS) method that is based on cyclists' perceptions from traveling on roadways, while Harkey et al. (1998) relied on video data review to obtain the perspectives of cyclists.

More recently, other studies used spatial data to evaluate zonal attractiveness to biking (Bike Score Methodology [WWW Document], 2019; Larsen et al., 2013; Lin and Wei, 2018; Winters et al., 2013). Table 1 gives a summary of previous studies that developed bike zone assessment tools. Lin and Wei (2018) employed an analytic network process to manage the interdependences among criteria and zones and employed grey numbers (A grey number represents the possible values in a range rather than the exact value) to measure possible ranges of criteria performances and handle various performances within a zone. Lin and Wei (2018) relied on the literature and stakeholder interviews to determine the criteria weighting. Larsen et al. (2013) relied on an

Table 1
Summary of previous studies that developed zonal bikeability assessment tools.

Study/location	Data used for index development	Factors	Spatial unit of Analysis	Weighting Methodology
Lin and Wei (2018)/ Taipei, Taiwan	• Evidence from previous research	Bike network configuration; street network; pedestrian facility; public transportation facility; green areas; land use mix; and cyclist comfort perception	TAZ	Analytic Network Process
Winters et al. (2013)/ Vancouver, Canada	• Opinion survey • Travel behavior analyses • Focus groups	Bike network configuration; bike-friendly destinations density; and topography	10-m grid	Equal weight
Larsen et al. (2013)/ Montreal, Canada	• Cyclists online survey • Cyclist-vehicle crashes • Origin-destination survey	Street network configuration; observed and potential biking trips expected on different links; survey respondents for upgrading bike priorities; and cyclist-vehicle crashes	300-m grid	Equal weight
Bike Score®/ North America	• Walk Score community votes • Expert advice from Professors	Bike lane score, hill score, destinations and street connectivity, and bike commuting mode share	*	Equal weight

* Each factor has a different spatial unit of analysis that is described in the "Comparison to Bike Score®" Section.

online survey to develop a tool that helps planners to determine the optimal location for new bike facilities, while accounting for cyclist-vehicle crashes, but without accounting for traffic exposure. Winters et al. (2013) employed an opinion survey and focus group data to develop a tool that identifies areas that are more or less suitable to biking by employing five assessment factors: bike network density, route separation (as a dummy variable), bike-friendly streets connectivity, bike-friendly destination density, and topography. More recently, the Bike Score® was developed to measure whether a location is good for biking on a scale from 0 to 100 based on four equally weighted components: bike lane score, hill score, destinations and street connectivity, and bike commuting mode share (Bike Score Methodology [WWW Document], 2019). The Bike Score® methodology is based on votes from the Walk Score® community and expert advice from researchers at the University of British Columbia (Bike Score Methodology [WWW Document], 2019).

2.4. Tools for cyclist crash risk assessment

Most traditional ranking/identification methods of high-crash zones have relied on historical traffic crash records to obtain an estimate of safety for diverse traffic entities. These simple methods included the crash frequency method (Deacon et al., 1974), crash rate method, the rate-quality control method (Stokes and Mutabazi, 1996), the crash severity method, and summation of the ranks, the safety index method (Tamburri and Smith, 1970). Even though these approaches are simple, they suffer from limitations such as the “regression to the mean” issue (Hauer, 1997), and the incapability for accounting for cash data dispersion. Therefore, several studies adopted a model-based ranking approach. The Empirical Bayes (EB) approach is considered the state-of-the-practice for identification and ranking of the crash-prone location and has been utilized in many recent studies (e.g. Cheng et al. (2018) and Yang and Loo (2016)). A useful indicator to measure the expected safety benefits is to employ the EB approach to calculate the Potential for Safety Improvement (PSI) (Sayed and Rodriguez, 1999).

Alternatively, to account for uncertainty in inferences based on statistical data analysis, the Full Bayes approach (FB) is usually employed. Aside from the PSI method of ranking mentioned above, Schlüter et al. (1997) described three methods of rankings employing the hierarchical Bayesian model, namely the posterior probability, the predictive probability, and the Posterior Mean (PM). The PM ranking is comparing the sites based on the expected crash frequency given the observed data. Miranda-Moreno and Fu (2007) explored the differences between EB and FB approaches, where they employed both approaches with the Posterior Mean (PM) method to rank sites. They showed that FB is superior in the case of a small dataset, however for a large dataset (e.g. more than 300 sites), the FB and EB approaches had similar performance.

A few studies have ranked hazardous locations based on cyclist crashes. For example, van den bossche et al. (2002) employed a Bayesian binomial to model cyclist crashes and employed PM to identify and rank the most hazardous locations. Osama and Sayed (2019) employed a multivariate full Bayesian spatial mixed crash model (CM) to model cyclist and pedestrian crashes while accounting for the motorized and non-motorized traffic exposure measures. They identified and ranked active transportation crashes hotspots using the Mahalanobis distances, where they showed that the Mahalanobis distances are more efficient than the traditional PSI method for site ranking.

3. Method

3.1. Index development

Different disciplines (e.g. social science and economics) have given indices substantial attention, where various techniques have been developed to handle different aspects of the index development. A number of techniques have been explored, such as aggregation techniques, multiple regression analysis, principal components and factor analysis, efficiency frontier, and experts’ opinion (budget allocation).

This section describes the followed approach to developing a zone-based Bike Composite Index (BCI). BCI consists of a Bike Attractiveness Index (BAI) to represent the attractiveness of the zone for biking and a Bike Safety Index (BSI) to represent cyclist crash risk in the zone.

Indices are developed not only to accommodate but also encourage biking, particularly in North America. In addition, these indices are broadly-accessible, user-friendly tools to incorporate a variety of important influencing factors on a single scale. Generally, indices are developed to 1) summarize multi-dimensional characteristics with a view to supporting decision-makers; 2) access the progress over time; 3) facilitate the communication with the general public. However, indices’ users and developers should be careful of sending a misleading message that could lead to inappropriate policies and potential misuse to support the desired policy.

Bike Kilometers Travelled (BKT) by zone is used to develop the BAI, by employing multiple regression analysis as an index development tool, where the regression coefficients are considered the sub-index weights (Elvik, 2007; Porter and Stern, 1999). It might be argued that if the concepts to be measured (e.g. biking attractiveness and cyclist crash risk) could be represented by a single indicator (e.g. Bike kilometers traveled, cyclist-vehicle crashes), then there would be no need for developing BAI and BSI. However, the set of sub-index considered as input in the regression model could be related to various policy actions. The model could then quantify the relative effect of each policy action on the target.

Bike Kilometers Travelled (BKT) by zone partially represents biking propensity if the zones are constructed to contain similar populations (as is typical with Traffic Analysis Zones). A BKT prediction model was estimated using lognormal regression, as the model should not yield a negative result and the BKT distribution is right-skewed. The model form is shown in Eq. (1), where b_0 and b_m are estimated model parameters, m is the number of zonal characteristic variables, X_{mi} is the zonal characteristic variables, and u_i is an error term.

$$\ln(BKT_i) = b_0 + \Sigma_m b_m X_{mi} + u_i \tag{1}$$

In the model development process, zonal characteristic variables that capture the same indicator are not allowed in the same model. For example, network connectivity indicators (intersection density, network density, and network coverage), are not included in the same model. Extensive previous research showed the importance of network weighted slope and network connectivity on zonal biking attractiveness (Hood et al., 2011; Marshall and Garrick, 2010; Osama et al., 2017; Schoner and Levinson, 2014; Winters et al., 2013). As a result, these two dimensions were forced in the developed models: network weighted slope and a network connectivity indicator. Various models were developed using the criteria described above. Out of the developed models, only models where all the independent variables are significant at 10 % were included as candidates, while the other tested models were discarded. Then, from the list of candidate models, the model with the lowest five-fold cross-validation Root Mean Square Error (RMSE) was selected to enhance the model transferability.

The Bike Attractiveness Index (BAI) is derived from the selected BKT model using the parameter estimates as weights, as shown in Eq. (2).

$$BAI'_i = b_0 + \Sigma_m b_m X_{mi} \tag{2}$$

Where, BAI'_i is the raw Bike Attractiveness Index. It should be noted that, In Eq. (7), the natural logarithm is applied to the right-hand side of the equation. As the below mentioned BAI and BSI aggregation process requires the BAI'_i distribution to be as close as possible to a normal distribution.

Similar to the BAI development process, multiple regression analysis is employed to develop the Biking Safety Index (BSI). Zonal cyclist-vehicle crash model is used to develop the BSI while controlling for traffic exposure to quantify the cyclist crash risk. Let Y_i denote the number of crashes at site i (i = 1...,n) and assume that crashes at the n sites are independent and that

$$Y_i|\theta_i \sim Poisson(\theta_i) \tag{3}$$

where θ_i is the Poisson parameter. A crash prediction model was estimated using a generalized linear model with a Negative Binomial error distribution, consistent with previous safety research (Hauer et al., 1988; Sawalha and Sayed, 2001). To address over-dispersion for unobserved heterogeneity, it is assumed that

$$\theta_i = \mu_i \exp(u_i) \tag{4}$$

and

$$\ln(\mu_i) = c_o + c_1 \ln(VKT_i) + c_2 \ln(BKT_i) + \Sigma_m d_m X_{mi} \tag{5}$$

where the term $\exp(u_i)$ represents a multiplicative random effect as shown in Eq. (6), c_0, c_1, c_2, and d_m are estimated model parameters, VKT_i is the vehicle exposure variable (Vehicle Kilometers Travelled in the zone), BKT_i is the bike exposure variable, X_{mi} represents the zonal characteristic variables.

$$\exp(u_i)|\varkappa \sim Gamma(\varkappa, \varkappa) \tag{6}$$

Where $\varkappa$ is the inverse dispersion parameter. The same model development process to select zonal variables was used as the specification for the BAI described above, while forcing in two dimensions: vehicle kilometers traveled and bike kilometers traveled, to capture the cyclist crash risk, as shown in Eq. (5).

The Bike Safety Index (BSI) is derived from the crash prediction model. Then bike exposure variable (BKT) was removed, as shown in equation Eq. (7). To develop a safety index that is not biased to zones with more cyclists. In other words, BKT_i should be removed to capture the cyclist-vehicle crash risk instead of the crash frequency. In Eq. (7) the crash model is multiplied by -1, to invert the scale making high numbers represent lower crash risk this will show usefulness in the scaling process described below.

$$BSI'_i = (c_o + c_1 \ln(VKT_i) + \Sigma_m d_m X_{mi}) * -1) \tag{7}$$

Where, BSI'_i is the raw Bike Attractiveness Index. It should be noted that, In Eq. (7), the natural logarithm is applied to the right-hand side of the equation. The below mentioned BAI and BSI aggregation process requires BSI_i distribution to be as close as possible to a normal distribution.

Next, the Bike Attractiveness Index (BAI) and the Bike Safety Index (BSI) are aggregated to develop the Bike Composite Index (BCI). In this case, the reviewed practice tends to assume that items are weighted equally, and the obligation of proof should be on differential weighting (equal weighting is the standard). However, to maximize the statistical information from both, Principal Component Analysis (PCA) has been utilized. This is an established method to define the weights for the composite index (e.g. Internal Market Index (Markt, 2001)).

The Bike Attractiveness Index (BAI) and the Bike Safety Index (BSI) were aggregated into the Bike Composite Index (BCI) using Principal Component Analysis (PCA). PCA's objective is to take a number of variables (BAI and BSI) and find a linear combination of these variables to produce uncorrelated components. The statistical rationale for the transformation of BAI and BSI variables towards a normal distribution prior to PCA is to facilitate a meaningful quantification of variance and to reduce heteroskadacity. Transformation of variables towards normal distributions before employing PCA is a common practice (e.g. Wold et al. (1987); Baxter (1995)).

In order to avoid one variable (BAI or BSI) having an unjustified influence on the principal components, it is common to standardize the variables (BAI and BSI) to have means of zero and unit variances at the start of the analysis. Therefore, the BAI and BSI were standardized using Eq. (8). The resulted first component loadings (W_1 and W_2) from the PCA were used as the Bike Composite Index weighting values, as shown in Eq. (9).

$$z''_i = \frac{w_i - \omega}{\sigma} \tag{8}$$

$$BCI'_i = W_1 * BAI''_i + W_2 * BSI''_i \tag{9}$$

Where w_i is observations, ω is the mean, σ is the standard deviation, z''_i is the standardized values. In addition, BCI'_i is the raw Bike Composite Index, BAI''_i is the standardized values of the raw Bike Attractiveness Index (BAI'_i), and BSI''_i is the standardized values of the raw Bike Safety Index (BSI'_i).

Finally, BAI'_i, BSI'_i, and BCI'_i are scaled from 0 to 100 using Eq. (10), for a better presentation.

$$z_i = 100 * \frac{z'_i - \min(z'_i)}{\max(z'_i) - \min(z'_i)} \tag{10}$$

Where z'_i are the index raw values, and z_i are the indices scaled from 0 to 100 values. Note that the intercept in Eq. (2&7) may be discarded, as the indices are scaled.

3.2. Data sources

The BAI and BSI models are based on data from 134 Traffic Analysis Zones (TAZ) in the city of Vancouver, Canada. The TAZ data were from the following sources:

1 The Insurance Corporation of British Columbia (ICBC), a public automobile insurance company, provided crash data for a five-year period (2009–2013) with location information. Only cyclist-vehicle crashes are included in the analysis. All three reported severity levels (fatality, injury, and property damage only) were included. Cyclist-vehicle crashes are aggregated at the different TAZ according to the reported location. Crashes reported occurring on TAZ boundaries were distributed to adjacent TAZ as fractions based on BKT proportions between zones (rounded to whole numbers) (Osama and Sayed, 2016).
2 TransLink, the transportation planning agency for metropolitan Vancouver, provided geocoded files of the bike network, street network, and TAZ boundaries in 2013. TransLink provided Annual Average Daily Traffic (AADT) for the city of Vancouver network for the year 2011. In addition, TransLink provided population, employment, and household data for each TAZ.
3 Acuere Analytics provided the Vancouver Biking Data Model (VCDM), which gives estimates of the annual average daily bike traffic (AADB) in 2011 on the entire City of Vancouver bike network. The VCDM is based on bike counts from 2005 through 2011, including more than 810,000 hourly volumes over 7 years (El Esawey et al., 2015). The VCDM calculated AADB for 1645 links out of the 2328 single-direction bike links.
4 The open data catalog of the City of Vancouver (Vancouver, 2013) provided transportation system data (transit stops and traffic signals), terrain data (as a 1-m Digital Elevation Model), and zoning data for the City of Vancouver.

It should be noted that some cyclist-vehicle crashes are not reported, and the crash data excluded cyclist-cyclist and cyclist-pedestrian crashes. Furthermore, the AADB was computed on links of the bike network, while excluding local streets and other street network facilities. Lastly, zoning is not the same as land use; however, in this study, zoning date is used as an available proxy for actual land use.

3.3. Analysis variables

Definitions of the variables that are used in the analysis and their summary statistics are presented in Table 2. The variables are divided into three main categories: crashes, traffic exposure, and zonal

Table 2
Data summary statistics.

Variable	Mean	SD*	Min*	Max*
Crashes				
Cyclist-vehicle crashes over five years	12.71	13.48	0.00	78.00
Traffic Exposure				
VKT (Vehicle Kilometers Traveled) in thousands of Kilometers	4.29	3.33	0.19	22.29
BKT (Bike Kilometers Travelled) in thousands of Kilometers	1.05	2.11	0.00	21.46
Zonal Characteristics				
Bike Network				
Centrality Indicators				
Betweenness Centrality	0.01	0.01	0.00	0.07
Degree Centrality	0.05	0.03	0.01	0.20
Complexity Indicators				
Complexity	1.39	1.41	0.00	6.00
Pi Index	190.42	107.60	0.00	791.31
Connectivity Indicators				
Bike Network Density	5.38	3.78	0.00	21.91
Bike Network Coverage	0.34	0.19	0.00	1.01
Intersection Density	74.28	33.8	6.08	235.95
Directness Indicators				
Bike Network Average Link Length	0.13	0.05	0.00	0.57
Bike Network Linearity	0.68	0.27	0.00	1.00
Miscellaneous Indicators				
Total Length of Bike Network Links (km)	3.37	2.53	0.00	17.41
Bike Links with On-street Parking Proportion (Length of Bike Links with On-street Parking/Bike Network Length)	0.31	0.26	0.00	1.00
Off-Street Bike Links Proportion (Length of Off-Street Bike Links/Bike Network Length)	$0.01*10^{-2}$	$0.08*10^{-3}$	0.00	$0.09*10^{-2}$
Topography				
Bike Network Slope	2.53	0.9	0.64	6.66
Street Network				
Signal Density (Number of Signals/Zone Area in km^2)	14.27	18.5	0.00	110.55
Bus Stops Density (Number of bus Stops/Zone Area in km^2)	24.29	23.62	0.00	162.25
Arterial Streets Proportion (of all street links in the TAZ, by length)	0.22	0.22	0.00	1.00
Collector Streets Proportion (of all street links in the TAZ, by length)	0.13	0.10	0.00	0.55
Local Streets Proportion (of all street links in the TAZ, by length)	0.64	0.21	0.00	0.88
Land Use				
Recreational Density (Recreational Areas/Zone Area)	0.10	0.13	0.00	0.91
Residential Density (Residential Areas/Zone Area)	0.34	0.2	0.00	0.67
Industrial Density (Residential Areas/Zone Area)	0.03	0.08	0.00	0.61
Institutional Density (Residential Areas/Zone Area)	0.03	0.05	0.00	0.43
Commercial Density (Commercial Areas/Zone Area)	0.05	0.11	0.00	0.58
Land Use Mix	0.45	0.13	0.05	0.74
Demographics				
Employment Density (Employment/Zone Area in km^2)	12236.3	26399.1	84.5	170910.1
Household Density (Households/Zone Area in km^2)	4214.73	4328.55	0.00	21418.85
Population Density (Population/Zone Area in km^2)	8391.82	6995.86	0.00	33658.91

* Standard Deviation (SD), Minimum value (Min), and Maximum value (Max).

characteristics. Cyclist-vehicle crashes are aggregated at the TAZ level according to their geospatial information. Two exposure measures are incorporated: Bike Kilometers Travelled (BKT) and Vehicle Kilometers Travelled (VKT). BKT is obtained by employing the Vancouver Cycling Data Model (VCDM). VCDM provided the cyclists' trips on the city of Vancouver bike network links (El Esawey et al., 2015). To obtain each bike link BKT, trip counts at each segment is multiplied by the corresponding bike link length, then aggregated for each TAZ. Finally, zonal characteristic variables were further divided into bike network, street network, land use, and demographics.

3.3.1. Bike network

Bike network variables are divided into five indicators: centrality, complexity, connectivity, directness, topography, and miscellaneous indicators. To calculate network indicators, the bike network is characterized as a set of links and nodes. The links represent the bike network infrastructure, while the nodes represent the intersections between network links (street and bike network links). ArcGIS software was employed to divide the bike network into TAZ's according to their location. Links that go through multiple zones were split on the zone boundary.

3.3.1.1. Centrality. A network with high centrality indicates low interconnectivity and accessibility. Centrality measurement includes degree and betweenness centralities (Porta et al., 2006; Jiang, 2009; Zhang et al., 2011). Degree centrality measures to what extent a node is connected directly to other nodes (Freeman, 1978). Degree centrality is calculated using Eq. (11), where $a_{ij} = 1$ only if node i and node j are connected by a link or more, and is equal to zero otherwise, and n is the number of nodes in a network.

$$C_i^D = \frac{1}{(n-1)} \sum_{j=1}^{n} a_{ij} \tag{11}$$

The Betweenness centrality is great if the node is navigated by many of the shortest paths linking each two-node (Freeman, 1978). Betweenness centrality is calculated using Eq. (12), where $g_{jk(i)}$ represents the number of node pairs j and k that contain point i on the shortest path connecting them and g_{jk} represents the number of node pairs j and k.

$$C_i^B = \frac{1}{(n-1)(n-2)} \sum_{j}^{n} \sum_{k}^{n} \frac{g_{jk(i)}}{g_{jk}}\, i \neq j \neq k \tag{12}$$

Eq. (13) is used to aggregate degree and betweenness centralities to the zone, where C_i^X is the centrality of zone i and $C_{i^*}^X$ is the largest possible value of C_i^X for the zone.

$$C^X = \frac{\sum_{i=1}^{n} [C_{i^*}^X - C_i^X]}{\max\sum_{i=1}^{n} [C_{i^*}^X - C_i^X]} \tag{13}$$

3.3.1.2. Complexity. Rodrigue et al. (2013) introduced two network metrics to quantify network complexity: Pi index and complexity. The pi index is the ratio of network diameter to network length. Network diameter is the length of the shortest path between the most distanced nodes of a zone network and network length is the total length of links at each zone. Complexity is the number of independent cycles, calculated using Eq. (14), where e is the number of links in a network, n is the number of nodes in a network, and P is the number of sub-graphs. Sub-graph is a subset of a graph that is isolated from the other subsets (there are no links between each subset).

$$Complexity = e - n + P \tag{14}$$

3.3.1.3. Connectivity. Three measures of connectivity are used: network density, network coverage, and intersection density. Network density is the ratio of the total length of bike links in a TAZ to the TAZ area (Berrigan et al., 2010; Zhang et al., 2012; Schoner and Levinson, 2014; Osama and Sayed, 2016; Saha et al., 2018). The degree of network coverage is calculated as the ratio of the number of bike links to the number of street network links in the TAZ (Osama and Sayed, 2016; Yigitcanlar and Dur, 2010). Intersection density is the ratio of the number of intersections to the area of the TAZ, including all intersections among bike network links as well as between bike network links and street network links (Cervero and Kockelman, 1997; Osama and Sayed, 2016; Reilly and Landis, 2003; Zhang et al., 2012).

3.3.1.4. Directness. Two measures are used to assess network directness: linearity and average link length. To measure linearity a hypothetical length (modified bike network length) is calculated that represents the total length of all bike network links if all the links were straight while preserving the node location and topology. Then linearity is the ratio of the modified bike network length to the actual total length of the bike network links within each TAZ (Osama and Sayed, 2016). The average link length is calculated as the total length of bike network links in a TAZ divided by the number of links (Kansky, 1963).

3.3.1.5. Miscellaneous. The total length of the bike network is calculated by summing up bike network links, regardless of their types, i.e. off-street, on-street, etc., within each TAZ. Bike Links with on-street parking proportion was calculated as a ratio of the total bike links length at each TAZ. The off-street bike facilities proportion to the bike network was calculated at the TAZ level.

3.3.1.6. Topography. The average weighted slope of the zonal bike network is calculated as the ratio between the aggregate lengths weighted slope and the length of the network at each TAZ as shown in Eq. (15).

$$WSlope = \frac{\sum_1^n (l_1 * s_1 + l_2 * s_2 + \ldots + l_n * s_n)}{\sum_1^n (l_1 + l_2 + \ldots + l_n)}, \tag{15}$$

where l represent the link length, s represent the link's slope, n represent the number of links at each TAZ. The total length of the network is calculated by summing up all the bike network links at the TAZ level. The average weighted slope of the network in each TAZ is calculated according to the following steps. First, the absolute grades along each link are averaged to compute the average slope of each link, using the contour map of the city of Vancouver. Afterward, the slope at each link is given a weight relative to its length. Finally, the average weighted slope of the links (in %) is calculated for each TAZ as shown in Eq. (15).

3.3.2. Street network

Street network variables included traffic signal density and bus stop density at each TAZ. Additionally, arterial, collector, and local street length proportions to the street network were calculated at each TAZ.

3.3.3. Land use

In this study, the City of Vancouver zoning data is used to produce the land use data. It should be noted, however, that the City of Vancouver zoning data, which represents the city permission for a certain area, is not the exact same as land use, which represents the actual type of activity running in each area. Five categories of land use densities were calculated: residential, recreational, industrial, institutional, and commercial. Other land use categories were discarded namely; agriculture, cemetery, and undeveloped. Additionally, different types of residential areas such as high rise apartments and low rise apartments were treated the same and aggregated under residential areas. The land use mix was calculated using Eq. (16), where θ_j is the land use category proportions for each TAZ (Frank et al., 2006).

$$Land\ Use\ Mix_i\ (LUM_i) = -\sum \frac{\theta_j * \ln(\theta_j)}{\ln(5)} \tag{16}$$

3.3.4. Demographics

Demographic variables included employment, household, and population densities at each TAZ. These variables were provided at each TAZ by Translink.

4. Results

This section presents the developed indices. The reported indices in the following section are the scaled indices unless otherwise mentioned.

4.1. Bike Attractiveness Index (BAI)

Table 3 shows the estimated Bike Kilometers Travelled (BKT)

Table 3
Bike Kilometers Travelled (BKT) Model.

	BKT Model					BKT Model Standardized Variables			
	Estimate	SD[§]	P-value	Confidence Interval		Estimate	SD[§]	Confidence Interval	
				2.50 %	97.50 %			2.50 %	97.50 %
Intercept	−1.131	0.468	0.017	−2.057	−0.205	0.000	0.076	−0.15	0.150
Land Use Mix	2.410	0.689	< 0.001	1.047	3.773	0.272	0.078	0.118	0.427
Recreation Density	1.268*	0.678	0.064	−0.073	2.608	0.146*	0.078	−0.008	0.300
Bike Network Density	0.063	0.024	0.011	0.015	0.111	0.202	0.078	0.046	0.357
Degree Centrality	−10.693	2.817	< 0.001	−16.267	−5.119	−0.293	0.077	−0.446	−0.140
Bike Network Slope	−0.229	0.102	0.027	−0.431	−0.027	−0.174	0.078	−0.328	−0.020

* Significantly different from zero at 10 %; all other variables were significantly different from zero at 5 %.
§ Standard Deviation (SD).

model, which includes five variables (five-fold cross-validation RMSE of 1.07). Standardized values are reported to evaluate the magnitude of each variable's influence on BKT. All estimates are significant at 10 % or lower.

Based on the standardized model estimates in Table 3, land use mix and network centrality have the highest influence on BKT, while recreational density and bike network slope have the lowest influence. The raw Bike Attractiveness Index (BAI) model derived from the BKT regression is given in Eq. (17).

$$\begin{aligned} BAI' = &-1.131 + 2.410{*}\mathit{Land\ use\ Mix} + 1.268{*}\mathit{Recreational\ Density} \\ &+ 0.063{*}\mathit{Bike\ Network\ Density} - 10.693{*}\mathit{Degree\ Centrality} \\ &- 0.229{*}\text{Bike Network}\ \mathit{Slope} \end{aligned} \tag{17}$$

Fig. 1 illustrates the Bike Attractive Index (BAI) by TAZ as well as its elements: land use mix, network density, bike network average weighted slope, and degree centrality. The TAZ colors in Fig. 1 use green for values that correspond to higher attractiveness and red for values that correspond to lower attractiveness. The color breaks are the indicators five quintiles. The BAI heat map shows the TAZ number on each TAZ. Fig. 1 shows how land use mix, network density, bike network average weighted slope, and degree centrality contributes to the BAI. Fig. 1 shows that there is a spatial correlation between Zones' BAI. Also, it is noted that the BAI is low in some of the downtown zones (zones from 1 to 34), where it is expected to be high as many biking trips may be conducted in the downtown.

4.2. Bike Safety Index (BSI)

Table 4 gives the developed crash prediction model, which includes seven variables (AIC = 841.68, five-fold cross-validation RMSE of 9.10).

Bike and vehicle exposure are positively associated with cyclist-vehicle crashes as expected and similar to many previous studies (Miranda-Moreno et al., 2011; Strauss et al., 2013; Hamann and Peek-Asa, 2013; Kaplan and Prato, 2015). Based on the standardized model estimates, signal density has the highest influence on cyclist-vehicle crashes while the network coverage has the lowest influence. The raw Bike Safety Index (BSI) derived from the crash prediction model is given in Eq. (18).

$$\begin{aligned} BSI' = &-2.829 - 0.187{*}\ln(VKT) - 0.016{*}\mathit{Signal\ Density} \\ &+ 1.409{*}\mathit{Recreational\ Density} - 0.113{*}\mathit{Complexity} \\ &+ 3.000{*}\mathit{Average\ Link\ Length} + 0.729{*}\mathit{Bike\ Network\ Coverage} \end{aligned} \tag{18}$$

Fig. 2 illustrates the Bike safety Index (BSI) by TAZ as well as its elements: signal density, recreational density, complexity, average link length, vehicle kilometers traveled, and network coverage. Similar to Fig. 1, the TAZ colors in Fig. 2 use green for values that correspond to higher safety for cyclists and red for values that correspond to lower safety for cyclists. The color breaks are the indicators five quintiles. The BSI heat map shows the TAZ number on each TAZ. Fig. 2 shows how signal density, recreational density, network complexity, average link length, vehicle kilometers traveled, and network coverage contribute to the BSI. The downtown suffers from low BSI due to its high signal

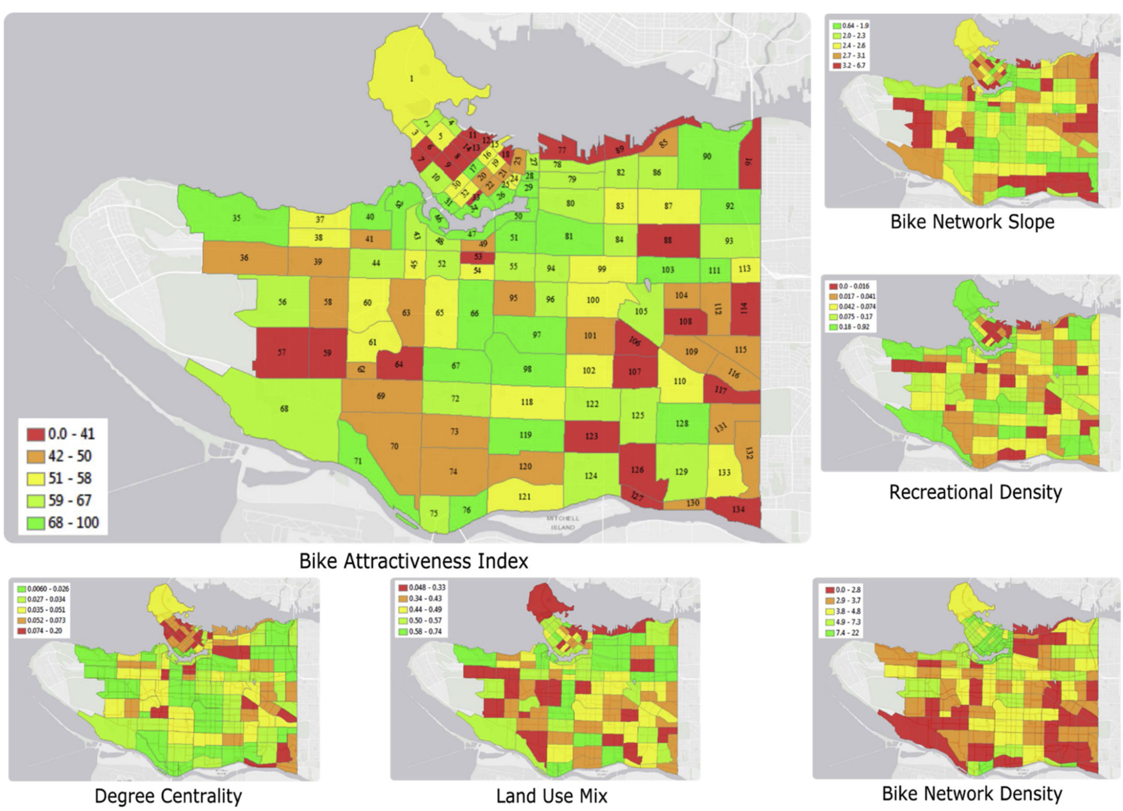

Fig. 1. Bike Attractiveness Index and its components.

Table 4
Cyclist-Vehicle Crash Model.

	Crash Model					Crash Model Standardized Variables			
	Estimate	SD	P-value	Confidence Interval		Estimate	SD	Confidence Interval	
				2.50 %	97.50 %			2.50 %	97.50 %
Intercept	2.829	0.276	< 0.001	2.252	3.411	2.216	0.056	2.107	2.326
VKT	0.187	0.086	0.03	0.012	0.363	0.161	0.074	0.010	0.312
BKT	0.614	0.057	< 0.001	0.501	0.730	0.727	0.068	0.593	0.864
Signal Density	0.016	0.004	< 0.001	0.009	0.023	0.292	0.065	0.159	0.431
Recreational Density	−1.409	0.557	0.011	−2.534	−0.28	−0.192	0.076	−0.345	−0.038
Complexity	0.113	0.041	0.006	0.029	0.198	0.159	0.058	0.041	0.279
Average Link Length	−3.001	1.319	0.023	−5.673	−0.362	−0.155	0.068	−0.293	−0.019
Bike Network Coverage	−0.729*	0.397	0.067	−1.500	0.047	−0.136*	0.074	−0.28	0.009

*Significantly different from zero at 10 %, all other variables were significantly different from zero at 5 %.
§Standard Deviation (SD).

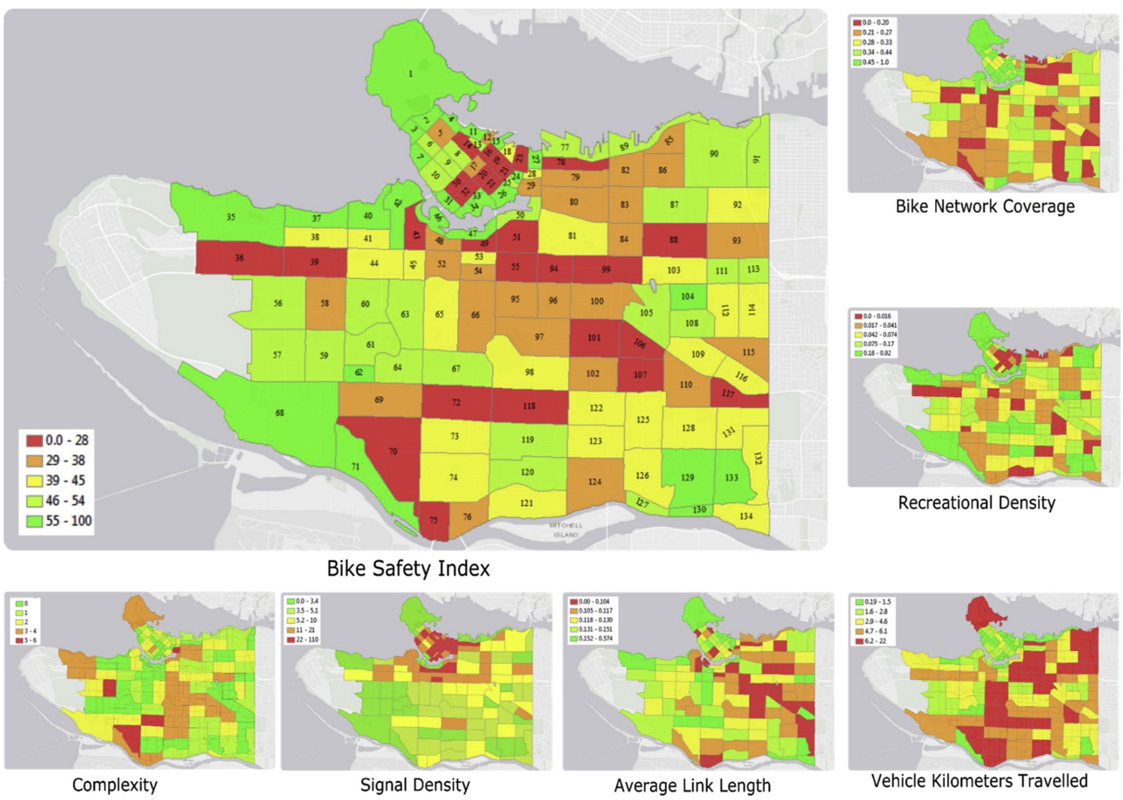

Fig. 2. Bike Safety Index and its components.

Table 5
A sample of the highest and lowest Bike Safety Index (BSI).

TAZ	Bike Safety Index (Ranking)	VKT	Signal Density	Recreational Density	Complexity	Average Link Length	Bike Network Coverage
			Safest Zones				
1	100 (1)	13.64	0.00	0.92	4	0.33	0.61
47	94.7 (2)	0.44	0.00	0.36	0	0.13	0.82
3	85.1(3)	0.19	0.00	0.27	0	0.09	0.55
133	83.9(4)	1.41	0.92	0.53	0	0.17	0.19
			Least Safe Zones				
19	15.3 (132)	0.97	75.37	0.02	2	0.13	0.58
4	14.9 (132)	3.29	54.60	0.00	1	0.13	0.31
16	4.7 (133)	1.35	110.55	0.00	1	0.16	0.79
30	0 (134)	2.65	78.50	0.00	3	0.09	0.66

Zones 1, 47, 3, and 133 are highly ranked on the BSI mainly because of their high recreational area density, and low signal density. On the other hand, Zones 30, 16, 14, and 19 were ranked the least safe zones for cyclists in the City of Vancouver. This is mainly due to their high signal density and almost zero recreational density.

density and low recreational densities. Fig. 2 shows a spatial auto-correlation in the BSI and its 9four safest zones and the four least safe zones according to the Bike Safety Index (BSI) are shown in Table 5, along with the BSI constitutes.

4.3. Bike Composite Index (BCI)

The Pearson correlation between BAI and BSI is low (r = 0.11), which highlights the need for a composite index including attractiveness and safety. Fig. 3 illustrates the relationship between the BAI and BSI, along with BCI and TAZ area. The dispersion (and low Pearson correlation between BAI and BSI) supports the importance of including both safety and attractiveness dimensions in the composite index.

The Principle component analysis loading for the BAI and BSI are 0.73 and 0.68, respectively. The first component explained 56 % of the variability in BAI and BSI, which is reasonable given the low Pearson

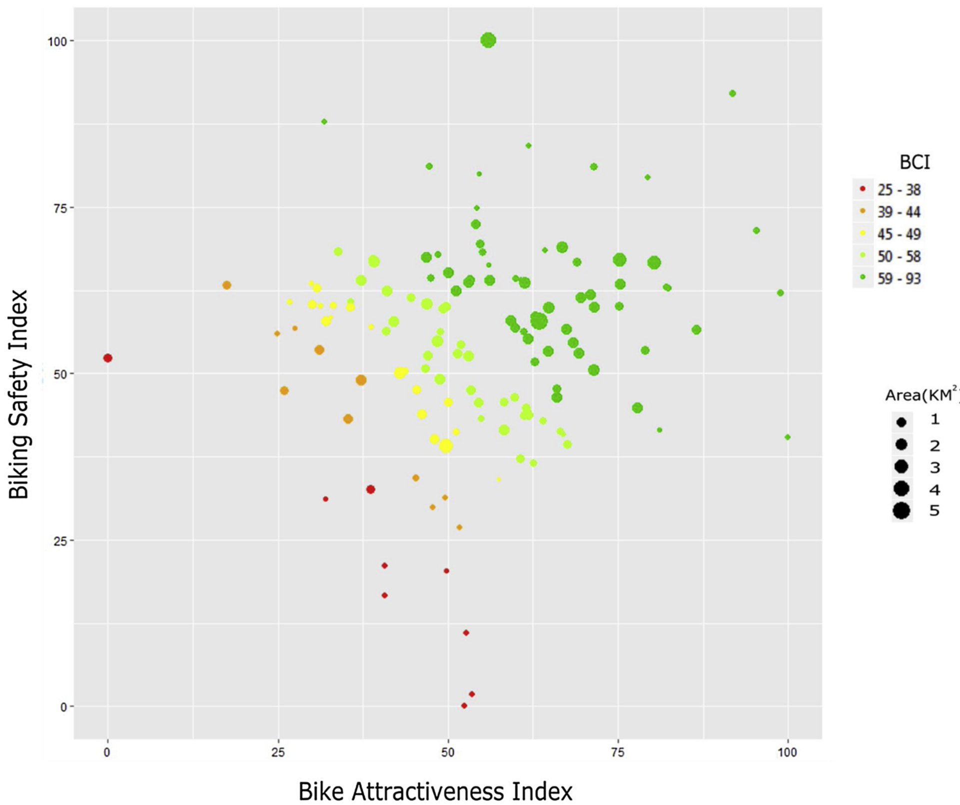

Fig. 3. Bike Attractiveness Index Vs. Bike Safety Index, with BCI and TAZ area as the point size and color.

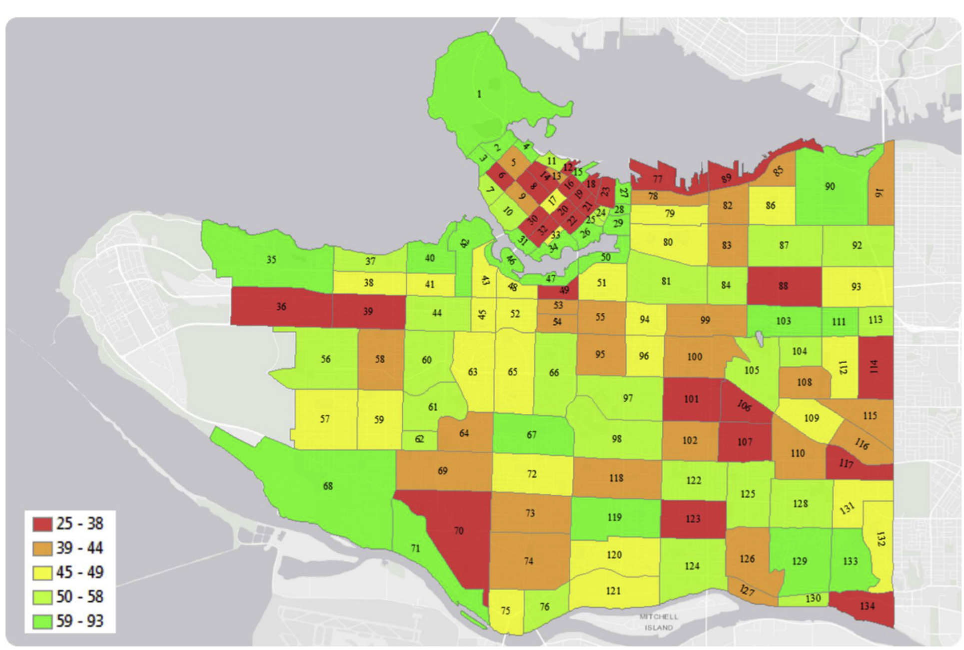

Fig. 4. Bike Composite Index.

correlation between BAI and BSI (r = 0.11). Since the loadings are almost equal, the Bike Composite Index (BCI) is aggregated in equal weighting, as given in Eq. (19).

$$BCI_i = 0.5 * BAI''_i + 0.5 * BSI''_i \tag{19}$$

Fig. 4 illustrates the Bike Composite Index (BCI) values by TAZ. The color breaks are the Bike Composite Index five quintiles. Fig. 4 shows the TAZ number on each TAZ. Since the BCI is in equal weighting format, and Vancouver's downtown suffers a low biking attractiveness (BAI) and cyclist-vehicle crash risk (BSI), the BCI is low in most of the downtown zones. It should be noted that the results do not differ significantly from the 0.73 and 0.68 weightings, compared to equal weighting format.

5. Discussion

5.1. Bike Attractiveness Index (BAI)

The bike attractiveness model estimates show that the land use mix is positively associated with BKT. Similarly, Dill and Voros (2007) found that residents living in mixed land use environments have a higher probability to commute by bike. Recreational density is positively associated with BKT. This result is reasonable, as Daley and Rissel (2011) showed that recreational density encourages road users to conduct more biking trips. The average weighted slope is negatively associated with BKT, this is intuitive and consistent with previous studies (Hood et al., 2011; Osama et al., 2017; Winters et al., 2016), as steeper slopes work as a deterrent for cyclists. Network density is positively associated with BKT, in line with previous studies in which network connectivity and density were positively associated with biking levels (Berrigan et al., 2010; Marshall and Garrick, 2010; Osama et al., 2017; Schoner and Levinson, 2014). Degree centrality is negatively associated with BKT. This is consistent with Marshall and Garrick (2010) who found that gridded street networks (low centrality networks) were associated with higher walking and biking.

It is noted that the BAI is low in some of the downtown zones (zones from 1 to 34), where it is expected to be high as many biking trips may be conducted in the downtown. This is because the BAI objective is to capture attractiveness to biking, not the probability of biking, without considering the density of activities origin and destinations. Fig. 1 shows a spatial autocorrelation in the BAI and its components, therefore BKT model was redeveloped after accounting for the spatial autocorrelation, using the Conditional Autoregressive (CAR) technique. The newly developed CAR model estimates' were all significant at 10 % or lower, and have similar estimates, compared to the reported model in Table 3. As a result, to enhance the BAI transferability, and to develop a parsimonious model the spatial model is not employed to develop the BAI.

5.2. Bike Safety Index (BSI)

Table 4 shows that signal density is positively associated with cyclist-vehicle crashes which is also consistent with previous studies (Strauss et al., 2013; Wei and Lovegrove, 2013). Higher traffic signal density implies the existence of wider intersections that increase the crash risk for cyclists. Higher recreational density is found associated with a decrease in cyclist-vehicle crashes. This is plausible because these areas typically provide off-street bike lanes for cyclists and therefore it reduces the potential conflict between cyclists and motorists. Bike network complexity has a positive association with cyclist-vehicle crashes. The positive association can be attributed to that a complex network configuration creates more potential conflict points between motorists and cyclists. The average link length has a negative association with cyclist-vehicle crashes. The results imply that longer links without discontinuities (fewer intersections) are safer for cyclists which is consistent with Osama and Sayed (2016). Bike network coverage has a negative association with cyclist-vehicle crashes; this is intuitive as bike lanes are safer for cyclists than street facilities without bike lanes.

The downtown suffers from low BSI due to its high signal density and low recreational densities, as shown in Fig. 2. Fig. 2 shows a spatial autocorrelation in the BSI and its constituents, therefore the cyclist-vehicle crash model is redeveloped after accounting for the spatial autocorrelation, using Conditional Autoregressive (CAR) technique. The newly developed CAR spatial model estimates' were significant at 10 % or lower excluding the bike network coverage and have similar estimates compared to the reported model in Table 4. As a result, to enhance the BSI transferability, and to develop a parsimonious model the spatial model is not employed to develop the BSI.

5.3. Bike Composite Index (BCI)

The fact that the annual average daily bike traffic (AADB) includes only the bike network links should increase the correlation between the BAI and BSI. This is because bike traffic volume might be underestimated in zones with high local street proportions by the BKT and subsequently the BAI. Furthermore, this would result in an underestimation of the cyclist-vehicle crash risk (BSI). As the AADB limitations cause underestimation to both BAI and BSI, this limitation may not weaken the correlation between the BAI and BSI.

5.4. Comparison to the Bike Score®

The Bike Score® is selected to be compared to the study's developed indices as it is published for the City of Vancouver. The Bike Score® was developed based on expert advice from researchers at the University of British Columbia as well as votes from the Walk Score community (Bike Score Methodology [WWW Document], 2019).

Fig. 5 shows the aggregated TAZ Bike Score® values and the base Bike Score® map for the City of Vancouver. The Bike Score® was developed to measures whether a location is suitable for biking and is calculated from equal weightings of bike lane score, hill score, destinations and street connectivity, and bike commuting mode share (Bike Score Methodology [WWW Document], 2019). Bike lane score includes all on-street and off-street bike lanes/paths extracted from OpenStreetMap, where a decay function is used to each segment (no value is given to segments further than 1,000 m from the studied location). Hill score employs the National Elevation Data set to search for the steepest grade within a 200-meter radius of the studied location, where a grade of 10 %–2 % is given a score of 0 - 100. Destinations and street connectivity measure the network distances to a set of amenities then calculates connectivity measures (e.g. intersection density). Bike commuting mode share from the Census data was added, by creating a 1 km moving window over the census tract level data and normalize bicycle mode share.

Thirty Bike Score® values per zone for the City of Vancouver were retrieved from www.walkscore.com manually (by entering each location address). Then Bike Score® is aggregated for each TAZ using the maximum score, the minimum score, and the mean of five evenly dispersed points' scores in the TAZ. The average standard deviation within TAZ for the City of Vancouver is 6.95. As presented in Fig. 5, high spatial autocorrelation in the scores is evident, due to the use of a 1 km decay function over bike lane score, 200 m buffer over hill Score, and 1 km moving window over the Bike commuting mode share.

As the Bike Score® is similar to the BAI in definition, therefore it is compared mainly to the BAI. There is a 0.19 Pearson correlation between BAI and the zonal Bike Score®. Additionally, Cohen's kappa is a -0.0001, which shows poor interrater reliability. As the two scores (BAI, Bike Score®) are poorly correlated and have poor interrater reliability, the difference in rankings between the two scores is used to compare the two scores. Furthermore, Bike Score® is weakly correlated with the

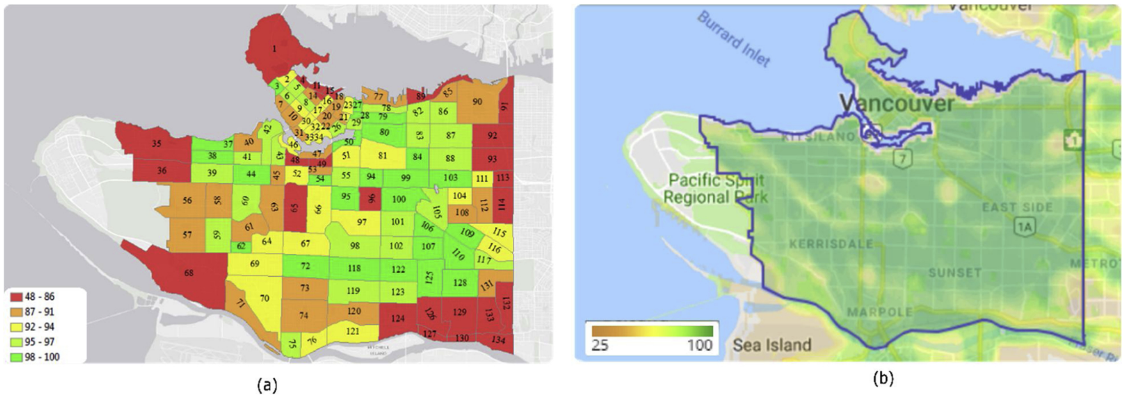

Fig. 5. (a) aggregated Bike Score® by TAZ and (b) base Bike Score®.

Table 6
Zones with the largest differences between Bike Attractiveness Index and Bike Score® rankings.

TAZ	Bike Score® (Ranking)	Bike Attractiveness Index (Ranking)	Land Use Mix	Recreational Density	Bike Network density	Degree Centrality	Bike Network Slope
			Greatest positive differences zones				
35	72.6 (127)	80.4 (9)	0.63	0.29	3.65	0.025	2.08
71	86.2 (106)	77.8 (12)	0.57	0.56	0.55	0.022	2.55
92	85 (112)	70.9 (19)	0.59	0.16	5.21	0.022	2.95
			Lowest negative differences zones				
95	99.8 (3)	45.3 (99)	0.26	0.00	3.36	0.031	1.68
107	97.8 (22)	35.4 (117)	0.32	0.02	4.00	0.068	2.44
123	97.2 (28)	32.1 (121)	0.26	0.00	2.94	0.054	2.54

Bike Safety Index (r = -0.31) and Bike Composite Index (r = −0.07). Table 6 presents the three zones with the greatest positive differences in rankings between the BAI and Bike Score® and the three zones with the lowest negative differences in rankings between the BAI and Bike Score®, and their BAI constitutes values.

BAI gives high rankings to zones that have recreational areas, whereas Bike Score® does not. This is due to two reasons: first, the Bike score® does not include recreational areas in its calculation directly, while the BAI accounts for zones' recreational density. Second, the BAI accounts for the bike links slope instead of the 200 m buffer used in the hill score component of the Bike Score. Therefore, the Bike Score® gives low ranking to zones that have recreational areas, as recreational areas specifically parks, such as Stanley Park (zone 1), usually have steeper slopes than other land use categories.

Moreover, the Bike Score® extract bike lane data from OpenStreetMap (OSM). The process of crowdsourcing geographic data accuracy might be questionable as the OSM data is not administered by a public agency. Whereas, the current study employed a bike network provided by the City of Vancouver. The different sources of bike network data may add discrepancies to the developed scores.

Bike Score® accounted for destinations that might attract cyclists. In contrast, the BAI valued two zoning characteristics, namely recreational facilities presence and land use diversity. Previous studies have found mixed land use to be correlated with biking levels (Dill and Voros, 2007; Jones et al., 2010).

5.5. Study limitations and areas of future research

There are several limitations of this study and areas for future research. This study assumes that the zonal BKT represents attractiveness for biking. Also, this study considered only cyclist-vehicle crashes due to data limitations. Investigating other types of crashes such as cyclist-cyclist or cyclist-pedestrian crashes would be beneficial. Excluding cyclist-cyclist and/or cyclist-pedestrian crashes may overestimate the Bike Safety Index for zones with high conflict points between pedestrian and cyclists (e.g. recreational areas). Neglecting single-cyclist crashes (falls) is a limitation for crash prediction, as they can cause a significant portion of injuries (Schepers et al., 2015). However, because the BSI is normalized, it is only sensitive to this limitation if the single-cyclist, cyclist-cyclist, or cyclist-pedestrian crashes are spatially clustered. This also could lead to an overestimate of BSI in recreational areas, if they are poorly lit and/or heavily vegetated.

For future research, the elevation data could be updated using high-resolution LIDAR (El Masri and Bigazzi, 2019) or crowd-sourced data (McKenzie and Janowicz, 2017). The transferability of the developed indices should be tested in other cities. The data in this study are cross-sectional tracking biking levels over time and cyclist crashes could capture the causal relationships among zonal characteristics, biking levels, and cyclist safety. As traffic analysis zones change over time, future research should investigate other units of analysis (e.g. 10 m grid cells). Furthermore, the study methodology could be applied to develop statistically calibrated indices that employ data from open street maps, which might not be as accurate as the models developed in this study (with locally collected data). However, the model would be more easily applied to different cities.

6. Conclusion

Previous bikeability zonal indices did not account for the zonal cyclist-vehicle crash risk. This study developed a Bike Composite Index (BCI) that consists of two sub-indices representing bike attractiveness and bike safety. The developed Bike Attractiveness Index (BAI) consists of bike network weighted slope, bike network centrality, bike network density, land use mix, and recreational density. The Bike Safety Index (BSI) consists of signal density, recreational density, and vehicle kilometers traveled as well as bike network coverage, average link length, and complexity. The Pearson correlation between the BAI and BSI is low, which highlights the need for a composite index including attractiveness and safety. One possible use of the developed indices is in decision-making for investments in cycling infrastructure that prioritize areas with high potential demand (i.e., the density of activities) but low biking infrastructure attractiveness (BAI) or safety (BSI).

CRediT authorship contribution statement

Mohamed Bayoumi Kamel: Conceptualization, Data curation, Visualization, Methodology, Formal analysis, Writing - original draft. **Tarek Sayed:** Conceptualization, Validation, Methodology, Writing - review & editing, Supervision. **Alexander Bigazzi:** Validation, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aap.2020.105439.

References

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2016. Macroscopic modeling of pedestrian and bicycle crashes: a cross-comparison of estimation methods. Accid. Anal. Prev. 93, 147–159. https://doi.org/10.1016/j.aap.2016.05.001.

Banister, D., 2008. The sustainable mobility paradigm. Transp. Policy 15, 73–80. https://doi.org/10.1016/j.tranpol.2007.10.005.

Baxter, M.J., 1995. Standardization and transformation in principal component analysis, with applications to archaeometry. J. R. Stat. Soc. Ser. C Appl. Stat. 44, 513–527.

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

Bike Score Methodology [WWW Document], 2019. Walk Score. URL https://www.walkscore.com/bike-score-methodology.shtml (Accessed 4.29.19). .

Carr, L.J., Dunsiger, S.I., Marcus, B.H., 2010. Walk score™ as a global estimate of neighborhood walkability. Am. J. Prev. Med. 39, 460–463. https://doi.org/10.1016/j.amepre.2010.07.007.

Caulfield, B., Brick, E., McCarthy, O.T., 2012. Determining bicycle infrastructure preferences – a case study of Dublin. Transp. Res. Part Transp. Environ. 17, 413–417. https://doi.org/10.1016/j.trd.2012.04.001.

Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: density, diversity, and design. Transp. Res. Part Transp. Environ. 2, 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6.

Cervero, R., Denman, S., Jin, Y., 2019. Network design, built and natural environments, and bicycle commuting: evidence from British cities and towns. Transp. Policy 74, 153–164. https://doi.org/10.1016/j.tranpol.2018.09.007.

Chen, P., 2015. Built environment factors in explaining the automobile-involved bicycle crash frequencies: a spatial statistic approach. Saf. Sci. 79, 336–343. https://doi.org/10.1016/j.ssci.2015.06.016.

Cheng, W., Lin, W.H., Jia, X., Wu, X., Zhou, J., 2018. Ranking cities for safety investigation by potential for safety improvement. J. Transp. Saf. Secur. 10, 345–366. https://doi.org/10.1080/19439962.2017.1279250.

Daley, M., Rissel, C., 2011. Perspectives and images of cycling as a barrier or facilitator of cycling. Transp. Policy 18, 211–216. https://doi.org/10.1016/j.tranpol.2010.08.004.

Deacon, J.A., Zegeer, C.V., Deen, R.C., 1974. Identification of Hazardous Rural Highway Locations.

Dill, J., Voros, K., 2007. Factors affecting bicycling demand: initial survey findings from the Portland, Oregon. Region. Transp. Res. Rec. J. Transp. Res. Board 2031, 9–17. https://doi.org/10.3141/2031-02.

El Esawey, M., Lim, C., Sayed, T., 2015. Development of a cycling data model: city of Vancouver case study. Am. J. Civ. Eng. 42, 1000–1010. https://doi.org/10.1139/cjce-2015-0065.

El Masri, O., Bigazzi, A.Y., 2019. Road grade estimates for bicycle travel analysis on a street network. Transp. Res. Part C Emerg. Technol. 104, 158–171. https://doi.org/10.1016/j.trc.2019.05.004.

Elvik, R., 2007. State-of-the-art approaches to road accident black spot management and safety analysis of road networks. Transportøkonomisk institutt Oslo.

Frank, L.D., Sallis, J.F., Conway, T.L., Chapman, J.E., Saelens, B.E., Bachman, W., 2006. Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality. J. Am. Plann. Assoc. 72, 75–87.

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. Soc. Netw. 1, 215–239. https://doi.org/10.1016/0378-8733(78)90021-7.

Giles-Corti, B., Foster, S., Shilton, T., Falconer, R., 2010. The co-benefits for health of investing in active transportation. New South Wales Public Health Bull. 21, 122–127. https://doi.org/10.1071/NB10027.

Hamann, C., Peek-Asa, C., 2013. On-road bicycle facilities and bicycle crashes in Iowa, 2007–2010. Accid. Anal. Prev. 56, 103–109. https://doi.org/10.1016/j.aap.2012.12.031.

Handy, S.L., Xing, Y., 2011. Factors correlated with bicycle commuting: a study in six small U.S. cities. Int. J. Sustain. Transp. 5, 91–110. https://doi.org/10.1080/15568310903514789.

Harkey, D.L., Reinfurt, D.W., Knuiman, M.W., Stewart, J.R., 1998. Development of the Bicycle Compatibility Index: a Level of Service Concept. Final Report. .

Hauer, E., 1997. Observational before/after Studies in Road Safety. Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety.

Hood, J., Sall, E., Charlton, B., 2011. A GPS-based bicycle route choice model for San Francisco, California. Transp. Lett. 3, 63–75. https://doi.org/10.3328/TL.2011.03.01.63-75.

Jiang, B., 2009. Street hierarchies: a minority of streets account for a majority of traffic flow. Int. J. Geogr. Inf. Sci. 23, 1033–1048.

Jones, M.G., Ryan, S., Donlon, J., Ledbetter, L., Ragland, D.R., Arnold, L.S., 2010. Seamless Travel: Measuring Bicycle and Pedestrian Activity in San Diego County and Its Relationship to Land Use, Transportation, Safety, and Facility Type.

Kamel, M.B., Sayed, T., Osama, A., 2019. Accounting for mediation in cyclist-vehicle crash models: a Bayesian mediation analysis approach. Accid. Anal. Prev. 131, 122–130. https://doi.org/10.1016/j.aap.2019.06.009.

Kansky, K.J., 1963. Structure of Transportation Networks: Relationships between Network Geometry and Regional Characteristics.

Kaplan, S., Prato, C.G., 2015. A spatial analysis of land use and network effects on frequency and severity of cyclist–motorist crashes in the Copenhagen region. Traffic Inj. Prev. 16, 724–731. https://doi.org/10.1080/15389588.2014.1003818.

Landis, B., Vattikuti, V., Brannick, M., 1997. Real-time human perceptions: toward a bicycle level of service. Transp. Res. Rec. J. Transp. Res. Board 119–126.

Larsen, J., Patterson, Z., El-Geneidy, A., 2013. Build it. But where? The use of geographic information systems in identifying locations for new cycling infrastructure. Int. J. Sustain. Transp. 7, 299–317. https://doi.org/10.1080/15568318.2011.631098.

Lin, J.-J., Wei, Y.-H., 2018. Assessing area-wide bikeability: a grey analytic network process. Transp. Res. Part Policy Pract. 113, 381–396. https://doi.org/10.1016/j.tra.2018.04.022.

Markt, D., 2001. Internal Market Scoreboard.

Marshall, W.E., Garrick, N.W., 2010. Effect of street network design on walking and biking. Transp. Res. Rec. J. Transp. Res. Board 2198, 103–115. https://doi.org/10.3141/2198-12.

McKenzie, G., Janowicz, K., 2017. ISED: constructing a high-resolution elevation road dataset from massive, low-quality in-situ observations derived from geosocial fitness tracking data. PLoS One 12, e0186474. https://doi.org/10.1371/journal.pone.0186474.

Mekuria, M.C., Furth, P.G., Nixon, H., 2012. Low-stress Bicycling and Network Connectivity.

Miranda-Moreno, L.F., Fu, L., 2007. Traffic Safety Study: Empirical Bayes or Full Bayes?.

Miranda-Moreno, L.F., Strauss, J., Morency, P., 2011. Disaggregate exposure measures and injury frequency models of cyclist safety at signalized intersections. Transp. Res. Rec. 2236, 74–82. https://doi.org/10.3141/2236-09.

Moudon, A.V., Lee, C., Cheadle, A.D., Collier, C.W., Johnson, D., Schmid, T.L., Weather, R.D., 2005. Cycling and the built environment, a US perspective. Transp. Res. Part Transp. Environ. 10, 245–261. https://doi.org/10.1016/j.trd.2005.04.001.

Narayanamoorthy, S., Paleti, R., Bhat, C.R., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transp. Res. Part B Methodol. 55, 245–264. https://doi.org/10.1016/j.trb.2013.07.004.

Osama, A., Sayed, T., 2016. Evaluating the impact of bike network indicators on cyclist safety using macro-level collision prediction models. Accid. Anal. Prev. 97, 28–37. https://doi.org/10.1016/j.aap.2016.08.010.

Osama, A., Sayed, T., 2019. A novel approach for identifying, diagnosing, and treating active transportation safety issues. Transp. Res. Rec., 0361198119847614. https://doi.org/10.1177/0361198119847614.

Osama, A., Sayed, T., Bigazzi, A.Y., 2017. Models for estimating zone-level bike kilometers traveled using bike network, land use, and road facility variables. Transp. Res. Part Policy Pract. 96, 14–28. https://doi.org/10.1016/j.tra.2016.11.016.

Porta, S., Crucitti, P., Latora, V., 2006. The network analysis of urban streets: a primal approach. Environ. Plan. B Plan. Des. 33, 705–725. https://doi.org/10.1068/b32045.

Porter, M.E., Stern, S., 1999. The New Challenge to America's Prosperity: Findings from The Innovation Index. Council on Competitiveness Publ. Off, Washington DC.

Pucher, J., Buehler, R., 2010. Walking and cycling for healthy cities. Built Environ. 36, 391–414. https://doi.org/10.2148/benv.36.4.391.

Reilly, M., Landis, J., 2003. The Influence of Built-form and Land Use on Mode Choice.

Reynolds, C.C., Harris, M.A., Teschke, K., Cripton, P.A., Winters, M., 2009. The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. Environ. Health 8, 47. https://doi.org/10.1186/1476-069X-8-47.

Rodrigue, J.-P., Comtois, C., Slack, B., 2013. The Geography of Transport Systems, third edition. Routledge, London ; New York ed.

Safety of vulnerable road users (DSTI/DOT/RTR/RS7(98)1/FINAL No. 68074), 1998. Scientific Expert Group on the Safety of Vulnerable Road Users (RS7). Organisation for Economic Co-operation and Development, Paris.

Saha, D., Alluri, P., Gan, A., Wu, W., 2018. Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models. Accid. Anal. Prev. 118, 166–177. https://doi.org/10.1016/j.aap.2018.02.014.

Sayed, T., Rodriguez, F., 1999. Accident prediction models for urban unsignalized intersections in British Columbia. Transp. Res. Rec. 1665, 93–99.

Schepers, P., Agerholm, N., Amoros, E., Benington, R., Bjørnskau, T., Dhondt, S., Geus, Bde, Hagemeister, C., Loo, B.P.Y., Niska, A., 2015. An international review of the frequency of single-bicycle crashes (SBCs) and their relation to bicycle modal share. Inj. Prev. 21, e138–e143. https://doi.org/10.1136/injuryprev-2013-040964.

Schlüter, P.J., Deely, J.J., Nicholson, A.J., 1997. Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. J. R. Stat. Soc. Ser. A Stat. 46, 293–316.

Schneider, R.J., Stefanich, J., 2015. Neighborhood characteristics that support bicycle commuting: analysis of the top 100 US census tracts. Transp. Res. Rec. 2520, 41–51.

Schoner, J.E., Levinson, D.M., 2014. The missing link: bicycle infrastructure networks and ridership in 74 US cities. Transportation 41, 1187–1204. https://doi.org/10.1007/s11116-014-9538-1.

Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid. Anal. Prev. 45, 382–391. https://doi.org/10.1016/j.aap.2011.08.003.

Stokes, R.W., Mutabazi, M.I., 1996. Rate-quality control method of identifying hazardous road locations. Transp. Res. Rec. 1542, 44–48.

Strauss, J., Miranda-Moreno, L.F., Morency, P., 2013. Cyclist activity and injury risk analysis at signalized intersections: a Bayesian modelling approach. Accid. Anal. Prev. 59, 9–17. https://doi.org/10.1016/j.aap.2013.04.037.

Tamburri, T.N., Smith, R.N., 1970. The safety index: a method of evaluating and rating safety benefits. Highw. Res. Rec.

Teschke, K., Harris, M.A., Reynolds, C.C.O., Winters, M., Babul, S., Chipman, M., Cusimano, M.D., Brubacher, J.R., Hunte, G., Friedman, S.M., Monro, M., Shen, H., Vernich, L., Cripton, P.A., 2012. Route infrastructure and the risk of injuries to bicyclists: a case-crossover study. Am. J. Public Health 102, 2336–2343. https://doi.org/10.2105/AJPH.2012.300762.

Tilahun, N.Y., Levinson, D.M., Krizek, K.J., 2007. Trails, lanes, or traffic: valuing bicycle facilities with an adaptive stated preference survey. Transp. Res. Part Policy Pract. 41, 287–301. https://doi.org/10.1016/j.tra.2006.09.007.

van den bossche, F., WETS, G., LESAFFRE, E., 2002. A Bayesian Hierarchical Approach to Model the Rank of Hazardous Intersections for Bicyclists using the Gibbs Sampler.

Vancouver, C., 2013. Open Data Catalogue [WWW Document]. Of, URL https://vancouver.ca/your-government/open-data-catalogue.aspx (Accessed 4.27.11). .

Vandenbulcke, G., Thomas, I., Int Panis, L., 2014. Predicting cycling accident risk in Brussels: a spatial case–control approach. Accid. Anal. Prev. 62, 341–357. https://doi.org/10.1016/j.aap.2013.07.001.

Wei, F., Lovegrove, G., 2013. An empirical tool to evaluate the safety of cyclists: community based, macro-level collision prediction models using negative binomial regression. Accid. Anal. Prev. 61, 129–137. https://doi.org/10.1016/j.aap.2012.05.018.

Wilkinson, W.C., 1994. Selecting Roadway Design Treatments to Accommodate Bicycles.

Winters, M., Brauer, M., Setton, E.M., Teschke, K., 2013. Mapping bikeability: a spatial tool to support sustainable travel. Environ. Plan. B Plan. Des. 40, 865–883. https://doi.org/10.1068/b38185.

Winters, M., Teschke, K., Brauer, M., Fuller, D., 2016. Bike Score®: associations between urban bikeability and cycling behavior in 24 cities. Int. J. Behav. Nutr. Phys. Act. 13, 18.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis chemometrics and intelligent laboratory systems 2. IEEE Conference on Emerging Technologies & Factory Automation Efta Volume 704–706.

World Health Organization, 2009. Global Status Report on Road Safety: Time for Action. World Health Organization, Geneva.

Yang, B.Z., Loo, B.P.Y., 2016. Land use and traffic collisions: a link-attribute analysis using Empirical Bayes method. Accid. Anal. Prev. 95, 236–249. https://doi.org/10.1016/j.aap.2016.07.002.

Yigitcanlar, T., Dur, F., 2010. Developing a sustainability assessment model: the sustainable infrastructure, land-use, environment and transport model. Sustainability 2, 321–340.

Zhang, Y., Wang, X., Zeng, P., Chen, X., 2011. Centrality characteristics of road network patterns of traffic analysis zones. Transp. Res. Rec. J. Transp. Res. Board 16–24.

Zhang, Y., Bigham, J., Li, Z., Ragland, D., Chen, X., 2012. Associations between road network connectivity and pedestrian-bicyclist accidents. In: 91st Annual Meeting of the Transportation Research Board, Washington DC. Washington, DC.