
Faculty of Science

Faculty Publications

What does crowdsourced data tell us about bicycling injury? A case study in a mid-sized Canadian city

Jaimy Fischer, Trisalyn Nelson, Karen Laberee, Meghan Winters

July 2020

© 2020 Jaimy Fischer et al. This is an open access article distributed under the terms of the Creative Commons Attribution License. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

This article was originally published at:

<https://doi.org/10.1016/j.aap.2020.105695>

Citation for this paper:

Fischer, J., Nelson, T., Laberee, K., & Winters, M. (2020). What does crowdsourced data tell us about bicycling injury? A case study in a mid-sized Canadian city. *Accident Analysis & Prevention*, 145, 1-8. <https://doi.org/10.1016/j.aap.2020.105695>.



Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

What does crowdsourced data tell us about bicycling injury? A case study in a mid-sized Canadian city



Jaimy Fischer^a, Trisalyn Nelson^b, Karen Laberee^{c,*}, Meghan Winters^a

^a Faculty of Health Sciences, Simon Fraser University, Burnaby, V5A 1S6, Canada

^b School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, 85281, USA

^c Department of Geography, University of Victoria, 3800 Finnerty Road, Victoria, BC, V8P 5C2, Canada

ARTICLE INFO

Keywords:

Bicycling safety
Citizen science
Crowdsourced
Injury

ABSTRACT

With only ~20 % of bicycling crashes captured in official databases, studies on bicycling safety can be limited. New datasets on bicycling incidents are available via crowdsourcing applications, with opportunity for analyses that characterize reporting patterns. Our goal was to characterize patterns of injury in crowdsourced bicycle incident reports from BikeMaps.org. We extracted 281 incidents reported on the BikeMaps.org global mapping platform and analyzed 21 explanatory variables representing personal, trip, route, and crash characteristics. We used a balanced random forest classifier to classify three outcomes: (i) collisions resulting in injury requiring medical treatment; (ii) collisions resulting in injury but the bicyclist did not seek medical treatment; and (iii) collisions that did not result in injury. Results indicate the ranked importance and direction of relationship for explanatory variables. By knowing conditions that are most associated with injury we can target interventions to reduce future risk. The most important reporting pattern overall was the type of object the bicyclist collided with. Increased probability of injury requiring medical treatment was associated with collisions with animals, train tracks, transient hazards, and left-turning motor vehicles. Falls, right hooks, and doorings were associated with incidents where the bicyclist was injured but did not seek medical treatment, and conflicts with pedestrians and passing motor vehicles were associated with minor collisions with no injuries. In Victoria, British Columbia, Canada, bicycling safety would be improved by additional infrastructure to support safe left turns and around train tracks. Our findings support previous research using hospital admissions data that demonstrate how non-motor vehicle crashes can lead to bicyclist injury and that route characteristics and conditions are factors in bicycling collisions. Crowdsourced data have potential to fill gaps in official data such as insurance, police, and hospital reports.

1. Introduction

Bicycling infrastructure planning and safety studies commonly rely on incident data generated through official reporting systems such as insurance claims, police reports, and hospital records. However, researchers have found these sources substantially underestimate the burden of bicycling incidents (Langley et al., 2003; Lopez et al., 2012; Nelson et al., 2015; Watson et al., 2015; Winters and Branion-Calles, 2017). For example, a study in Vancouver, British Columbia demonstrated that less than 20 % of incidents involving bicyclists were captured by official reports (Winters and Branion-Calles, 2017). Official sources overrepresent bicycle-motor vehicle crashes and incidents with severe injury outcomes (Tin Tin et al., 2013; Winters and Branion-Calles, 2017), and underrepresent non-fatal bicyclist crashes (Amoros et al., 2006; Stutts and Hunter, 1998; Tin Tin et al., 2013; Watson et al.,

2015; Winters and Branion-Calles, 2017). The lack of universal standards across reporting systems, let alone different jurisdictions, also limit data quality, and details on crash and injury circumstances are often minimal (City of Boston, 2013; Lopez et al., 2012; Nelson et al., 2015; Teschke et al., 2012).

Alternatively, crowdsourcing offers researchers the opportunity to gather rich and diverse data with fewer obstacles compared to traditional data collection methods (Romanillos et al., 2015). Crowdsourced data are contributed by citizen volunteers via platforms such as web and mobile applications and are emerging as a valuable resource in active transportation planning (Misra et al., 2014; Romanillos et al., 2015; Schlossberg and Brehm, 2009; Smith, 2015). For instance, crowdsourced bicycling data have provided novel information in a variety of areas, including public bike share and bicyclist amenities (Misra et al., 2014; Romanillos et al., 2015), route choice and safety

* Corresponding author.

E-mail addresses: jaimyf@sfu.ca (J. Fischer), trisalyn.nelson@asu.edu (T. Nelson), klaberee@uvic.ca (K. Laberee), mwinters@sfu.ca (M. Winters).

<https://doi.org/10.1016/j.aap.2020.105695>

Received 25 November 2019; Received in revised form 24 June 2020; Accepted 14 July 2020

0001-4575/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

(Baker et al., 2017; Misra et al., 2014; Romanillos et al., 2015; Schlossberg and Brehm, 2009), complete streets (Schlossberg and Brehm, 2009), transportation projects (Misra et al., 2014), bicycle volumes (Griffin and Jiao, 2015; Jestico et al., 2016), and bicycling infrastructure provision (Ferster et al., 2019; Hochmair et al., 2015). Likewise, crowdsourced data on bicycling safety bode well as a supplement to official sources, most notably for capturing less severe and non-motor vehicle crashes and near misses (Branion-Calles et al., 2017; Goodchild, 2007; Jestico et al., 2017; Nelson et al., 2015).

BikeMaps.org is a global crowdsourced mapping platform for citizen reporting of bicycling incidents. Citizen mappers use the BikeMaps.org website or mobile app to mark a location where a bicycling crash or near miss occurred and provide details on rider experience and safety behaviors, bicycling environment, bicycling infrastructure, personal and trip characteristics, and injury. Citizens also describe the incident by way of a mandatory open-ended text field (“Please give a brief description of the incident.”). The detailed descriptions add nuanced depth to reports and help contextualize injury hotspots, issues specific to locale, and different crash or near miss circumstances. In cities that promote BikeMaps.org, crowdsourced incidents can provide data on themes that are difficult to draw from official sources (Ferster et al., 2017a;2017b). For example, BikeMaps.org data contributed to bicycling safety knowledge on multiuse trails by identifying trail-road intersection characteristics that were associated with near misses and collisions (Jestico et al., 2017). In another study, compared to automobile insurance records, BikeMaps.org data provided more information about crashes in locations with bicycle facilities, and near miss reports provided data on situations that bicyclists found intimidating, such as interactions with automobiles on streets with no bicycle infrastructure (Branion-Calles et al., 2017).

We aimed to characterize injury patterns in BikeMaps.org collision reports. Such investigation is necessary to better understand what crowdsourced incident data capture about bicycling safety and is a requisite step toward their wider use in research and practice. We extracted 281 collisions mapped on BikeMaps.org and analyzed 21 explanatory variables representing personal, trip, route, and crash characteristics. We used a balanced random forest classifier to classify three outcomes: (i) collisions resulting in injury requiring medical treatment; (ii) collisions resulting in injury but the bicyclist did not seek medical treatment; and (iii) collisions that did not result in injury. We report the ranked variable importance in classifying injury outcomes. By knowing what conditions are most associated with injury we can target interventions to minimize those conditions, reduce injury, and increase safety.

2. Materials and methods

2.1. Study area

The study area includes Greater Victoria core municipalities (City of Victoria, District of Saanich, Township of Esquimalt, and the District of Oak Bay) on the southern tip of Vancouver Island, British Columbia, Canada (Fig. 1). The four municipalities have a combined population of 235,689 and a relatively high bicycling journey to work mode share with approximately ~8.7 % of trips made by bicycle (Statistics Canada, 2017). The region experiences a Mediterranean climate characterized by dry summers and mild, wet winters, with temperatures ranging between a daily average minimum of 1.3 °C (34 °F) in the winter and maximum of 22.4 °C (72 °F) in summer (Government of Canada, 2019). Total annual precipitation is 880 mm with the majority falling during the winter months.

The region has an established bicycling network with cycle tracks, painted on-street bike lanes, signed shared roadways, and a regional trail network with nearly 100 km of multiuse pathways separated from motor vehicles. The Galloping Goose Regional Trail, Lochside Regional Trail, and E&N Rail Trail are the primary elements of the regional trail

network. These mostly paved, multiuse paths are heavily utilized by commuter and recreational bicyclists alike and are notable as arteries for moving bicyclist commuter traffic between the urban core and outlying residential areas.

2.2. BikeMaps.org data

BikeMaps.org is a tool for citizen reporting of bicycling collisions, falls, and near misses. Citizen mappers identify the locations of bicycling incidents with a pin on the online map and complete follow-up questions in three attribute categories: incident details, conditions, and personal details (see Nelson et al. (2015) for rationale of attributes collected and BikeMaps.org for full list). Citizens also provide detailed incident descriptions in an open-ended text section. Seven out of nine attributes within the incident detail category are mandatory, including the detailed description, while all attributes in the conditions and personal details categories are optional.

We used a geospatial dataset of bicycling incidents reported to BikeMaps.org for the study area. These data represent 281 collisions, of which 211 were collisions with objects or motor vehicles and 63 were falls. Incidents occurred from 2005 to 2019, with 91 % reported between 2013 and 2019.

Mandatory incident details include the question “Were you injured?” to which citizen mappers answered yes or no. If yes, they were required to answer if: overnight stay in hospital was required; they visited the hospital emergency department; they saw a family doctor; or medical treatment was not required. In our dataset injury circumstances varied, with 78 crashes requiring a hospital visit or admission, 44 requiring a visit to a family doctor, 78 where the bicyclist was injured but did not seek medical treatment, and 81 crashes where no injury was reported. To facilitate classification, we collapsed reported injuries into three categories: (i) collisions resulting in injury requiring medical treatment; (ii) collisions resulting in injury but the bicyclist did not seek medical treatment; and (iii) collisions that did not result in injury.

2.3. Explanatory variables

To identify the circumstances important for classifying bicycling injury, we included 21 explanatory variables in our model. Eighteen were collected or derived from BikeMaps.org reports (Table 1). Precipitation and visibility at the time of the incident were attributed to BikeMaps.org reports using Dark Sky weather API (Dark Sky API, 2020), and three additional explanatory variables representing built environment characteristics were manually attributed (Table 2): (a) road classification (major, collector, local, off-street); (b) if the incident was at intersection (Y/N) and (c) bicycle ridership (high, medium, low). Bicycling volumes are an important variable for quantifying exposure (Fournier et al., 2016; Lovelace et al., 2016; Lusk et al., 2013; Nelson et al., 2015; Reynolds et al., 2009), and including ridership in our study characterizes how exposure influences injury.

2.4. Analysis

We used a balanced random forest classifier to classify three injury outcomes derived from the BikeMaps.org incident details question “Were you injured?”: (i) collisions resulting in injury requiring medical treatment (more severe collisions); (ii) collisions resulting in injury but the bicyclist did not seek medical treatment (less severe collisions); and (iii) collisions that did not result in injury (minor collisions). Our rationale was to characterize injury patterns and examine the relative importance of explanatory variables for each outcome. Results show variable importance and the direction of relationship between explanatory variables and each outcome class, and we used detailed incident descriptions from BikeMaps.org reports to aid in interpretation.

Random forests are an extension of classification and regression tree (CART) methods that use an ensemble of many (hundreds to thousands)

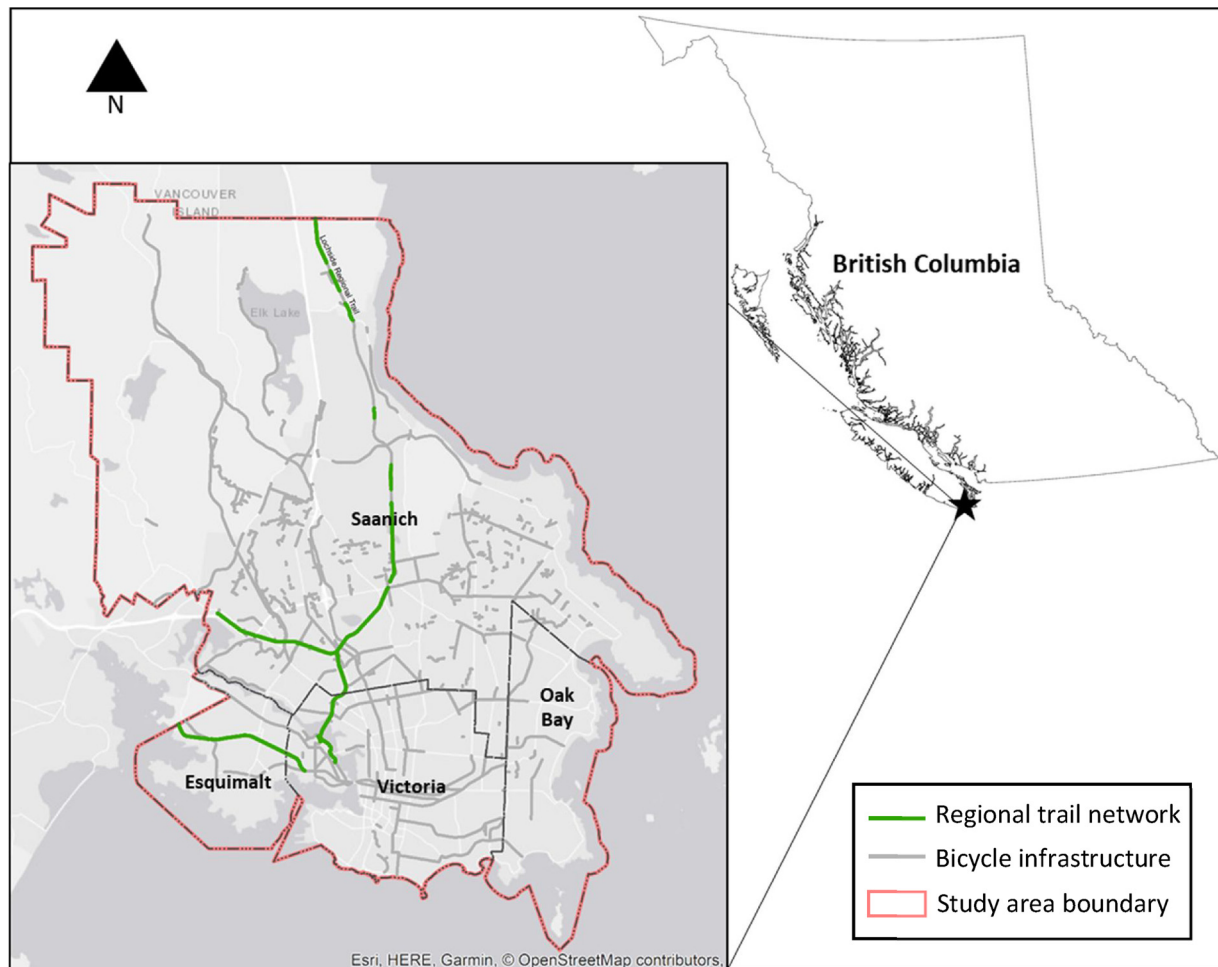


Fig. 1. Study area: Victoria, BC core municipalities and established regional bicycling network.

uncorrelated decision trees to reach a final ranking of variable importance. The model learns patterns in the data using bootstrap sampling and makes final classifications by averaging results across all trees. Each decision tree is trained on a different sample of observations (a bootstrap sample) and only a random subset of explanatory variables. This approach reduces correlation between trees (Hastie et al., 2013) and addresses overfitting, even when there are more explanatory variables than observations (Berk, 2016). Overfitting is also limited because the final class assigned to each observation is based on the ‘majority vote’—or the most commonly assigned class across all the trees it was used to fit. The balanced random forest variation improves modeling of an imbalanced dataset by down sampling the number of observations from the majority class to match that of the minority class in each bootstrapped sample (Branion-Calles et al., 2016; Chen et al., 2004). Some BikeMaps.org report variables had incomplete attributes thus we imputed missing values using proximity (RfImpute function).

Both the relative variable importance and model performance are estimated internally using ‘out-of-bag’ (OOB) data, which are the data left out of the bootstrapped samples (Branion-Calles et al., 2016; Cutler et al., 2007; Hastie et al., 2013). For all trees, OOB data are passed down and classification accuracy is recorded. Misclassification error is summed across all the trees and used as a measure of model accuracy. Variable importance reflects the mean decrease in OOB accuracy across all trees when the values of a given variable are randomly permuted in OOB data, with variables causing the largest decrease in model accuracy ranking as the most important (Branion-Calles et al., 2016; Cutler et al., 2007; Hastie et al., 2013). Variables with zero or negative values offer no explanatory power to the model (Branion-Calles et al., 2016).

We used partial dependence plots to visualize the direction of relationship between explanatory variables and each of the three injury classes (Hastie et al., 2013) and gain insight on common issues related to injury in BikeMaps.org reports. Plots show the probability of an outcome for the different values of an explanatory variable (i.e., variable attributes), holding all other variables constant. The y-axis is the log odds of an injury classification and the x-axis shows the change in class probabilities across the range of variable attributes. Higher logits mean that the outcome class is more probable for the corresponding variable attribute (Branion-Calles et al., 2016). We focused on the six most important explanatory variables based on our variable importance plot (Fig. 2). For model performance we used accuracy overall and for each class, and the kappa score, which quantifies the agreement between the predicted and actual classes, adjusted for agreement that might be due to random chance (Branion-Calles et al., 2016; Fatourehchi et al., 2008; Strobl et al., 2007).

3. Results

We analyzed 281 bicycling collisions reported to BikeMaps.org. The proportion of incidents for each outcome class was: 43.4 % (n = 122) for collisions resulting in injury requiring medical treatment; 27.8 % (n = 78) for collisions resulting in injury but the bicyclist did not seek medical treatment; and 28.8 % (n = 81) for minor collisions that did not result in injury. Most incidents were collisions with motor vehicles, pedestrians, bicyclists, or animals (75 %, n = 211), and the remaining 25 % (n = 70) were single bicycle crashes. The median age for BikeMaps.org incident reports was 39 years for males, and 35 years for

Table 1
BikeMaps.org variables used in Random Forest classification.

Variable	Completeness (%)	Variable attributes
BikeMaps.org incident details		
Time of day incident occurred(When was the incident?)	100 %	AM peak (6–9am, weekdays) PM peak (3–6pm, weekdays) inter-peak hours
Day incident occurred(When was the incident?)	100 %	Weekday weekend
Season incident occurred(When was the incident?)	100 %	Winter Spring Summer Fall
What type of incident was it?	100 %	Collision with moving object or vehicle collision with stationary object or vehicle fall
What sort of object did you collide with?	100 %	Animal another bicyclist curb other pedestrian pothole roadway sign/post train tracks vehicle, angle vehicle, head on vehicle, open door vehicle, passing vehicle, rear end vehicle, side vehicle, turning left vehicle, turning right
Were you injured? (outcome)	100 %	Injury, medical treatment injury, no medical treatment no injury
What was the purpose of your trip?	99 %	Commute exercise or recreation personal business social reason
BikeMaps.org conditions		
What were the road conditions?	87 %	Dry wet loose sand or gravel icy snowy don't remember
How were the sight lines?	85 %	No obstructions view obstructed glare or reflection obstruction on road don't remember
Were there cars parked on the roadside?	85 %	Yes no
^a Where were you riding your bike?	100 %	Street with no bicycle facility on a local street bikeway on a painted bicycle lane on a protected on-street bicycle path (cycle track) on a multiuse path
What was the terrain like?	87 %	Uphill downhill flat don't remember
How were you moving?	87 %	Heading straight turning left turning right don't remember
BikeMaps.org personal details		
Age category(What is your birth year?)	73 %	Under 30 31–40 41–50 50+
Please select your sex	74 %	Male female other
Do you bike at least once per week?	77 %	Yes no don't know
Were you wearing a helmet?	77 %	Yes no don't know
BikeMaps.org environmental		
Precipitation (“Dark Sky API: Documentation Overview,” n.d.)	100 %	Precipitation no precipitation
Visibility	100 %	Day night

^a Verified with external datasets from City of Victoria Open Data Catalogue (City of Victoria, 2017) and partners at the Regional Transportation Authority.

females, with males accounting for 70 % of all reports. Past analysis of data representativeness shows that BikeMaps.org has a high proportion of reports from people 25–34 years of age and that females tended to report more incidents in the urban core, especially collisions (46 %) (Ferster et al., 2017b). Completion of optional attributes ranged from 73 to 99 % (Table 1).

We modeled bicycling incidents across three injury classes: crashes that required medical attention, crashes resulting in minor injuries, and crashes with no reported injury. The six most important explanatory variables in the model were (in ranked order): the type of object the bicyclist collided with; the incident type; the terrain the bicyclist was riding on; bicyclist movement (i.e. traveling straight or turning); bicyclist age group; and bicycle ridership at the incident site (Fig. 2). Our random forest model had an overall accuracy of 42.4 % with accuracy varying between injury severity classes. Accuracy was 48.4 % for collisions resulting in injury requiring medical treatment, 32.1 % for collisions resulting in injury but the bicyclist did not seek medical treatment, and 43.2 % for collisions that did not result in injury. The model had a kappa score of 0.12, indicating model performance 12 % better than a random classifier.

Partial dependence plots for the six most important explanatory variables greater detail the circumstances specific to each injury class.

Table 2
Additional variables used in Random Forest classifier.

Variable	Variable attributes	Relevance	Data source
Road classification	Major ^a collector local off-street ^b	Related to route safety and bicycling risk (Allen-Munley and Daniel, 2006; Teschke et al., 2012).	Regional transportation authority
Intersection	Yes no	Intersections are common locations for cycling incidents involving motor vehicles (Dozza and Werneke, 2014).	Regional transportation authority
Bicycle ridership	High medium low	Characterizes exposure (Nelson et al., 2015).	Jestico et al. (2016); Regional transportation authority

^a Major streets included arterials (most with > 2 demarcated lanes); and collectors (most with > 2 demarcated lanes) (Reynolds et al., 2009).

^b Included as a road classification in the dataset provided by the Regional Transportation Authority. Represents routes that are closed to motorized vehicles and are physically separated from traffic on segments between intersections.

In the next section we describe the results from partial dependence plots for all injury classes used in the model, and Fig. 3 illustrates partial dependence plots for the most severe collision class (collisions resulting in injury requiring medical treatment).

3.1. Collisions resulting in injury requiring medical treatment

The most important variable in the model was the object that a bicyclist collided with, and partial dependence plots show that collisions with animals, infrastructure (such as fixed signs/posts, train tracks, transient hazards), and left-turning motor vehicles had the highest probabilities of being classified as an injury requiring medical treatment (Fig. 3a). The type of incident was the second most important variable and collisions with both moving and stationary objects had higher probabilities (Fig. 3b). Terrain, bicyclist movement, and age group were also important, with downhill terrain, bicyclist traveling straight, and the 41–50 age group more likely for this class (Fig. 3c, d, and e). Ridership was the sixth most important variable and moderate ridership at the site of the incident was associated with more severe injury (Fig. 3f).

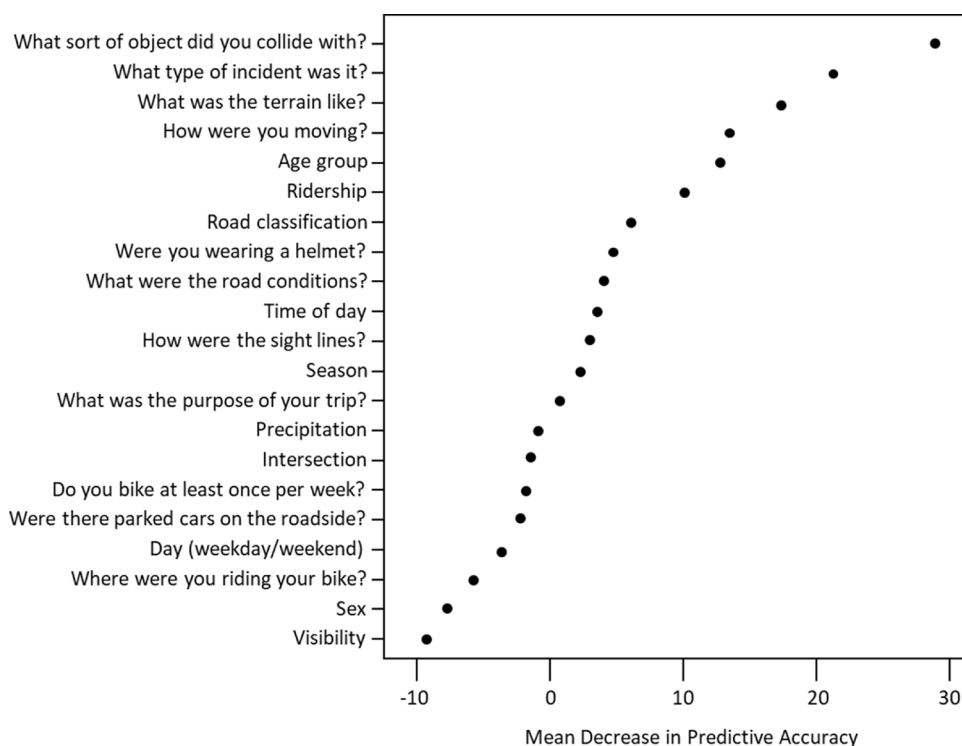


Fig. 2. Variable importance plot. Measured using OOB predictions, the plot shows the relative importance of each BikeMaps.org explanatory variable for correctly classifying injury outcomes. Variables causing the largest decrease in predictive accuracy rank higher in importance and those with zero or negative values offer no explanatory power to the model.

3.2. Collisions resulting in injury, but the bicyclist did not seek medical treatment

Conflicts with other bicyclists, right hooks, and doorings had the highest probabilities of being classified as collisions where the bicyclist was injured but did not seek medical treatment. Falls were most likely for this outcome class, and downhill terrain, if the bicyclist was turning right, the over 50 age group, and high ridership were also important.

3.3. Collisions that did not result in injury

Conflicts with pedestrians and passing motor vehicles had the highest probabilities for minor collisions where no injury occurred. Minor collisions were also associated with moving objects or vehicles. Uphill or flat terrain and/or bicyclists turning left were also likely circumstances in this class, as were the under 40 age group and moderate ridership at the site of the incident.

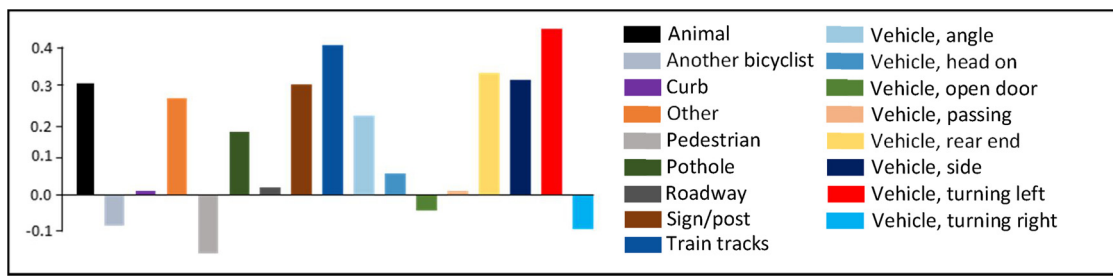
4. Discussion

Incomplete data has limited our understanding of how to improve safety for bicyclists. Crowdsourcing is a promising supplement to official bicycling safety data and is of growing interest for active transportation planning and monitoring. Modeling the variables captured in crowdsourced BikeMaps.org data allowed us to summarize key patterns in circumstances and conditions that led to bicycling injury, and we found that crowdsourced data provided rich detail that aided interpretation of results. The most important variable overall was the object the bicyclist collided with, and collisions with animals, surface features, and left-turning motor vehicles had the highest class probabilities for injuries requiring medical treatment. Crashes with moving and stationary objects or vehicles were more likely to be classified as an injury requiring medical treatment than falls were, as were downhill terrain and traveling along a straight trajectory. The 41–50 age group and moderate ridership were also important in classifying more severe collisions. While not all these variables can be modified (i.e., bicyclist age), surface features, left-turning motor vehicles, and even exposure to

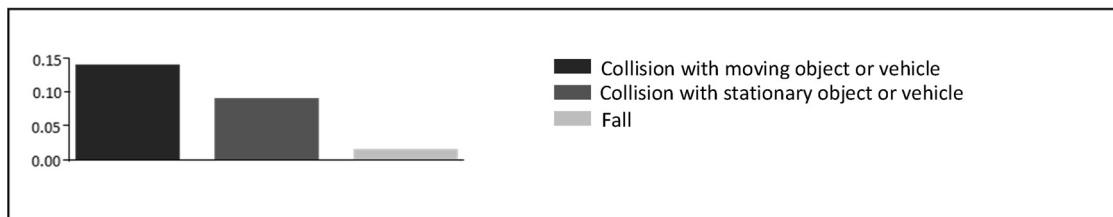
animals can be impacted by interventions aimed to increase bicycling safety. Here, the rank ordering of variable importance to injury helps decision makers focus change on conditions that will lead to the strongest safety improvements.

Collisions with motor vehicles were important in classifying injury. This finding aligns with previous research that implicates collisions with motor vehicles in more severe injury outcomes in bicyclists and highlights how infrastructure planning should continue to prioritize separating bicycles from motor vehicle traffic (Cripton et al., 2015). In this study, nearly two thirds (63 %, n = 176) of collisions involved a moving motor vehicle, and of these, about half (47 %, n = 82) required medical attention. The other half were less severe or minor collisions resulting from circumstances such as close passes, doorings, and evasive maneuvers; such incidents are unlikely to be captured in official safety data.

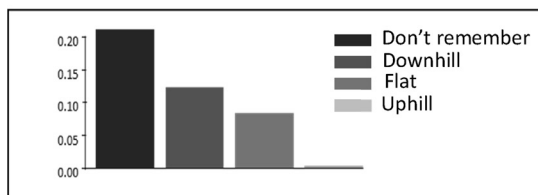
Compared to crashes with motor vehicles, single bicycle crashes are underrepresented in official data but are a significant contributor to injury (Beck et al., 2019; Schepers et al., 2015). Our research corroborates findings that single bicycle crashes—where no direct collision with a motor vehicle, pedestrian, bicyclist, or animal occurred—are common in places with high bicycle mode share (Schepers and den Brinker, 2011; Schepers and Klein Wolt, 2012). In our study, 25 % of crashes were single bicycle, and their importance in classifying injury patterns is evident. For example, collisions with road surface features like train tracks and signs/posts, transient hazards (e.g., lumber in the bike lane), and falls due to road surface conditions were more likely to be classified as an injury requiring medical treatment. A recent international review estimates that between 60–95 % of crashes requiring hospital admissions are single bicycle (Schepers et al., 2015). The characteristics of bicycle routes may influence the potential for single bicycle crashes (Schepers and den Brinker, 2011; Schepers and Klein Wolt, 2012; Teschke et al., 2014) and crowdsourced bicycling incident data may help direct attention to routes where these incidents are occurring. As such, from a planning perspective, there is potential in crowdsourcing to aid planning, engineering, and public works departments in safety improvements by highlighting issues with infrastructure and route maintenance.



(a) What sort of object did you collide with?



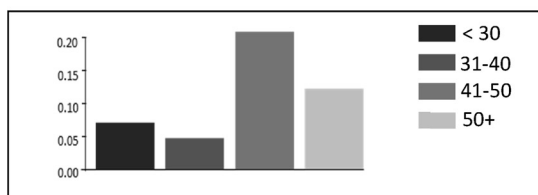
(b) What type of incident was it?



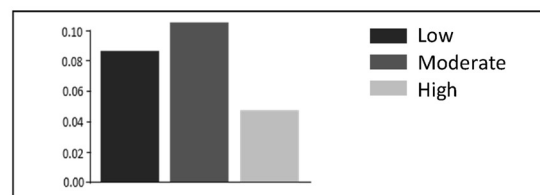
(c) What was the terrain like?



(d) How were you moving?



(e) Age group



(f) Ridership

Fig. 3. Partial dependence plots for the injury requiring medical treatment class. The y-axis is log odds of the class probability, and plots are interpreted as the probability of a class prediction over the range of variable attribute values. Higher values on the y-axis mean the class prediction was more probable for that value of the explanatory variable. For example, in (a) the probability of an injury requiring medical treatment was highest for collisions with animals, train tracks, and left-turning motor vehicles.

Downhill grades were more probable for the injury requiring medical treatment class than flat or uphill terrain. This result aligns with other research that found downhill grades were associated with injury risk due to faster motor vehicle and bicyclist speeds (Cripton et al., 2015; Teschke et al., 2012). In our study bicyclists reported a variety of incidents that occurred while traveling downhill, including collisions with motor vehicles crossing bicyclists' paths (left or right turns, pulling out of side streets) and falls related to poor surface conditions such as potholes, gravel, and ice. Bicyclists traveling straight were also important to the severe collision class. Two-thirds of BikeMaps.org reports used in this study indicate the bicyclist was traveling straight and of these incidents the majority involved motor vehicles. Of those traveling straight: 27 % were right hook collisions, where a motor vehicle overtook a through-traveling bicyclist from behind and turned right directly in front of them; 19 % were collisions with a left-turning vehicle; and 22 % were with a passing vehicle. Official reports rarely capture turning movement, and thus BikeMaps.org data may provide an opportunity for new research in this area. Our age results differ from previous studies that showed an increased risk of more severe injury for

older bicyclists (Cripton et al., 2015). One possible explanation for this difference is that younger people are more likely to report an incident to BikeMaps.org. This younger reporting pattern also aligns with regional travel behaviour, with older adults making fewer bicycling trips than the rest of the population (Ferster et al., 2017b).

An important aspect of this study is the inclusion of bicyclist volumes, which allows us to characterize how level of exposure changes likelihood of an injury occurring in a particular location. We incorporated a ridership variable in our model, derived from crowd-sourced data, which has an accuracy of 62 % (Boss et al., 2018; Griffin and Jiao, 2015; Jestico et al., 2016). Moderate ridership was more probable for the injury requiring medical treatment class, but this result may be a marker of infrastructure. Ridership is high along Victoria's regional trail network, which is comprised mainly of paved multiuse paths, and along major corridors that link with these infrastructures. With this in mind, our results suggest that it is important to differentiate between infrastructure types: while higher ridership may increase bicyclist visibility with respect to motor vehicle traffic, it may be irrelevant in preventing non-motor vehicle collisions such as those due to

environmental conditions or conflicts with animals, pedestrians, and other bicyclists.

Typically, official data lack detailed narratives or access to narratives is limited. In this study, we found the detailed descriptions in BikeMaps.org reports valuable in providing context to modeling results. For example, our model found that collisions with train tracks and animals were more likely to be classified as an injury requiring medical treatment than most other collision objects. In Victoria, tracks cross a high ridership bicycle route in two places, and citizens report issues specific to these locations: the bike lane crosses the tracks in the intersection at an angle; and the paint treatment in the bike lane becomes slippery when road conditions are wet or icy. A citizen report summarizes: *“The bike lane is painted green in the curve that crosses the railroad tracks. The wet surface made it slippery, which caused me to skid while following the curve.”* Another writes, *“The painted green section around the train tracks freezes quickly and is very dangerous.”* Deer-human conflicts are common in the Greater Victoria region, which may explain why collisions with animals were an important injury-reporting pattern. The mainstream media have already reported on the impact that deer in the region have on bicyclist safety (CTV Vancouver Island, 2018; Harnett, 2017; Times Colonist, 2019). Such reports are an example of where crowdsourced data can help fill gaps because insurance data fail to capture incidents with no motor vehicle involvement. Again, detailed descriptions provide context. For example, a bicyclist who reported a collision with a deer wrote: *“I was on a morning group ride when a deer ran right in front of us. Resulting in a broken collar bone for me and my friend suffered the same plus a broken thumb.”* Another citizen reported, *“Head on collision with a deer. Suffered punctured lung, cracked elbow and a lot of road rash. Was in hospital for two days after surgery”*. These examples demonstrate how the detailed descriptions in BikeMaps.org reports reveal bicycling safety issues specific to locale that are unlikely to be unaccounted for in conventional data. Crowdsourced data can lend insight on undetected issues and provide context that may be useful for targeting interventions to reduce future risk.

5. Strengths and limitations

BikeMaps.org has captured more than 7000 bicycling incidents reported in over 40 countries. While the random forest ensemble method is designed to address overfitting, our model predictions may be impractical to extrapolate to study areas with different bicycling environments. (Jeong et al., 2016). As such, we suggest our approach is most suitable for descriptive analysis of BikeMaps.org reports in locations ample data. Another limitation is that while it would be interesting to compare injury patterns across official and crowdsourced data sources, insurance claims, which are the main source of bicycling incident data available for the study area, have limited details on injury or crash circumstances and thus comparison was not possible. We used variable importance and partial dependence plots to describe our model results. Variable importance reveals global relationships in the model but is less comprehensive than metrics such as Shapley values or Local Interpretable Model-agnostic Explanations (LIME) (Molnar, 2018; Ribeiro et al., 2016). For our purpose we found that using variable importance alongside partial dependence plots was sufficient to identify unique patterns within our dataset, but future work might benefit from including metrics derived from local measures. The BikeMaps.org platform is available in English, French, and Icelandic, which could limit reporting from those who do not speak or write in these languages. Results may also be influenced by self-selection bias as younger adult males are more likely to submit a BikeMaps.org report in the study area compared to the general bicycling population (Ferster et al., 2017b). As with any crowdsourced data, uncertainty surrounding data quality may be an issue, thus highlighting the necessity to consider such data as supplementary to data such as traditional reports.

6. Conclusions

In crowdsourced data on bicycling safety, incidents with the highest probability of classifying an injury requiring medical attention were collisions with animals, infrastructure such as fixed signs or posts, train tracks, transient hazards, and left-turning motor vehicles, while incidents most likely to be classified as less severe were collisions resulting from conflicts with other bicyclists, right hooks, doorings, and falls. We found that some of the circumstances linked to injury are modifiable: left-turning motor vehicles, route maintenance and surface conditions, and exposure to animals along bicycling routes can all be targeted to improve bicycling safety. From a planning, engineering, and maintenance perspective, ranking BikeMaps.org variable importance can help prioritize resources and safety improvements. In the Victoria region specifically, we recommend that hazardous paint treatments around train tracks be fixed and that public works focus on route maintenance, especially during the darker winter months when conditions are more formidable. We also suggest that deer-human conflicts continue to be monitored and mitigated as these incidents can be fatal but may not be captured in official data. Our method demonstrates how BikeMaps.org data can be used to better understand the circumstances and conditions that lead to bicycling injury and highlights reporting patterns in crowdsourced bicycle incident data. The approach used in this study can be applied in other settings and to different incident severity classes such as near misses. As BikeMaps.org data have been reported in more than 40 countries globally the methods demonstrated here can be used by planners, researchers, advocates, and other local stakeholders who are working to improve bicycling safety and increase ridership.

Funding

This work was supported by the Public Health Agency of Canada [#1516-HQ-000064] and the Arizona State University Foundation.

CRediT authorship contribution statement

Jaimy Fischer: Conceptualization, Methodology, Writing - original draft, Visualization, Investigation. **Trisalyn Nelson:** Conceptualization, Funding acquisition, Methodology, Writing - review & editing. **Karen Laberee:** Writing - review & editing, Project administration. **Meghan Winters:** Supervision, Validation, Writing - review & editing, Resources.

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.

Acknowledgements

This BikeMaps.org research and outreach has been funded by a grant from the Public Health Agency of Canada (PHAC). We acknowledge Taylor Denouden, Darren Boss, Colin Ferster, and Ayan Mitra in creating and maintaining the technology used to collect BikeMaps.org incident data, and the Capital Regional District for their support of outreach. We thank all members of the BikeMaps.org team whose outreach has helped BikeMaps.org reach a broad number of bicyclists in numerous locations. We also thank everyone who took the time to report an incident on BikeMaps.org.

References

Allen-Munley, C., Daniel, J., 2006. Urban bicycle route safety rating model application in Jersey City, New Jersey. *J. Transp. Eng.* 132, 499–507. <https://doi.org/10.1061/>

- (ASCE)0733-947X(2006)132:6(499).
- Amoros, E., Martin, J.L., Laumon, B., 2006. Under-reporting of road crash casualties in France. *Accid. Anal. Prev.* 38 (4), 627–635. <https://doi.org/10.1016/j.aap.2005.11.006>.
- Baker, K., Ooms, K., Verstockt, S., Brackman, P., De Maeyer, P., Van de Walle, R., 2017. Crowdsourcing a cyclist perspective on suggested recreational paths in real-world networks. *Cartogr. Geogr. Inf. Sci.* 44 (5), 422–435. <https://doi.org/10.1080/15230406.2016.1192486>.
- Beck, B., Stevenson, M.R., Cameron, P., Oxley, J., Newstead, S., Olivier, J., Boufous, S., Gabbe, B.J., 2019. Crash Characteristics of On-road Single-bicycle Crashes: an Under-recognised Problem. pp. 1–5. <https://doi.org/10.1136/injuryprev-2018-043014>.
- Berk, R.A., 2016. *Statistical Learning from a Regression Perspective*, 2nd ed. Springer BikeMaps.org [WWW Document], n.d.
- Boss, D., Nelson, T., Winters, M., Ferster, C.J., 2018. Using crowdsourced data to monitor change in spatial patterns of bicycle ridership. *J. Transp. Heal.* 9, 226–233. <https://doi.org/10.1016/j.jth.2018.02.008>.
- Branion-Calles, M.C., Nelson, T.A., Henderson, S.B., 2016. A geospatial approach to the prediction of indoor radon vulnerability in British Columbia. Canada. *J. Expo. Sci. Environ. Epidemiol.* 26 (6), 554–565. <https://doi.org/10.1038/jes.2015.20>.
- Branion-Calles, M., Nelson, T., Winters, M., 2017. Comparing crowdsourced near miss and collision cycling data and official bike safety reporting. *Trans. Res. Rec.* 2662 (1), 1–11. <https://doi.org/10.3141/2662-01>.
- Chen, C., Liaw, A., Breiman, L., 2004. Using random forest to learn imbalanced data. *Univ. California, Berkeley* 1999, 1–12 [doi:ley.edu/sites/default/files/tech-reports/666.pdf](http://doi.org/10.1136/tech-reports/666.pdf).
- City of Boston, 2013. Boston Cyclist Safety Report. (Accessed 9 Sep 2019). https://www.cityofboston.gov/news/uploads/16776_49_15_27.pdf.
- City of Victoria, 2017. Open Data Catalogue. (Accessed 4 July 2017). <http://opendata.victoria.ca/>.
- Cripton, P.A., Shen, H., Brubacher, J.R., Chipman, M., Friedman, S.M., Harris, M.A., Winters, M., Reynolds, C.C.O., Cusimano, M.D., Babul, S., Teschke, K., 2015. Severity of urban cycling injuries and the relationship with personal, trip, route and crash characteristics: analyses using four severity metrics. *BMJ Open* 5 (1), e006654. <https://doi.org/10.1136/bmjopen-2014-006654>.
- CTV Vancouver Island, 2018. Aggressive Buck Charges and Knocks Cyclist to Ground in Oak Bay. (Accessed 31 July 2019). <https://vancouverisland.ctvnews.ca/aggressive-buck-charges-and-knocks-cyclist-to-ground-in-oak-bay-1.4192915>.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. *Ecology* 88 (11), 2783–2792. <https://doi.org/10.1890/07-0539.1>.
- Dark Sky API, 2020. Documentation Overview <https://darksky.net/dev/docs> (Accessed 21 July 2020).
- Dozza, M., Werneke, J., 2014. Introducing naturalistic cycling data: what factors influence bicyclists' safety in the real world? *Transp. Res. Part F. Traffic Psychol. Behav.* 24, 83–91. <https://doi.org/10.1016/j.trf.2014.04.001>.
- Fatourehchi, M., Ward, R.K., Mason, S.G., Huggins, J., Schlögl, A., Birch, G.E., 2008. Comparison of evaluation metrics in classification applications with imbalanced datasets. *Proc. - 7th Int. Conf. Mach. Learn. Appl. ICMLA 2008* 777–782. <https://doi.org/10.1109/ICMLA.2008.34>.
- Ferster, C., Nelson, T., Laberee, K., Vanlaar, W., Winters, M., 2017a. Promoting crowd-sourcing for urban research: cycling safety citizen science in four cities. *Urban Sci.* 1 (2), 21. <https://doi.org/10.3390/urbansci1020021>.
- Ferster, C., Nelson, T., Winters, M., Laberee, K., 2017b. Geographic age and gender representation in volunteered cycling safety data: a case study of BikeMaps. *Appl. Geogr.* 88, 144–150. <https://doi.org/10.1016/j.apgeog.2017.09.007>.
- Ferster, C., Fischer, J., Manaugh, K., Nelson, T., 2019. Using OpenStreetMap to inventory bicycle infrastructure: a comparison with open data from cities using OpenStreetMap to inventory bicycle infrastructure: a comparison. *Int. J. Sustain. Transp.* 0, 1–10. <https://doi.org/10.1080/15568318.2018.1519746>.
- Fournier, N., Christofa, E., Knodler, M.A., 2016. A mixed methods investigation of bicycle exposure in crash rates. *Accid. Anal. Prev.* 130, 54–61. <https://doi.org/10.1016/j.aap.2017.02.004>.
- Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *Geo J.* 69 (4), 211–221. <https://doi.org/10.1007/s10708-007-9111-y>.
- Government of Canada, 2019. https://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?searchType=stnName&txtStationName=Victoria+%&searchMethod=contains&txtCentralLatMin=0&txtCentralLatSec=0&txtCentralLongMin=0&txtCentralLongSec=0&stnID=118&dispBack=0 Accessed 8 November 2019.
- Griffin, G., Jiao, J., 2015. Crowdsourcing Bicycle Volumes: exploring the role of volunteered geographic information and established monitoring methods. *Compend. Transp. Res. Board Annu. Meet.* 27 (1), 1–19.
- Harnett, C.E., 2017. Oak Bay Police Chief Struck by Deer While Riding Bike Recovering From Broken Bones. (Accessed 31 July 2019). <https://www.timescolonist.com/news/local/oak-bay-police-chief-struck-by-deer-while-riding-bike-recovering-from-broken-bones-1.23054001>.
- Hastie, T., Tibshirani, R., Friedman, J., 2013. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer.
- Hochmair, H.H., Zielstra, D., Neis, P., 2015. Assessing the completeness of bicycle trail and lane features in OpenStreetMap for the United States. *Trans. GIS* 19 (1), 63–81. <https://doi.org/10.1111/tgis.12081>.
- Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.-M., Gerber, J.S., Reddy, V.R., Kim, S.-H., 2016. Random forests for global and regional crop yield predictions. *PLOS One* 11 (6), e0156571. <https://doi.org/10.1371/journal.pone.0156571>.
- Jestic, B., Nelson, T., Winters, M., 2016. Mapping ridership using crowdsourced cycling data. *J. Transp. Geogr.* 52, 90–97. <https://doi.org/10.1016/j.jtrangeo.2016.03.006>.
- Jestic, B., Nelson, T.A., Potter, J., Winters, M., 2017. Multiuse trail intersection safety analysis: a crowdsourced data perspective. *Accid. Anal. Prev.* 103, 65–71. <https://doi.org/10.1016/j.aap.2017.03.024>.
- Langley, J., Dow, N., Stephenson, S., Kypril, K., 2003. Missing cyclists. *Inj. Prev.* 9 (4), 376–379. <https://doi.org/10.1136/ip.9.4.376>.
- Lopez, D.S., Sunjaya, D.B., Chan, S., Dobbins, S., Dicker, R.A., Francisco, S., 2012. Using trauma center data to identify missed bicycle injuries and their associated costs. *J. Trauma Acute Care Surg.* 6, 1602–1606. <https://doi.org/10.1097/TA.0b013e318265fc04>.
- Lovelace, R., Roberts, H., Kellar, I., 2016. Who, where, when: the demographic and geographic distribution of bicycle crashes in West Yorkshire. *Transp. Res. Part F Traffic Psychol. Behav.* 41, 277–293. <https://doi.org/10.1016/j.trf.2015.02.010>.
- Lusk, A.C., Morency, P., Miranda-Moreno, L.F., Willett, W.C., Dennerlein, J.T., 2013. Bicycle guidelines and crash rates on cycle tracks in the United States. *Am. J. Public Health* 103 (7), 1240–1248. <https://doi.org/10.2105/AJPH.2012.301043>.
- Misra, A., Gooze, A., Watkins, K., Asad, M., Le Dantec, C., 2014. Crowdsourcing and its application to transportation data collection and management. *Transp. Res. Res. J. Transp. Res. Board* 2414, 1–8. <https://doi.org/10.3141/2414-01>.
- Molnar, C., 2018. *Iml: an R package for interpretable machine learning*. *J. Open Src. Soft.* 3 (27), 786. <https://doi.org/10.21105/joss.00786>.
- Nelson, T.A., Denouden, T., Jestic, B., Laberee, K., Winters, M., 2015. BikeMaps.org: a global tool for collision and near miss mapping. *Front. Public Health Serv. Syst. Res.* 3, 53. <https://doi.org/10.3389/fpubh.2015.00053>.
- Reynolds, C.C.O., 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 Perspect.* 8 (1), 47. <https://doi.org/10.1186/1476-069X-8-47>. `rflmpute` function | R Documentation [WWW Document], n.d.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why should I trust you?: explaining the predictions of any classifier. *Proc. 22nd Int. Conf. Data Min., ACM* 1135–1144. <https://doi.org/10.1145/2939672.2939778>.
- Romanillos, G., Zaltz Austwick, M., Ettema, D., De Kruijff, J., 2015. Big data and cycling. *Transp. Res.* 36 (1), 114–133. <https://doi.org/10.1080/01441647.2015.1084067>.
- Schepers, P., den Brinker, B., 2011. What do cyclists need to see to avoid single-bicycle crashes? *Ergonomics* 54 (4), 315–327. <https://doi.org/10.1080/00140139.2011.558633>.
- Schepers, P., Klein Wolt, K., 2012. Single-bicycle crash types and characteristics. *Cycl. Res. Int.* 2, 119–135.
- Schepers, P., Agerholm, N., Amoros, E., Benington, R., Bjørnskau, T., Dhondt, S., de Geus, B., Hagemester, 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 (E1), e138–e143. <https://doi.org/10.1136/injuryprev-2013-040964>.
- Schlossberg, M., Brehm, C., 2009. Participatory geographic information systems and active transportation. *Transp. Res. Res. J. Transp. Res. Board* 2105, 83–91. <https://doi.org/10.3141/2105-11>.
- Smith, A., 2015. *Crowdsourcing Pedestrian and Cyclist Activity Data*. White Pap. Ser. January, 34.
- Statistics Canada, 2017. *Census Profile, 2016 Census. Journey to Work Data for Victoria (City), Oak Bay (DM), Saanich (DM), and Esquimalt (DM) Census Subdivisions*. (Accessed 31 July 2019). <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CSD&Code1=5917034&Geo2=PR&Code2=59&SearchText=Victoria&SearchType=Begin&SearchPR=01&B1=Journey%20to%20work&TABID=1&type=0>.
- Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics* 8, 25. <https://doi.org/10.1186/1471-2105-8-25>.
- Stutts, J.C., Hunter, W.W., 1998. Police reporting of pedestrians and bicyclists treated in hospital emergency rooms. *Transp. Res. Res. J. Transp. Res. Board* 1635, 88–92. <https://doi.org/10.3141/1635-12>.
- 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 (12), 2336–2343. <https://doi.org/10.2105/AJPH.2012.300762>.
- Teschke, K., Frendo, T., Shen, H., Harris, M.A., Reynolds, C.C., Cripton, P.A., Brubacher, J., Cusimano, M.D., Friedman, S.M., Hunte, G., Monro, M., Vernich, L., Babul, S., Chipman, M., Winters, M., 2014. Bicycling crash circumstances vary by route type: a cross-sectional analysis. *BMC Public Health* 14 (1), 16–19. <https://doi.org/10.1186/1471-2458-14-1205>.
- Times Colonist, 2019. Cyclist Taken to Hospital After Running Into Deer on Munn Road. (Accessed 31 July 2019). <https://www.timescolonist.com/news/local/cyclist-taken-to-hospital-after-running-into-deer-on-munn-road-1.23872623>.
- Tin Tin, S., Woodward, A., Ameratunga, S., 2013. Completeness and accuracy of crash outcome data in a cohort of cyclists: a validation study. *BMC Public Health* 13 (1), 420. <https://doi.org/10.1186/1471-2458-13-420>.
- Watson, A., Watson, B., Vallmuur, K., 2015. Estimating under-reporting of road crash injuries to police using multiple linked data collections. *Accid. Anal. Prev.* 83, 18–25. <https://doi.org/10.1016/j.aap.2015.06.011>.
- Winters, M., Branion-Calles, M., 2017. Cycling safety: quantifying the under reporting of cycling incidents in Vancouver, British Columbia. *J. Transp. Heal.* 7, 48–53. <https://doi.org/10.1016/j.jth.2017.02.010>.