

Crowdsourced data as a tool for cycling research on ridership trends and safety in the  
Capital Regional District

by

Benjamin Andrew Jestico  
B.Sc., University of Victoria, 2014

A Thesis Submitted in Partial Fulfillment  
Of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Geography

© Benjamin Andrew Jestico, 2016  
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy  
or other means, without the permission of the author.

Crowdsourced data as a tool for cycling research on ridership trends and safety in the  
Capital Regional District

by

Benjamin Andrew Jestico  
B.Sc., University of Victoria, 2014

**SUPERVISORY COMMITTEE:**

---

Dr. Trisalyn A. Nelson, Supervisor  
(Department of Geography, University of Victoria)

---

Dr. Meghan Winters, Member  
(Faculty of Health Sciences, Simon Fraser University)

**ABSTRACT**

Supervisory Committee:

---

Dr. Trisalyn A. Nelson, Supervisor  
(Department of Geography, University of Victoria)

---

Dr. Meghan Winters, Member  
(Faculty of Health Sciences, Simon Fraser University)

The benefits of cycling are well known and many communities are investing in cycling infrastructure in order to encourage and promote ridership. Safety is a primary concern for new cyclists and remains a barrier for increasing ridership. Understanding what influences cyclist safety requires knowing how many cyclists are riding in an area. Lack of ridership data is a common challenge for cycling research and limits our ability to properly assess safety and risk. The goal of our research was to incorporate new data available through crowdsourcing applications to advance cycling research on ridership and safety in the Capital Regional District (CRD), British Columbia (BC), Canada.

To meet our goal, our first analysis assessed how crowdsourced fitness app data can be used to map and to quantify the spatial and temporal variation of ridership. Using a dataset from a popular fitness app Strava, we compared how manual cycling counts conducted at intersections during peak commuting hours in Victoria compared to the number of crowdsourced cyclists during these same count periods. In order to estimate ridership at unsampled manual count locations, we used Poisson regression to model the association between manual counts and infrastructure variables found to influence ridership. Our results found that there was a linear association ( $r^2$  between 0.4 and 0.58) between crowdsourced cyclists and manual count cyclists, which amounted to one

crowdsourced cyclist representing 51 riders. Crowdsourced cyclist volumes, traffic speeds, on street parking, slope, and time of year were found to significantly influence the amount of cyclists in different count locations with a predictive accuracy of 62%. Overall, crowdsourced data from fitness apps are a biased sample of ridership; however, in urban areas in mid-size North American cities, cyclists using fitness apps may choose similar routes as commuter cyclists.

Our second analysis used crowdsourced data on cyclist incidents to determine the factors that influence incident reporting at multiuse trail and roadway intersections. Using incident reports from BikeMaps.org, we characterized attributes of reported incidents at intersections between multiuse trails and roads and also examined infrastructure features at these intersections that are predictors of incident frequency. We conducted site observations at 32 multiuse trail-road intersections in the CRD to determine infrastructure characteristics that influence safety. Using Poisson regression we modeled the relationship between the number of incidents (collision and near misses) and the infrastructure characteristics at multiuse trail-road intersections. We found that collisions were more commonly reported (over near misses) at multiuse trail-road intersections than road-road intersections (38% versus 27%), and incidents involving an injury were more common (35% versus 21%). Cycling volumes, vehicle volumes, and lack of vehicle speed reduction factors were associated with incident frequency. Our analysis was able to use crowdsourced cycling incident data to provide valuable evidence on the factors that influence safety at intersections between multiuse trails and roadways where diverse transportation modes converge.

Through this thesis we help to overcome limitations for cycling research and planning by demonstrating how crowdsourced ridership and safety data can help fill gaps and supplement available data. Our methodology integrates the high spatial and temporal resolution of crowdsourced cycling data with the detailed attributes provided by traditional ridership counts. We also demonstrate how volunteered safety data can allow new questions on safety to be explored. Improving data available for cycling research allows for a more comprehensive understanding of the factors that influence ridership and safety and, in turn, informs decisions targeted at increasing cycling.

## TABLE OF CONTENTS

SUPERVISORY COMMITTEE: .....	ii
ABSTRACT.....	iii
TABLE OF CONTENTS.....	vi
LIST OF TABLES.....	ix
LIST OF FIGURES .....	x
ACKNOWLEDGEMENTS.....	xi
CO-AUTHORSHIP STATEMENT.....	xii
1.0 INTRODUCTION .....	1
1.1 Research context .....	1
1.2 Research focus .....	2
1.3 Research goals and objectives .....	4
References.....	7
2.0 MAPPING RIDERSHIP USING CROWDSOURCED CYCLING DATA .....	10
2.1 Abstract.....	10
2.2 Introduction.....	11
2.3 Materials and Methods.....	14
2.3.1 Study area .....	14
2.3.2 Victoria Cycling Counts .....	14
2.3.3 Crowdsourced cycling data from the Strava fitness app.....	15
2.3.4 Summary of explanatory variables .....	15
2.3.5 Crowdsourced data and manual count comparisons.....	16
2.3.6 Modeling analysis.....	16
2.3.7 Model error analysis .....	17
2.3.8 Categorical breakdowns of cycling volumes .....	18

2.3.9 Mapping cycling volumes.....	18
2.4 Results.....	18
2.4.1 Modeling analysis results.....	19
2.4.2 Modeling error analysis results.....	20
2.4.3 Categorical breakdown of cycling volume results.....	20
2.4.4 Mapping cycling volumes results .....	20
2.5 Discussion.....	21
2.6 Limitations .....	24
2.7 Conclusions.....	25
Acknowledgements.....	26
References.....	35
3.0 MULTIUSE TRAIL INTERSECTION SAFETY ANALYSIS: A CROWDSOURCED DATA PERSPECTIVE.....	39
3.1 Abstract.....	39
3.2 Introduction.....	41
3.3 Materials .....	44
3.3.1 Study Setting.....	44
3.3.2 Crowdsourced cycling incident data from BikeMaps.org .....	44
3.3.3 Multiuse trail-road intersection data collection .....	45
3.4 Methods.....	45
3.4.1 Comparing incident report attributes at multiuse trail-road intersections to road-road intersections.....	45
3.4.2 Assessing intersection characteristics associated with incident reporting at multiuse trail-road intersections.....	46
3.5 Results.....	47

3.5.1 Comparing incident report attributes at multiuse trail-road intersections to road-road intersection results.....	47
3.5.2 Assessing intersection characteristics associated with incident reporting at multiuse trail-road intersection results.....	47
3.6 Discussion.....	48
3.7 Conclusions.....	52
3.8 Acknowledgements.....	53
References.....	62
4.0 CONCLUSIONS .....	66
4.1 Discussion and conclusions .....	66
4.2 Research contributions.....	69
4.3 Research opportunities.....	71
References.....	74

## LIST OF TABLES

Table 2.1 Victoria Sample Strava Rider Age and Gender Breakdown.....	27
Table 2.2 Explanatory variables considered for analysis.....	28
Table 2.3 Regression Estimates for GLM of cycling volumes along street segments .....	30
Table 2.4 Categorical breakdown analysis for thresholds of low, medium and high.....	31
Table 3.1 Select BikeMaps.org incident attributes that citizens are asked to detail when they map a cycling incident. ....	54
Table 3.2 Age and gender breakdown of incident reports at multiuse trail-road and road-road intersections .....	55
Table 3.3 Data collected at each multiuse trail-road intersection.....	56
Table 3.4 Incident report attributes at multiuse trail-road and road-road intersections in the CRD. ....	57
Table 3.5 Summary data for the number of incidents reported at multiuse trail-road intersections .....	58
Table 3.6 Poisson GLM regression results for infrastructure characteristics that influence incident reporting at multiuse trail-road intersections. ....	59

## LIST OF FIGURES

Figure 2.1 Study Area Victoria, BC Canada.....	32
Figure 2.2 Percentage of observations predicted within an amount of model error, via cross validation. For instance, 55% of predictions had errors of less than 30%.....	33
Figure 2.3 Peak period(AM and PM combined) predicted cycling volumes for Victoria, based on the GLM regression. ....	34
Figure 3.1 Study area in the Capital Regional District (CRD), BC, Canada.....	60
Figure 3.2 Sample intersections along the Galloping Goose Trail intersections.....	61

## ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Trisalyn Nelson for her encouragement, wisdom, support, and enthusiasm throughout my degree. Her expertise and insight kept me on track and focused during times when I was unsure of my research. I would like to extend my sincere gratitude to Dr. Meghan Winters for her guidance, knowledge, and insight that provided direction for this thesis. To the BikeMaps.org team, I feel so blessed to have been involved with such a unique project and I am extremely proud to be the first of many graduate students to work on the project. To the numerous volunteers and undergraduate students that helped in this research, I could not have done this without you and I cannot thank you enough. I would like to extend my thanks to the Capital Regional District, Bunt & Associates and the City of Victoria for their support throughout my research. My SPAR lab mates, thank you for all the laughs, pep talks, and encouragement when I needed it most. To my family, Mom, Dad, Steph, Brittany, and James, thank you for supporting and believing in me that one day I might actually finish school.

## **CO-AUTHORSHIP STATEMENT**

This thesis is the combination of two scientific manuscripts for which I am the lead author. Together Dr. Trisalyn Nelson and Dr. Meghan Winters developed the project structure, where utilizing crowdsourced cycling data to understand ridership trends and safety was identified as a key research opportunity for broadening the scientific knowledge of cycling data. For these two manuscripts, I led all research, data collection, data analysis, initial result interpretations and final manuscript writing. Dr. Trisalyn Nelson provided guidance in developing research questions and interpretation of results. Dr. Meghan Winters provided assistance with research insight, methodological considerations, and interpretation of results. Dr. Nelson and Dr. Winters supplied editorial comments and suggestions incorporated into the final manuscript.

## 1.0 INTRODUCTION

### 1.1 Research context

Cycling provides many benefits for both cyclists and communities. Cycling can reduce obesity, diabetes and heart disease by increasing daily activity (Pucher et al., 2010). Research has shown that countries whose citizens use active modes of transportation, such as cycling and walking, have lower levels of obesity (Bassett et al., 2008). Using active modes of transportation can have profound impacts on overall health and longevity (Pucher et al., 2010) and cities that are pedestrian and cycling friendly are associated with higher overall levels of happiness (Choi, 2013). At a city level, cycling presents an opportunity to shift modes of transportation away from motor vehicles (Winters and Teschke, 2010), which can reduce emissions and congestion (Cupples and Ridley, 2008). Cycling use is much lower in North American cities compared to other parts of the world (Pucher and Buehler, 2008), which creates a large potential to increase ridership (Pucher and Dijkstra, 2003) and the benefits of cycling in North America.

A primary barrier to increasing ridership is the perception that cycling is an unsafe mode of transportation (Nelson et al., 2015). Many people do not feel comfortable cycling (Dill and McNeil, 2012), especially near motor vehicles for fear of being involved in a crash (Parkin et al., 2007). However, most citizens are open and interested in the idea of cycling if conditions for safety were improved (Dill and McNeil, 2012). To evaluate safety, researchers often use traditional cycling crash statistics from police reports, insurance claims, or hospital records to determine where crashes occur and the conditions that may have led to a crash. However, crashes reported through traditional

means are often only reported for severe events that are relatively rare (Nelson et al., 2015). Minor or less severe events are often not reported but are important for safety and the perception of safety among potential riders.

A fundamental challenge for cycling research and planning is lack of data about where people ride and overall safety. Research has shown that cycling incidents are underreported and in some cases only 30% of incidents are recorded (Tin Tin et al., 2013; De Geus et al., 2012). Incidents that are less severe or those that do not involve a motorized vehicle are often not reported, but may provide valuable information for understanding and monitoring problem areas. Monitoring and assessing safety requires reliable ridership data in order to account for exposure or the ‘exposure to risk’ faced by each cyclist (Nelson et al., 2015; Reynolds et al., 2009). Ridership data are difficult to capture across time and space and existing methods can be expensive and time consuming to collect; however, communities investing in infrastructure rely on cycling data for monitoring progress, safety, planning, and prioritize cycling initiatives. Current methods for data collection could be vastly improved by considering new technologies for collecting and capturing data for cycling research.

## **1.2 Research focus**

New opportunities from crowdsourced applications present new information that can improve the data available for cycling research. Crowdsourced data (or Volunteered Geographic Information) allows citizens to collect geographic data about features in their environment (Goodchild, 2007). The popularity of smartphone apps continues to grow, which has created large amounts of data in a variety of different contexts (Kanhere,

2013). The ability to use Global Positioning Systems (GPS) through smartphone apps for route tracking purposes has become popular in the cycling community, with websites such as Strava, MapMyRide, and Garmin allowing users to map their own detailed cycling routes and monitor use (Jestico et al., 2016; Kessler, 2011). Crowdsourcing can also engage citizens through online platforms to be a part of the planning process for new cycling infrastructure (Le Dantec et al., 2015). Capturing information from local citizens through crowdsourcing could provide higher quality information due to their local knowledge of their own community (Kamel Boulos et al., 2011). Using crowdsourced data within the context of cycling research is a growing research field to monitor ridership (Griffin and Jiao, 2015; Jestico et al., 2016) and safety (Nelson et al., 2015). The crowdsourcing website BikeMaps.org allows citizens to map the location of a cycling incident and provide details such as time of day, weather conditions, sight lines, as well as demographics like age and gender (Nelson et al., 2015). Using the popularity of GPS smartphone based apps and web applications, crowdsourced data present new information for researchers but require more insight in order to understand how to use appropriately.

Using data generated from ‘the crowd’ can cause concerns about the overall quality of the data (Barbier et al., 2012). Potential biases and overall accuracy of user submitted content can be difficult to understand and quantify (Foody et al., 2013; Jackson et al., 2013). Citizens providing data may have hidden motivations for providing information, which could affect the overall quality of data (Coleman et al., 2009). Concerns using crowdsourced data are challenging to overcome, but can be alleviated by demonstrating applications in practice. To overcome these barriers and concerns around

the quality of crowdsourced data, research is needed to provide examples of how effectively use and evaluate these datasets.

### **1.3 Research goals and objectives**

Our research goal was to incorporate crowdsourced data to advance cycling research on ridership and safety. To meet our goal, we conducted two studies on two different types of crowdsourced data: one on ridership and one on cyclist safety.

The first objective (Chapter 2) was to examine how a crowdsourced dataset from the Strava fitness app could be used to estimate ridership volumes throughout the year in Victoria, BC, Canada. To meet our objective, we compared how manual cycling counts conducted at intersections during peak commuting hours in Victoria at different times in the year compared to the number of crowdsourced cyclists counted during the same count period. Drawing on existing research on factors that influence ridership, we integrated continuous GIS covariates including topography, traffic speeds, on street parking, and time of year along with crowdsourced data for all of Victoria to create prediction maps for ridership volumes at different times of the year. We discussed how our results compared with existing literature on factors that influence ridership. We explored how our results from a case study in a mid-size North American city explain route choice between fitness app users and all cyclists. In our study, crowdsourced cyclists using fitness apps choose similar routes as commuter cyclists in urban environments. We highlighted how crowdsourced data from fitness apps are a biased sample of ridership; however, combining with other GIS covariates, crowdsourced data can provide valuable information for predicting ridership volumes in areas where no traditional data sources

are available. The importance of this contribution is significant for cycling research, safety, and planning.

Our second objective (Chapter 3) was to investigate how a crowdsourced cycling incident dataset could be used to assess safety at intersections between multiuse trails and roads. We used a cycling incident dataset from BikeMaps.org and examined reported collisions and near misses at multiuse trail and road intersections along the Galloping Goose Trail in the Capital Regional District, BC, Canada. We compared attributes of incident reports at multiuse trail-road intersections to road-road intersections to examine differences. We conducted site visits at intersections along the Galloping Goose Trail to examine infrastructure characteristics associated with incident reporting. We then modeled the relationship between the number of incidents and infrastructure characteristics at multiuse trail-road intersections. Given that standard cycling incident records (such as police and insurance reports) were limited along multiuse trails, we showed how crowdsourced cycling incident data can be used to assess safety when other datasets are limited. We provided insight into the characteristics of multiuse trail-road intersections that are associated with incident reporting. We were able to showcase how crowdsourced incident data collected from citizens can provide valuable information for assessing safety at these intersections.

#### **1.4 Research study area**

The Capital Regional District (CRD) on Vancouver Island has some of the highest cycling ridership in Canada. The percentage of commuting trips by bike is 3.20% (Capital Regional District, 2011). Given the mild climate, with temperatures rarely

reaching below zero degrees Celsius, the CRD presents a key study area for researching cycling year round. Recently proposed and substantial investments in cycling infrastructure within the CRD also highlight the importance to understanding cycling trends. In order to understand and monitor ridership trends and safety, the CRD conducts manual cycling volume counts throughout the year. However, these lack spatial and temporal detail and could be improved by integrating new data sources through crowdsourcing.

## References

- Barbier, G., Zafarani, R., Gao, H., Fung, G., & Liu, H. (2012). Maximizing benefits from crowdsourced data. *Computational and Mathematical Organization Theory*, 18(3), 257–279. doi:10.1007/s10588-012-9121-2
- Bassett, D. R., Pucher, J., Buehler, R., Thompson, D. L., & Crouter, S. E. (2008). Walking, cycling, and obesity rates in Europe, North America, and Australia. *Journal of Physical Activity & Health*, 5(6), 795–814.
- Choi, J. (2013). An Analysis of Area Type and the Availability of Alternative Transportation Services on Subjective Well-Being: Are People Happiest in Cities?.
- Coleman, D. J., Georgiadou, Y., & Labonte, J. (2009). Volunteered Geographic Information : The Nature and Motivation of Producers\*. *International Journal of Spatial Data Infrastructures Research*, 4, 332–358. doi:10.2902/1725-0463.2009.04.art16
- Cupples, J., & Ridley, E. (2008). Towards a heterogeneous environmental and cycling responsibility : sustainability fundamentalism. *Area*, 40(2), 254–264.
- De Geus, B., Vandenbulcke, G., Int Panis, L., Thomas, I., Degraeuwe, B., Cumps, E., ... Meeusen, R. (2012). A prospective cohort study on minor accidents involving commuter cyclists in Belgium. *Accident Analysis and Prevention*, 45, 683–693. doi:10.1016/j.aap.2011.09.045
- Dill, J., & McNeil, N. (2012). Four Types of Cyclists? *Transportation Research Record*, 2387(2387), 129–138. doi:10.3141/2387-15
- District, C. R. (2011). Regional Pedestrian & Cycling Masterplan. Retrieved from <https://www.crd.bc.ca/project/pedestrian-cycling-master-plan>
- Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., & Boyd, D. S. (2013). Assessing the Accuracy of Volunteered Geographic Information arising from Multiple Contributors to an Internet Based Collaborative Project. *Transactions in GIS*, 17(6), 847–860. doi:10.1111/tgis.12033
- Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211–221. doi:10.1007/s10708-007-9111-y
- Griffin, G. P., & Jiao, J. (2015). Where does Bicycling for Health Happen? Analysing Volunteered Geographic Information through Place and Plexus. *Journal of Transport & Health*, 2(2), 238–247. doi:10.1016/j.jth.2014.12.001
- Jackson, S., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., & Stefanidis, A. (2013). Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information. *ISPRS International Journal of Geo-Information*, 2(2), 507–530. doi:10.3390/ijgi2020507

- Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*, 52, 90–97. doi:10.1016/j.jtrangeo.2016.03.006
- Kamel Boulos, M. N., Resch, B., Crowley, D. N., Breslin, J. G., Sohn, G., Burtner, R., ... Chuang, K.-Y. (2011). Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International Journal of Health Geographics*, 10, 67. doi:10.1186/1476-072X-10-67
- Kanhere, S. S. (2013). Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7753 LNCS, 19–26. doi:10.1007/978-3-642-36071-8-2
- Kessler, F. (2011). Volunteered Geographic Information: A Bicycling Enthusiast Perspective. *Cartography and Geographic Information Science*, 38(3), 258–268. doi:10.1559/15230406382258
- Le Dantec, C. A., Asad, M., Misra, A., & Watkins, K. E. (2015). Planning with Crowdsourced Data : Rhetoric and Representation in Transportation Planning. *CSCW '15 Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1717–1727. doi:http://dx.doi.org/10.1145/2675133.2675212
- Nelson, T. A., Denouden, T., Jestico, B., Laberee, K., & Winters, M. (2015). BikeMaps.org: A Global Tool for Collision and Near Miss Mapping. *Frontiers in Public Health*, 3(53), 1–8. doi:10.3389/fpubh.2015.00053
- Parkin, J., Wardman, M., & Page, M. (2007). Models of perceived cycling risk and route acceptability. *Accident; Analysis and Prevention*, 39(2), 364–71. doi:10.1016/j.aap.2006.08.007
- Pucher, J., & Dijkstra, L., 2003. Promoting safe walking and cycling to improve public health: lessons from The Netherlands and Germany. *American Journal of Public Health*, 93(9), 1509–16.
- Pucher, J., & Buehler, R. (2008). Making Cycling Irresistible: Lessons from The Netherlands, Denmark and Germany. *Transport Reviews*, 28(4), 495–528. doi:10.1080/01441640701806612
- Pucher, J., Buehler, R., Bassett, D. R., & Dannenberg, A. L. (2010). Walking and cycling to health: A comparative analysis of city, state, and international data. *American Journal of Public Health*, 100(10), 1986–1992. doi:10.2105/AJPH.2009.189324
- 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. *Environmental Health : A Global Access Science Source*, 8, 47. doi:10.1186/1476-069X-8-47

- 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, 420. doi:10.1186/1471-2458-13-420
- Winters, M., & Teschke, K. (2010). Route preferences among adults in the near market for bicycling: Findings of the cycling in cities study. *American Journal of Health Promotion*, 25(1), 40–47. doi:10.4278/ajhp.081006-QUAN-236

## **2.0 MAPPING RIDERSHIP USING CROWDSOURCED CYCLING DATA**

### **2.1 Abstract**

Cycling volumes are necessary to understand what influences ridership and are essential for safety studies. Traditional methods of data collection are expensive, time consuming, and lack spatial and temporal detail. New sources have emerged as a result of crowdsourced data from fitness apps, allowing cyclists to track routes using GPS enabled cell phones. Our goal is to determine if crowdsourced data from fitness apps data can be used to quantify and map the spatial and temporal variation of ridership. Using data provided by Strava.com, we quantify how well crowdsourced fitness app data represent ridership through comparison with manual cycling counts in Victoria, British Columbia, Canada. Comparisons are made for hourly, AM and PM peak, and peak period totals that are separated by season. Using Geographic Information Systems (GIS) and a Generalized Linear Model we modelled the relationships between crowdsourced data from Strava and manual counts and predicted categories of ridership into low, medium, and high for all roadways in Victoria. Our results indicate a linear association ( $r^2$  0.40 to 0.58) between crowdsourced data volumes and manual counts, with one crowdsourced data cyclist representing 51 riders. Categorical cycling volumes were predicted and mapped using data on slope, traffic speeds, on street parking, time of year, and crowdsourced ridership with a predictive accuracy of 62%. Crowdsourced fitness data are a biased sample of ridership, however, in urban areas the high temporal and spatial resolution of data can predict categories of ridership and map spatial variation. Crowdsourced fitness apps offer a new source of data for transportation planning and can increase the spatial and temporal resolution of official count programs.

## 2.2 Introduction

Increased concern over global climate change has resulted in a growing need for sustainable transportation modes that are emission free (Chapman, 2007). These can include active transportation such as cycling and walking, which provide a number of health benefits to participants while also reducing motor vehicle congestion and greenhouse gas emissions (Cupples and Ridley, 2008; Handy et al., 2014). Among active modes, cycling has perhaps the greatest potential for growth in North America (Pucher and Dijkstra, 2003), with many cities having low cycling rates as low as 1% (Pucher and Buehler, 2008).

Data on cycling volumes support decision making and research by enabling, for example, investigation of factors that influences ridership (Griswold et al., 2011; Niemeier, 1996) and quantification of exposure when assessing cycling safety (Nelson et al., 2015). Ridership data are difficult to obtain and often limited by traditional methods of data collection (Gosse and Clarens, 2014; Nordback et al., 2013). Traditional data collection methods typically include manual counts of cyclists during peak commuting periods, which are adjusted to provide an estimate of overall ridership. While traditional counts provide an indication of overall volumes, they lack spatial detail and temporal coverage (Ryus et al., 2014). More recently cities are installing permanent count stations (Griffin et al., 2014), which provide excellent data on ridership through time but continue to lack spatial detail. In an effort to better characterize ridership, annual average daily bicycle (AADB) volumes have been utilized to apply daily and monthly adjustment factors to explain fluctuations in cyclists volumes at different periods of the year (El Esawey, 2014). Stated preference surveys have also been employed, which ask cyclists to

provide input into characteristics that are important when choosing cycling routes (Forsyth et al., 2012; Sener et al., 2009; Stinson and Bhat, 2003). The existing suite of ridership survey methods can provide insight into cycling route choice, but can be challenging to implement over broad spatial scales (Griffin and Jiao, 2015) and are expensive to repeat through time.

Through the expansion of Global Positioning Systems (GPS) new methods for collecting detailed cycling route information have emerged. GPS enabled mobile devices, such as smartphones, allow individuals to track and map their location (Broach et al., 2012; Casello and Usyukov, 2014; Hood et al., 2011; Le Dantec et al., 2015). Researchers have used GPS tracks of cyclists to quantify variables that influence route choice such as slope, distance, bicycle facility, traffic speeds, and on-street parking (Broach et al., 2012; Casello and Usyukov, 2014; Hood et al., 2011; Menghini et al., 2010). GPS technology has also led to popular use of fitness apps, where individuals can track routes, distance, and speed when exercising. Data generated through fitness apps is a form of “crowdsourced data”. Crowdsourced data allows the public to engage and provide data for a wide variety of transportation areas (Misra et al., 2014), including valuable insight into cycling route choice included in bikeability assessments (Krykewycz et al., 2012). Strava is one of the largest cycling fitness apps in the world, with global coverage and over 2.5 million GPS routes uploaded weekly (Strava, 2015). Strava is marketed to athletes for training and fitness tracking; however, any type of cyclist may use the app. One study (Griffin and Jiao, 2015) examined Strava data in Austin, Texas and found that users tended to use roads with bicycle lanes, shoulders, paths, steep slopes, and in populated places (Griffin and Jiao, 2015).

While crowdsourced fitness app data present an opportunity to collect detailed space-time ridership data, as with all crowdsourced data, there are challenges to effective use. Crowdsourced data lack the quality assurance of traditional geographic data collection measures (Goodchild and Li, 2012). Additionally, concerns around potential biases from user submitted content are difficult to quantify without comparing against reference data sources (Jackson et al., 2013). Cyclists may be using fitness apps for commuting purposes; however, this might be a secondary objective for downloading the app. While challenges exist, the volume and space-time resolution of crowdsourced data may have additional information content that can be leveraged to improve the availability of ridership data.

Our goal is to determine if crowdsourced fitness app data can be used to quantify and map the spatial and temporal variation of ridership within a city. We analyzed a crowdsourced fitness app dataset from Strava from January 1, 2013 to December 31, 2013 in Victoria, BC, Canada and according to the following objectives. First, we quantified linear correlations in space-time cycling trends between standard ridership surveys and crowdsourced data. Second, we constructed a model to predict total bicycle volumes using crowdsourced data and Geographic Information Systems (GIS) metrics associated with ridership. Third, for Victoria we mapped predicted categorical ridership volumes (low, medium and high) along individual road segments by season.

## **2.3 Materials and Methods**

### **2.3.1 Study area**

Our study area is the city of Victoria, BC Canada. Victoria has a population of approximately 80,000 residents and boasts the highest percentage of the population that cycles to work in Canada at 5.9% (Statistics Canada, 2011; Statistics Canada, 2012) (Figure 2.1). Victoria is the urban core of the wider Capital Region District with a population of approximately 375,000 (Capital Regional District, 2014). Victoria temperatures on average range between from 0°C (32°F) in the winter and 24°C (75°F) in the summer and precipitation levels range from 19mm in summer to 233mm in winter (Government of Canada, 2015). As a reflection of Victoria's climate, many cyclists ride year round; however, ridership is highest in spring and summer months. The cycling facilities in the region consist of on-street bike lanes and multi-use paths, most notably the four metre wide paved Galloping Goose Regional Trail, a 60 km trail heavily used by both commuter and recreational cyclists. During the time period of this study no separated bike-only facilities existed.

### **2.3.2 Victoria Cycling Counts**

In 2013 manual counts of cyclists were completed at 18 locations in Victoria as part of the regional bike count program. Cyclists are counted in January, May, July, and October to capture variation in seasonal cycling volumes due to changes in weather conditions. 34 days of manual counts were collected (n=6, 8, 6 and 14 in each season). Manual counts ranged from 15-534 cyclists per hour, 59-1296 during peak periods, and 143-2169 daily. Count stations consisted of two-, three- or four-leg intersections and

included major roadways with and without bike lanes, quiet residential streets, on street parking, and paved multi-use trails. Cyclists were counted on one week day (Tuesday, Wednesday, or Thursday) during peak commuting traffic periods (7-9 am and 3-6 pm).

### **2.3.3 Crowdsourced cycling data from the Strava fitness app**

We obtained a crowdsourced cycling dataset from Strava for 2013 for all of Victoria, BC. The data provided by Strava included a road network shapefile where the number of Strava users cycling on a particular roadway could be queried. There were 3,650 unique cyclists using Strava, which collectively mapped 74,679 routes in the Victoria region (Table 2.1). Ridership counts were provided for each road segment at a one-minute temporal resolution. 77% of users were male and 19% were female and 4% did not specify a gender (Table 2.1). The high spatial and temporal coverage of Strava data in Victoria allows for counts to be obtained in the same locations and time periods as those collected through manual counts. Strava cyclists at the same locations ranged from 0-20 cyclists per hour, 0-38 during peak periods, and 0-59 daily. While only a small portion of the Strava data was used to directly compare with manual counts, much more was used in creating prediction maps for volumes of cyclists at unknown locations.

### **2.3.4 Summary of explanatory variables**

Seven explanatory variables using geospatial datasets were used in this study (Table 2.2). The variables were considered based on their relevance in previous research evaluating cycling route choice.

### 2.3.5 Crowdsourced data and manual count comparisons

To compare cyclist counts between crowdsourced data and manual count data, crowdsourced data were aggregated into hourly intervals and matched to days when manual counts were conducted. PostGreSQL was used to summarize and extract crowdsourced data counts for each individual road segment in Victoria. We then aggregated these volumes for each road segment to match manual count periods between 7am-9am and 3pm-6pm. Comparisons between the two datasets were made at an hourly level, AM period (7-9am combined) and PM period (3-6pm combined), and peak period totals (AM and PM periods combined).  $R^2$  values using simple linear regression for each time period provide an indication of the strength of the relationship between manually counted cyclist volumes and crowdsourced cyclist volumes. The  $r^2$  became stronger as the time window increased: for the hourly, AM and PM periods and peak period totals the  $r^2$  was 0.40, 0.56, and 0.58, respectively. As a result, all subsequent analysis focused on the peak period total volume of riders cycling during the AM and PM count periods.

### 2.3.6 Modeling analysis

We developed a test model to examine the feasibility of predicting cycling volumes in Victoria, using crowdsourced data and other explanatory variables listed in Table 2.2. We used a Generalized Linear Model (GLM) with a Poisson distribution, as ridership data are in the form of count data (Crawley, 2005; Zuur et al., 2010). GLMs have the added benefit of being flexible in terms of model parameters, which allow for varying distributions to be fitted (Zuur et al., 2007). In the GLM we aimed to predict cycling volume for all unsampled, individual road segments in Victoria. The 34 days of

manual cycling volumes at 18 count locations were the dependent variable. Explanatory variables offered to the model were those found to be significant in previous studies. Time of year was included to correspond with manual count dates. Crowdsourced cyclist volume data (e.g., Strava) was included as an explanatory variable, as data coverage for Victoria was nearly continuous. We modeled predicted cycling volumes at a daily level to provide a broad overview of cycling volumes across AM (7am-9am) and PM (3pm-6pm) peak traffic periods for each season (January, May, July, and October). Predictions were made across the entire road and trail network in Victoria, including paved multi-use paths.

We used step-wise backward selection to remove explanatory variables not associated with volumes at a significance of  $p < 0.05$  (Crawley, 2005). Collinearity between explanatory variables was examined using Variance Inflation Factors (VIF), and those above a threshold VIF of 4 were removed from the model to reduce the effects of collinearity between explanatory variables.

### **2.3.7 Model error analysis**

Model error was evaluated using cross validation. Data were randomly partitioned into 90% and 10% subsets, where the GLM prediction was fit on the 90% subset and tested on the 10% subset 100 times. This 10% subset represents a sample of data that were not used in building the model and as such can be used to compare how well the GLM predicts cycling volumes compared to the observed volumes within this 10% portion. By conducting cross validation 100 times, each with a random 90% and 10% of data, an average error was computed to determine the percent difference between

predicted cycling volumes (using 90% subset) and observed cycling volumes (10% testing subset).

### **2.3.8 Categorical breakdowns of cycling volumes**

As the aim of the model was to predict categories of cycling volumes, we also assessed the accuracy of predictions to low, medium, and high classes. Five different classifications breakdowns of low, medium, and high cycling volumes were assessed to compare predicted volumes to observed volumes. Kappa coefficients were calculated to provide a measure of classification accuracy and a suitable classification to use for mapping (Jensen, 2005).

### **2.3.9 Mapping cycling volumes**

We created maps using the prediction model derived for Victoria cycling volumes and the selected classification levels for all road and trail segments in Victoria. Given the variation in cycling volumes during the year, we created maps for each count season to provide a visual indication of the changes in volumes throughout the year. Volumes of cyclists were grouped in low, medium, and high categories for mapping.

## **2.4 Results**

The results section below summarizes the methodological process and highlights key findings for each section. First we explain the results of the GLM analysis and variables found to be associated with cycling volumes. Second we assess the accuracy of the model based on a cross validation approach using training and testing datasets. Third, we examine the results of different categorical breakdowns used to distinguish low,

medium and high cycling volumes based on the GLM predictions. Finally, we examine the results of the prediction maps generated for each season in Victoria.

#### **2.4.1 Modeling analysis results**

Results of the GLM for predicting cycling volumes included five explanatory variables (Table 2.3): Crowdsourced data volumes, segment slope, posted speed limit, time of year, and available on street parking. By taking the exponential of the log estimate compared to the model intercept, log estimates can be transformed into cycling volume change that is represented by a one unit increase in each variable or factor level. Count locations with more crowdsourced cyclists were associated with increased manual count volumes: given the regression coefficient, an increase of one crowdsourced cyclist would correspond to 51 more cyclists at a location, all other parameters held constant. For slope, a one percent increase in slope resulted in 72 fewer cyclists. Segments with posted limits of 50 km/h and 40 km/h had lower cyclist volumes than 20 km/h, while 30 km/h road segments were higher volume. Time of year significantly affected cycling volumes with May, July, and October all resulting in increased volumes compared to January. Seasonality mattered; in May there were 703 cyclists more than January, in July 986 more, and October was similar to January. The presence of on-street parking facilities was shown to deter cyclists, where segments with on-street parking having 237 fewer cyclists compared to areas with no on-street parking. Variables not retained in the model include pavement width (p-value=0.291), population density (p-value=0.863), and bike facilities (p-value=0.884).

### **2.4.2 Modeling error analysis results**

Through cross validation, 100 model iterations using a random 90% and 10% subset of data were conducted and had an overall average model error of 38%. On average, over half of the predictions (55%) had errors of less than 30% (Figure 2.2).

### **2.4.3 Categorical breakdown of cycling volume results**

We assessed five different categorical breakdown thresholds for predicted cycling volumes using low, medium, and high classes. Results compared the predicted volumes to observed volumes using categorical breakdowns and the associated accuracy of how well predictions were made in each category (Table 2.4). Scenario 3, where low volumes of cyclist were between 0-199, medium volumes of cyclist between 200 and 999, and high volumes 1000+, had the highest predictive accuracy of all low, medium and high categories with 76%, 77%, and 85% respectively. While the Kappa coefficient was slightly lower than other scenarios, categorical predictions were conducted in subsequent analysis and are most accurate using this scenario. Based on this, the Scenario 3 threshold ranges were used for subsequent predictive model mapping.

### **2.4.4 Mapping cycling volumes results**

The predicted cycling volume maps by season are shown in Figure 2.3, using classification breakdowns of low (0-199), medium (>200-999), and high (>1000) based upon highest accuracies. May and July had overall higher volumes of cyclists on all roadways than January and October. Most roadways that had high volumes of cyclists in January and October remained high throughout the year.

## 2.5 Discussion

We assessed the contribution of crowdsourced cycling volume data, collected through the Strava cycling app, for predicting cycling volumes in Victoria, the Canadian city with the highest work commute cycling mode share (Statistics Canada, 2011). Our findings suggest that crowdsourced data may be a good proxy for estimating daily, categorical cycling volumes. Although crowdsourced cyclists represent a small portion of all cyclists, comparison with manual counts revealed a linear relationship between crowdsourced cyclists and total ridership in Victoria. The associations were strongest when ridership was aggregated to peak period totals that included both AM and PM counts of cyclists where the regression analysis accounted for 58% of the variance between the two datasets. Crowdsourced data from Strava is generally marketed as a fitness app with users in Victoria logging an average trip distance of 30 km. In urban areas, recreation and commuting riders seem to use the same routes, at least during mid-week.

Based on the results of the GLM analysis, locations with more crowdsourced cyclists were shown to predict increases in overall cycling volumes. On average, an increase of one crowdsourced cyclist represented an increase of 51 cyclists compared to the baseline volume of cyclists at a count station. The presence of riders using crowdsourced fitness apps can be an important indicator of overall cycling activity during peak weekday commuting periods. While crowdsourced riders only represent a sample of the overall cycling population, this sample can significantly improve model prediction capabilities. Future studies should investigate associations between crowdsourced cyclist

volumes and cyclists volumes on weekends, or off-peak weekday periods, when higher proportions of recreational and fitness riders are expected.

The key predictors influencing Victoria cyclists' route choice are consistent with previous research results. Steeper slopes are deterrents (Broach et al., 2012; Hood et al., 2011), and in our study a one percent increase in slope resulted in 72 fewer cyclists on average. In our analysis, we restricted the model to the most urban area, and our results may indicate that in urban locations, recreational riders and commuters use the same routes. A similar study focusing on the larger city of Austin Texas, USA which included rural areas, found that a sample of Strava cyclists preferred to cycle in areas with steeper slopes, which were thought to provide a more physical challenge (Griffin and Jiao, 2015). In scenic or more rural areas outside of Victoria the route selection of Strava riders may vary more. Traffic speeds and on street parking have both been found to deter cyclists (Hood et al., 2011; Stinson and Bhat, 2003), and this was consistent with our results. Seasonality was significantly associated with cycling volume, as has been found elsewhere (Heinen et al., 2010; Miranda-Moreno and Nosal, 2011). While spring and summer months saw increased cycling volumes attesting to more favourable weather conditions, weather conditions in Victoria are less extreme than other locations in Canada that see colder temperatures and increased snowfall.

A surprising result was that presence of bike facilities was not significant in predicting cycling volumes. This may be due to the limited number of count stations with cycling facilities. Of the 18 count stations, only two on-road locations had bicycle facilities, which were painted bike lanes on all intersection legs, and only one count station was located along an off-street multi-use path, albeit a location with three times

the cycling volume of other locations. Substantial evidence shows that off-street paths are a preferred infrastructure type (Broach et al., 2012; Heinen et al., 2010a; Winters and Teschke, 2010) and that cyclists will detour a small amount (<400m) to areas where these are available, owing to their importance in route choice (Winters et al., 2011). Studies in other areas with more manual count locations along bike facilities may yield different conclusions.

Model error highlighted the error between predicted cycling volumes and observed cycling volumes at each count station. The overall average model error was 38% based on cross validation. Error results indicated that the majority of predicted volumes (55%) had errors less than 30%. Categorical breakdowns of low, medium and high volumes of cyclists highlighted that using a range of 0-199 for low, 200-999 for medium, and 1000+ for high had the highest predictive categorical accuracy of 76%, 77%, and 85%.

Predicted cycling volume maps outlined the change in cycling volumes at different times of the year. May and July saw overall levels of ridership increase over January and October. However, volumes along major cycling routes remained high throughout the year. Prediction maps move beyond identifying individual variables that influence cycling to provide visual depictions of changes in cyclist volume across space and time. The added benefit of these maps is their ability to provide important cycling exposure data for safety studies aiming to characterize risk. By mapping cycling volumes, we provide a visual context for discussions between various stakeholders that can aid future management decisions surrounding cycling infrastructure and planning.

This was the first study to evaluate the contribution of crowdsourced data to predicting cycling volumes, and was conducted in the Canadian city with the highest proportion of cycling commuters. The investigation lays out a model for how this widespread crowdsourcing datasets can bring added value to modeling cycling volumes. The limitations of this exploratory work could be enhanced by inclusion of origin and destination data and we invite repetition of this work in areas with extensive manual count programs.

## **2.6 Limitations**

This research provides a novel approach to incorporating crowdsourced data to predict cycling volumes, but there are limitations to note. The focus of this research is on urban environments and the results are most applicable to other similar mid-size North American cities. Results may differ in large metropolitan centres or rural environments, or if Strava riders' route choices differ from general cyclists more substantially. We used all 18 manual count locations available for Victoria. Count locations were determined by the municipality. The availability of data for more count stations might strengthen model predictions. Less proximal stations would also limit any effects of spatial autocorrelation. Given the evidence on the impacts of motor vehicle traffic volumes on cycling route choice (Kang and Fricker, 2013; Sener et al., 2009), traffic volume would have been a desirable covariate for models. However, traffic volume was not available for study locations.

## 2.7 Conclusions

Understanding ridership trends and cycling route choice is a critical component of cycling research and practice, in order to inform safety, planning, and policy related to cycling. This research aimed to incorporate crowdsourced data to predict cycling ridership volumes across Victoria, BC throughout the year. Our results found that within urban environments and in mid-size North American cities, cyclists using crowdsourced fitness apps choose similar routes as commuter cyclists. In more scenic and rural environments this result could differ. Crowdsourced cycling data present a new type of data that allows for continuous spatial and temporal coverage to be incorporated with manual counts, through modelling and GIS, to predict categories of cycling volumes. Integrating the spatial and temporal detail contained within crowdsourced cycling data can provide valuable insight to supplement existing techniques for assessing cycling route choice. We welcome this work to be repeated in other settings, especially comparisons across urban and rural settings, to understand if spatial and temporal route choice trends vary by setting.

**Acknowledgements**

This research was supported by the Social Science and Humanities Research Council of Canada (SSHRC). We would like to thank the Capital Regional District, City of Victoria, and Strava.com for their assistance in data collection and ongoing support throughout the project.

Table 2.1 Victoria Sample Strava Rider Age and Gender Breakdown

Age	Male	Female
Under 25	174 (6%)	32 (5%)
25-34	591 (21%)	185 (27%)
35-44	712 (26%)	151 (22%)
45-54	527 (19%)	83 (12%)
55-64	249 (9%)	53 (8%)
65-74	80 (3%)	9 (1%)
75-84	6 (0%)	1 (0%)
85-94	0 (0%)	0 (0%)
Age not specified	460 (16%)	169 (25%)
Total	2799 (77%)	683 (19%)
Gender not specified		166 (4%)

Table 2.2 Explanatory variables considered for analysis

Description	Source	Type	Variable Category(s)	Operationalization	Relevance
Strava fitness app data	Strava.com	Shapefile	Continuous	Count of the number of cyclists using Strava on each road segment for any chosen time period in 2013.	Potential for crowdsourced data in transportation and cycling (Misra et al., 2014; Krykewycz et al., 2012).
Slope (%)	Geological Survey of Canada	Raster DEM 30m resolution	Continuous	Slope map calculated from DEM and average slope (%) attributed to each road segment.	Cyclists found to be generally deterred by areas with hills and increased slope (Broach et al., 2012).
Population Density (population per km <sup>2</sup> )	Statistics Canada Census Tracts	Polygon	Continuous	Population density given by population per km <sup>2</sup> for each census tract in Victoria. Value attributed to road segments that are located in each tract.	Denser population areas shown to have more cyclists (Winters et al., 2010).
Pavement Widths (m)	City of Victoria	Shapefile	Continuous	Roadway width from curb to curb for each road segment.	Wide roadways shown to be deterrent for cyclists (Allen-Munley and Daniel, 2006).
On-street parking	City of Victoria	Shapefile	On-street parking permitted or not	On-street parking permitted or not on each road segment.	Parked vehicles on roadways deter cyclists (Stinson and Bhat, 2003).
Posted Traffic	City of Victoria	Shapefile	20km/h,	Posted traffic speed limit attributed to each road	Motor vehicle volumes deterrent and cyclists prefer low traffic speed areas. (Hood

Speed limit	segment.	et al., 2011; Landis et al., 1997)
	30km/h, 40km/h, 50km/h	
Bike Facilities (painted bike lanes and paved multi-use trails)	Capital Regional District	Shapetfile
	Yes	No
	Bike facility refers to a painted bike lane or multi-use trail. If either was present on a road segment or trail then denoted as 'Yes' or 'No'.	Cyclists prefer to use bike facilities especially off-street pathways (Stinson and Bhat, 2003; Winters et al., 2013)

Table 2.3 Regression Estimates for GLM of cycling volumes along street segments

Variable	Category	Estimate (log)	Cycling volume change per 1 unit increase	P-value
Crowdsourced cyclist volume	Continuous	0.050	+ 51	<0.001
Segment Slope (%)	Continuous	-0.078	- 72	0.002
Posted Speed Limit (reference 20km/h)	50km/h	-1.424	-740	<0.001
	40km/h	-1.942	-834	<0.001
	30km/h	0.261	+291	0.025
Month (reference January)	May	0.543	+703	<0.001
	July	0.700	+986	<0.001
	October	0.009	+9	0.938
On Street Parking (reference none)	Yes	-0.279	-237	0.007

Table 2.4 Categorical breakdown analysis for thresholds of low, medium and high

	Range	Category	Accuracy	Percent of links in category	Kappa Coefficient
Scenario 1	0-199	L	76%	52%	0.55
	200-799	M	74%	35%	
	800+	H	65%	13%	
Scenario 2	0-299	L	90%	75%	0.63
	300-699	M	54%	10%	
	700+	H	75%	14%	
Scenario 3 <sup>1</sup>	0-199	L	76%	52%	0.59
	200-999	M	77%	39%	
	1000+	H	85%	9%	
Scenario 4	0-149	L	46%	28%	0.44
	150-599	M	75%	57%	
	600+	H	88%	14%	
Scenario 5	0-399	L	98%	78%	0.71
	400-599	M	20%	8%	
	600+	H	88%	14%	

<sup>1</sup> Scenario 3 had the highest predictive accuracy of all low, medium, and high categories and as such was used for mapping overall ridership volumes

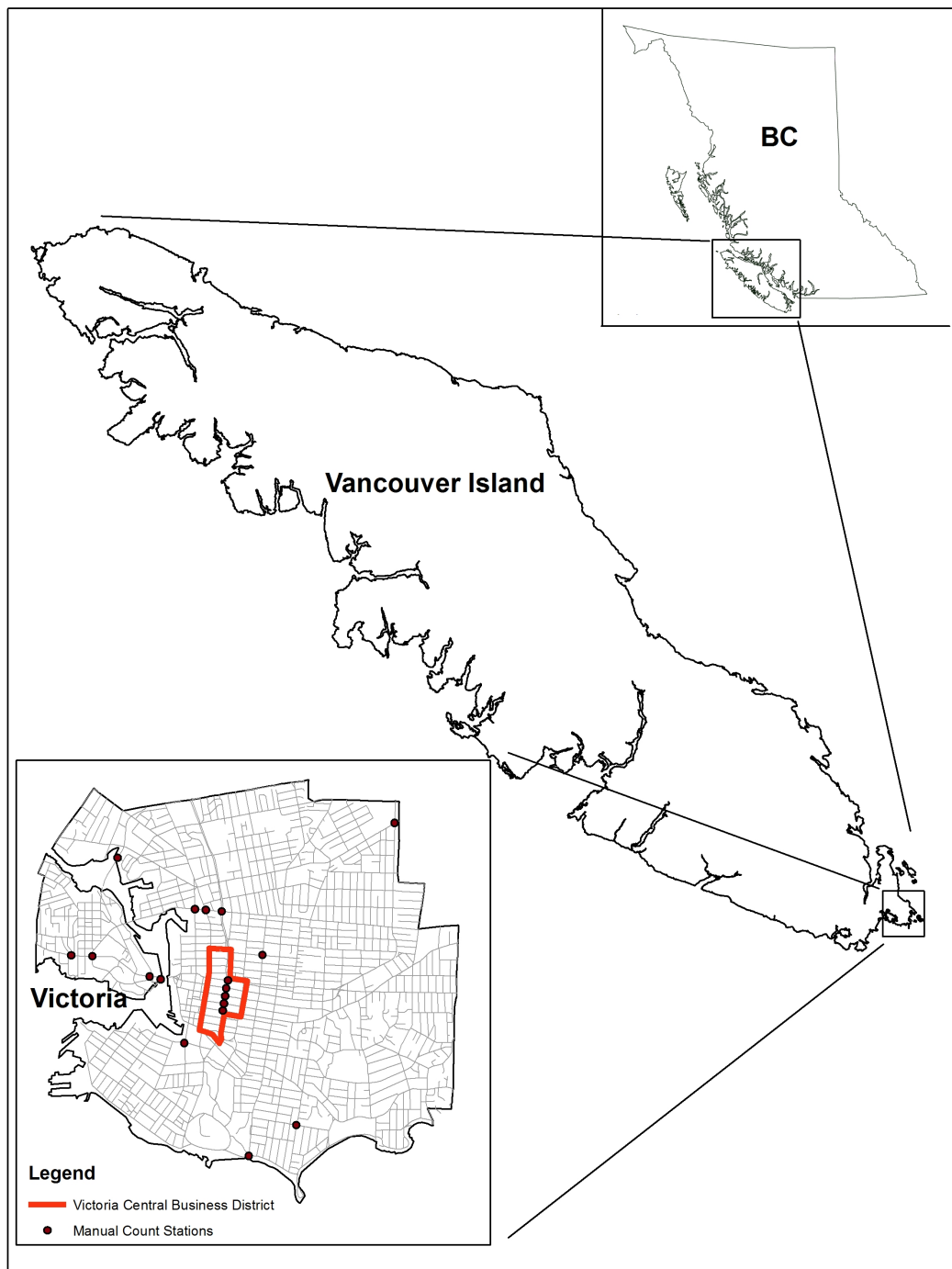


Figure 2.1 Study Area Victoria, BC Canada

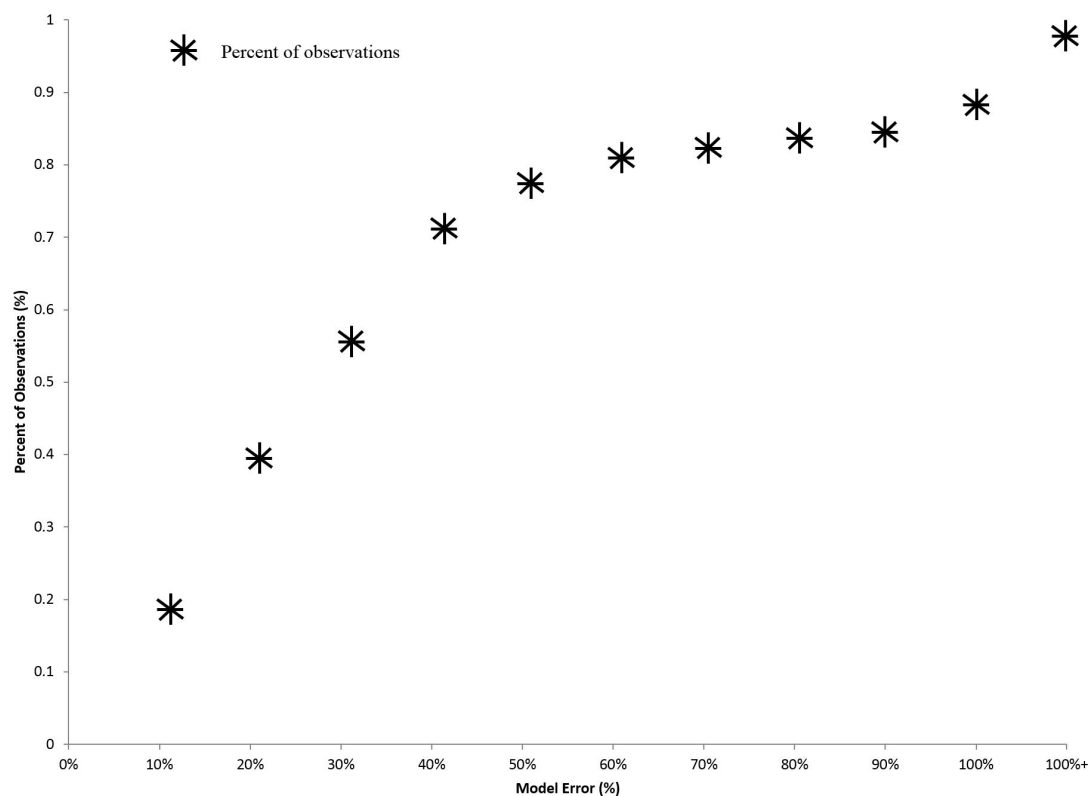


Figure 2.2 Percentage of observations predicted within an amount of model error, via cross validation. For instance, 55% of predictions had errors of less than 30%.

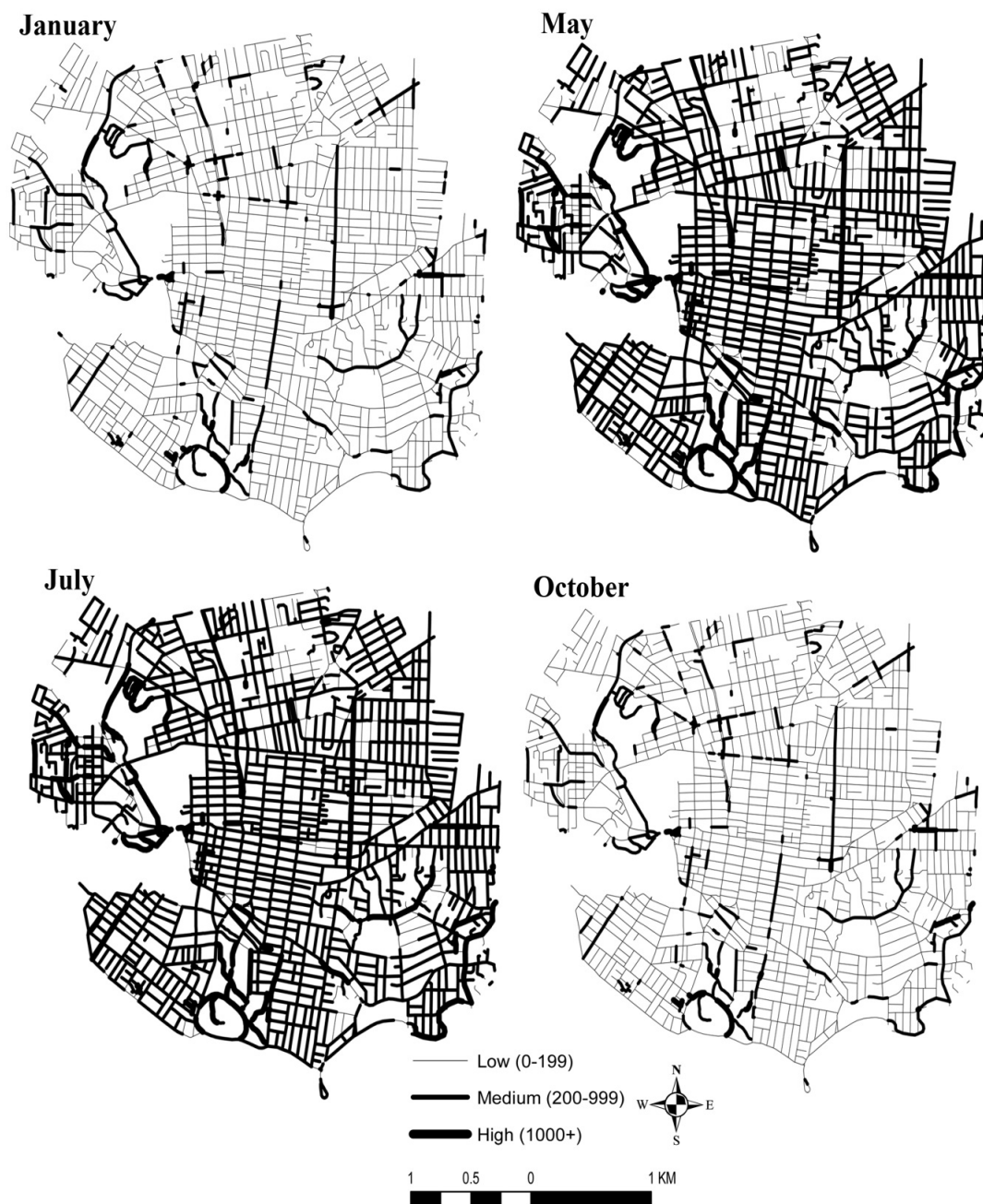


Figure 2.3 Peak period (AM and PM combined) predicted cycling volumes for Victoria, based on the GLM regression.

## References

- Allen-Munley, C., & Daniel, J. (2006). Urban Bicycle Route Safety Rating Model Application in Jersey City, New Jersey. *Journal of Transportation Engineering*, 132(6), 499–508. doi:doi 10.1061/(ASCE)0733-947X(2006)132(6):499
- Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, 46(10), 1730–1740. doi:10.1016/j.tra.2012.07.005
- Government of Canada, 2015. Canadian Climate Normals 1981-2010 Station Data. [http://climate.weather.gc.ca/climate\\_normals/results\\_1981\\_2010\\_e.html?stnID=116&radius=25&proxSearchType=city&coordsCity=48|25|123|22|Victoria&degreesNorth=&minutesNorth=&secondsNorth=&degreesWest=&minutesWest=&secondsWest=&proxSubmit=go&dCode=0](http://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?stnID=116&radius=25&proxSearchType=city&coordsCity=48|25|123|22|Victoria&degreesNorth=&minutesNorth=&secondsNorth=&degreesWest=&minutesWest=&secondsWest=&proxSubmit=go&dCode=0) (accessed 7.25.2015)
- Casello, J.M., Usyukov, V., (2014). Modeling Cyclists' Route Choice Based on GPS Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2430, 155–161. doi:10.3141/2430-16
- Chapman, L., (2007). Transport and climate change: a review. *Journal of Transport Geography*, 15(5), 354–367. doi:10.1016/j.jtrangeo.2006.11.008
- Crawley, M., (2005). *Statistics - An Introduction using R*, West Sussex, England: John Wiley & Sons, Inc.,
- Cupples, J., Ridley, E., (2008). Towards a heterogeneous environmental and cycling responsibility : sustainability fundamentalism. *Area* 40(2), 254–264.
- Capital Regional District, (2014). Estimates of Population Growth , Capital Region. <https://www.crd.bc.ca/about/data/regional-information/fact-sheets/population> (accessed 7.10.15).
- El Esawey, M., (2014). Estimation of Annual Average Daily Bicycle Traffic with Adjustment Factors. *Transportation Research Record: Journal of the Transportation Research Board*, 2443, 106–114. doi:10.3141/2443-12
- Forsyth, A., Krizek, K.J., Agrawal, A.W., Stonebraker, E., (2012). Reliability testing of the Pedestrian and Bicycling Survey (PABS) method. *Journal of Physical Activity & Health*, 9(5), 677–88.
- Goodchild, M.F., Li, L., (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, 110–120. doi:10.1016/j.spasta.2012.03.002

- Gosse, C.A., Clarens, A., (2014). Estimating Spatially and Temporally Continuous Bicycle Volumes Using Sparse Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2443, 115–122. doi:10.3141/2443-13
- Griffin, G., Nordback, K., Götschi, T., Stolz, E., Kothuri, S., (2014). Monitoring Bicyclist and Pedestrian Travel and Behavior, Current Research and Practice. Transportation Research Circular Number E-C183.
- Griffin, G.P., Jiao, J., (2015). Where does Bicycling for Health Happen? Analysing Volunteered Geographic Information through Place and Plexus. *Journal of Transport & Health*, 2(2), 238–247. doi:10.1016/j.jth.2014.12.001
- Griswold, J.B., Medury, A., Schneider, R.J., (2011). Pilot Models for Estimating Bicycle Intersection Volumes. *Transportation Research Record: Journal of the Transportation Research Board*, 2247, 1–7. doi:10.3141/2247-01
- Handy, S., Van Wee, B., Kroesen, M., (2014). Promoting Cycling for Transport: Research Needs and Challenges. *Transport Reviews*. 34(1), 4–24. doi:10.1080/01441647.2013.860204
- Heinen, E., van Wee, B., Maat, K., (2010). Commuting by Bicycle: An Overview of the Literature. *Transport Reviews*, 30(1), 59–96. doi:10.1080/01441640903187001
- Hood, J., Sall, E., Charlton, B., (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters: The International Journal of Transportation Research*, 3, 63–75. doi:10.3328/TL.2011.03.01.63-75
- Jackson, S., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., Stefanidis, A., (2013). Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information. *ISPRS International Journal of Geo-Information*, 2(2), 507–530. doi:10.3390/ijgi2020507
- Jensen, J.R., (2005). Introductory Digital Image Processing: A Remote Sensing Perspective, New Jersey, USA: Pearson Education Inc.
- Kang, L., Fricker, J.D., 2013. Bicyclist commuters' choice of on-street versus off-street route segments. *Transportation*. 40, 887–902. doi:10.1007/s11116-013-9453-x
- Krykewycz, G.R., Pollard, C., Canzoneri, N., He, E., 2012. Web-Based “Crowdsourcing” Approach to Improve Areawide “Bikeability” Scoring. *Transportation Research Record: Journal of the Transportation Research Board*, 2245, 1–7. doi:10.3141/2245-01
- Landis, B.W., Vattikuti, V.R., Brannick, M.T., (1997). Real-Time Human Perceptions: Toward a Bicycle Level of Service. *Transportation Research Record: Journal of the Transportation Research Board*, 1578, 119–126. doi:10.3141/1578-15

- Le Dantec, C.A., Asad, M., Misra, A., Watkins, K.E., (2015). Planning with Crowdsourced Data : Rhetoric and Representation in Transportation Planning. *CSCW '15 Proc. 18th ACM Conf. Comput. Support. Coop. Work Soc. Comput.* 1717–1727. <http://dx.doi.org/10.1145/2675133.2675212>
- Menghini, G., Carrasco, N., Schüssler, N., Axhausen, K.W., (2010). Route choice of cyclists in Zurich. *Transportation Research Part A: Policy and Practice*,. 44(9), 754–765. doi:10.1016/j.tra.2010.07.008
- Miranda-Moreno, L.F., Nosal, T., (2011). Weather or Not to Cycle. *Transportation Research Record: Journal of the Transportation Research Board*, 2247, 42–52. doi:10.3141/2247-06
- Misra, A., Gooze, A., Watkins, K., Asad, M., Le Dantec, C.A., (2014). Crowdsourcing and Its Application to Transportation Data Collection and Management. *Transportation Research Record: Journal of the Transportation Research Board*, 2414, 1–8. doi:10.3141/2414-01
- Nelson, T.A., Denouden, T., Jestico, B., Laberee, K., Winters, M., (2015). BikeMaps.org: A Global Tool for Collision and Near Miss Mapping. *Frontiers in Public Health*, 3(53), 1–8. doi:10.3389/fpubh.2015.00053
- Niemeier, D.A., (1996). Longitudinal Analysis of Bicycle Count Variability: Results and Modeling Implications. *Journal of Transportation Engineering*, 122(3), 200–206. doi:10.1061/(ASCE)0733-947X(1996)122:3(200)
- Nordback, K., Marshall, W.E., Janson, B.N., Stolz, E., (2013). Estimating Annual Average Daily Bicyclists. *Transportation Research Record: Journal of the Transportation Research Board*, 2339, 90–97. doi:10.3141/2339-10
- Pucher, J., Buehler, R., (2008). Making Cycling Irresistible: Lessons from The Netherlands, Denmark and Germany. *Transport Reviews*, 28(4), 495–528. doi:10.1080/01441640701806612
- Pucher, J., Dijkstra, L., (2003). Promoting safe walking and cycling to improve public health: lessons from The Netherlands and Germany. *American Journal of Public Health*, 93(9), 1509–16.
- Ryus, P., Ferguson, E., Laustsen, K.M., Scheider, R.J., Proulx, F.R., Hull, T., Miranda-Moreno, L. (2014). Guidebook on Pedestrian and Bicycle Volume Data Collection. National Cooperative Highway Research Program (NCHRP) Report 797.

- Sener, I.N., Eluru, N., Bhat, C.R., (2009). An analysis of bicycle route choice preferences in Texas, US. *Transportation* 36, 511–539. doi:10.1007/s11116-009-9201-4
- Statistics Canada, (2011). Proportion of workers commuting to work by car, truck or van, by public transit, on foot, or by bicycle, census metropolitan areas. <http://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-012-x/2011003/tbl/tb11a-eng.cfm> (accessed 7.10.15).
- Statistics Canada, (2012). Census subdivision of Victoria, CY-British Columbia. <https://www12.statcan.gc.ca/census-recensement/2011/as-sa/fogs-spg/Facts-csd-eng.cfm?LANG=Eng&GK=CSD&GC=5917034> (accessed 7.10.15).
- Stinson, M.A., Bhat, C.R., (2003). Commuter Bicyclist Route Choice: Analysis Using a Stated Preference Survey. *Transportation Research Record: Journal of the Transportation Research Board*, 1828, 107–115.
- Strava, (2015). Frequently Asked Questions. <http://metro.strava.com/faq/> (accessed 7.24.15).
- Winters, M., Brauer, M., Setton, E.M., Teschke, K., (2013). Mapping bikeability: a spatial tool to support sustainable travel. *Environment and Planning B: Planning and Design*, 40(5), 865–883. doi:10.1068/b38185
- Winters, M., Brauer, M., Setton, E.M., Teschke, K., (2010). Built environment influences on healthy transportation choices: bicycling versus driving. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 87(6), 969–93. doi:10.1007/s11524-010-9509-6
- Winters, M., Teschke, K., (2010). Route preferences among adults in the near market for bicycling: Findings of the cycling in cities study. *American Journal of Health Promotion*, 25(1), 40–47. doi:10.4278/ajhp.081006-QUAN-236
- Winters, M., Teschke, K., Grant, M., Setton, E.M., Brauer, M., (2011). How Far Out of the Way Will We Travel? *Transportation Research Record: Journal of the Transportation Research Board*, 2190, 1–10. doi:10.3141/2190-01
- Zuur, A.F., Ieno, E.N., Elphick, C.S., (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14. doi:10.1111/j.2041-210X.2009.00001.x
- Zuur, A.F., Ieno, E.N., Smith, G.M., (2007). *Statistics for Biology and Health: Analysing Ecological Data*. Springer-Verlag, New York, USA.

### **3.0 MULTIUSE TRAIL INTERSECTION SAFETY ANALYSIS: A CROWDSOURCED DATA PERSPECTIVE**

#### **3.1 Abstract**

Many cyclists and potential cyclists prefer to ride on facilities separated from motor vehicles. Multiuse trails separate cyclists from motor vehicles but are shared by other non-motorized users. Multiuse trails have been shown to have a higher risk of severe injury compared to cyclist-only facilities. However, the lack of data on less severe injuries or collisions not involving motor vehicles, which may be more common on multiuse trails, hampers research in this area. New methods for collecting incident data have emerged through crowdsourcing websites on cycling safety. We used a crowdsourced cycling incident dataset from BikeMaps.org for the Capital Regional District (CRD), BC, Canada. Our goal was to characterize the attributes of reported incidents at intersections between multiuse trails and roads and to examine infrastructure features at these intersections that are predictors of incident frequency. We extracted both collision and near miss incidents that occurred at intersections between 2005 and 2015 from BikeMaps.org. We conducted site observations at 32 intersections where a major multiuse trail intersected with roads. In our analysis, we first compared the attributes of reported incidents (collisions and near misses) at multiuse trail-road intersections to attributes of incidents at road-road intersections. Second, we used Poisson regression to model the relationship between the number of incidents (collisions and near misses) and the infrastructure characteristics at a multiuse trail-road intersections. Over the study period, 77 collisions and 192 near misses were reported at intersections in the CRD, with 14 of the collisions and 23 near misses occurring at unsignalized multiuse trail-road

intersections. Our results showed that at multiuse trail-road intersections a higher proportion of reports were collisions (38%, or 14/37 total reports), compared to reports at road-road intersections (27%, or 63/232 total reports). There was also a higher proportion of incidents that resulted in an injury at multiuse trail-road intersections than at road-road intersections (35% versus 21%). Cycling volumes, vehicle volumes, and a lack of vehicle speed reduction factors (e.g. raised crossings, speed bumps, and curb bulges) were all associated with incident frequency. Our findings indicate that by including crowdsourced cycling incident data, we can supplement traditional crash records and provide valuable evidence on the factors influencing safety at intersections between multiuse trails and roads, and more generally when cycling safety includes conflicts with diverse transportation modes.

*Keywords:* Cycling, Multiuse trail, Crowdsourc, Safety, Near miss, Collision

### 3.2 Introduction

Encouraging cycling is a goal for many communities due to its various health and environmental benefits (Handy et al., 2014). Evidence indicates separating cyclists and vehicles leads to increased ridership (Buehler and Dill, 2015). Substantial evidence indicates that cyclists prefer to cycle on off-street pathways away from motor vehicles (Heinen et al., 2010; Winters & Teschke, 2010). Separated bike only facilities have also been shown to reduce cycling injury risk compared to locations where no cycling infrastructure exists (Teschke et al., 2012).

Multiuse trails are attractive to riders and especially to novice riders (Winters and Teschke, 2010). However, their perceived safety may be overestimated (Winters et al., 2012). The diverse users on these trails include cyclists, pedestrians, rollerblades, and skateboarders. Injury risk on multiuse trails are higher than on bicycle only paths (Teschke et al., 2012; Harris et al., 2013), and injury rates are higher even when studies control for distance travelled (Moritz, 1998), or cycling volumes (De Rome et al., 2014).

Intersections are common locations for cycling incidents involving motor vehicles (Dozza and Werneke, 2014). Intersection characteristics associated with greater injury severity include motor vehicle speeds (Harris et al., 2013), cycling volumes (Miranda-Moreno et al., 2011), intersection type and design (Harris et al., 2013), route grade (Harris et al., 2013), and motor vehicle turning movements (Wang and Nihan, 2004). There has been little attention to cycling safety at intersections along multiuse trails. We found only one study, from the Netherlands, which found that at unsignalized intersections between bicycle-only paths and roadways, the most common cause of crash

was a cyclist or motorist failing to yield (Schepers et al., 2011). We found no studies that explicitly examine safety at multiuse trail-road intersections, where a mixture of road users converge.

A major challenge for assessing and conducting cycling research is lack of available data (Aultman-Hall & Kaltenecker, 1999; Zaki, Sayed, & Cheung, 2013). Generally, crash records from police reports, insurance claims, and hospital emergency records are used to assess cyclist safety, however these may only account for 30% of all cycling incidents (Tin Tin et al., 2013; De Geus et al., 2012). Incidents with anything other than motorized vehicles are vastly underreported, but may provide valuable information for understanding cyclist safety and monitoring trouble spots. Data gaps limit comprehensive safety assessments and are particularly limiting when assessing incidents between cyclists and non-motorized vehicles, such as other cyclists or pedestrians.

New methods for collecting data have become available through crowdsourcing (or Volunteered Geographic Information (VGI)), which allows citizens to collect geographic data about features in their environment (Goodchild, 2007). Crowdsourced data have been used to understand ridership trends both spatially and temporally (Griffin and Jiao, 2015; Jestico et al., 2016). A new tool, BikeMaps.org, is a source of crowdsourced data that captures cyclist reported safety incidents including both collisions and near misses. Crowdsourced data can supplement traditional crash records for a more comprehensive assessments of cyclist safety (Nelson et al., 2015). On the BikeMaps.org website and app, collisions and near misses can be mapped and attributed with descriptors to track and monitor incidents (Nelson et al., 2015). Incorporating near miss

data can allow for a proactive approach to identifying potentially dangerous areas for cyclists (Sanders, 2015).

BikeMaps.org was launched in October of 2014 and initially promoted in the Capital Regional District (CRD), which is a region that includes the City of Victoria in BC, Canada. We quickly observed that a disproportionate number of incidents were reported on multiuse trails, when compared to other areas of the same region. The Galloping Goose Trail is the primary multiuse trail and a commuter path with high volumes of cyclists throughout the year. Given that little research has been done to examine the safety of intersections between multiuse trails and roads, we use BikeMaps.org data to fill an important gap by examining the safety of multiuse trail-road intersections in a North American context.

We aim to characterize the attributes of reported incidents made at intersections between multiuse trails and roads and to also examine infrastructure features at these intersections that are predictors of incident frequency. In achieving this goal we have the following objectives. First, we investigate if incident report attributes (collision and near misses) made at multiuse trail-road intersections differed from incident report attributes made at road-road intersections in the CRD. Second, we aimed to understand what infrastructure characteristics at multiuse trail-road intersections were associated with higher incident frequency.

### **3.3 Materials**

#### **3.3.1 Study Setting**

The study area for this research is in the Victoria region, formally called the CRD on Vancouver Island, BC Canada (Figure 3.1). The CRD consists of 13 separate municipalities and has a population of approximately 375,000 residents (Capital Regional District, 2014). As a testament to a mild climate, overall cycling ridership in the CRD is among the highest in Canada at 3.2% (Capital Regional District, 2011). The CRD has set targets to increase ridership to 15% in the region by 2038 (Capital Regional District, 2011). The Galloping Goose Trail forms a backbone of the cycling network in the CRD and is a 60 km multiuse trail that is mostly paved in urban areas and heavily used by commuters and recreational cyclists throughout the year. Figure 3.2 shows two examples of intersections along the Galloping Goose Trail.

#### **3.3.2 Crowdsourced cycling incident data from BikeMaps.org**

BikeMaps.org is a website and mobile app where cyclists can map the location of a cycling crash or near miss (Nelson et al., 2015). For each incident, cyclists report time of day, weather conditions, sight lines, turning movements, injury severity (Nelson et al., 2015) (Table 3.1). A breakdown of age and gender is shown in Table 3.2. A near miss incident is self-defined by the cyclist involved, with guidance on the website suggesting a near miss when a cyclist had a close call or was almost involved in a collision. Near misses can have a significant impact on safety perceptions (Sanders, 2015) and can provide insight into potentially high risk areas. For more details on the BikeMaps.org tool refer to Nelson et al., 2015.

### **3.3.3 Multiuse trail-road intersection data collection**

Volume data are critical to account for exposure, and accurately assessing risk (Nelson et al., 2015; Reynolds et al., 2009). To quantify volumes, we conducted manual counts for cyclists and motor vehicles at 32 intersections where the Galloping Goose Trail intersected a road. We selected intersections within 15 km of Victoria as a reasonable estimate of how far cyclists would be willing to commute along the Galloping Goose Trail. As the Galloping Goose Trail extends further from downtown, the trail surface turns to gravel and becomes more of a recreational trail rather than commuter cycling trail. Trained observers completed counts for one hour during afternoon peak periods (4:15-5:15pm) with a standardized form to capture volumes for motor vehicles and cyclists. Counts provided a sample of commuting ridership during fall months in October-November 2015.

The observers also gathered data on select infrastructure characteristics at each intersection (Table 3.3), based on previous research evaluating cyclist safety.

## **3.4 Methods**

### **3.4.1 Comparing incident report attributes at multiuse trail-road intersections to road-road intersections.**

We investigated if incident report attributes made at multiuse trail-road intersections differed from incident report attributes made at road-road intersections in the CRD. We created two intersection groups, the first group consisted of incidents that occurred at intersections between the Galloping Goose Trail and roadway, and the second group consisted of incidents made at road-road intersections. Multiuse trail-road

intersections tend to have different characteristics in terms of geometry and design compared to conventional road-road intersections. For example, motor vehicles are restricted from turning onto a multiuse trail and can only proceed straight through the intersection.

We extracted BikeMaps.org collision and near miss incident reports that occurred at intersections in the CRD between 2005 and 2015; however, 89% of incidents were reported in the last two years (2013-2015). We selected incidents that were within 15 metres of an intersection, assuming that these actually occurred at the intersection. We compared the attributes from incident reports (as shown in Table 3.1) in each group. To determine if there were significant differences between the attributes in each group, we conducted a Chi-Squared test (McGrew, J. Chapman; Monroe, 2000).

### **3.4.2 Assessing intersection characteristics associated with incident reporting at multiuse trail-road intersections**

We also aimed to understand what characteristics at multiuse trail-road intersections were associated with higher incident frequency. Collision prediction studies commonly use Poisson and negative binomial regression for modeling cycling crashes (Reynolds et al., 2009; Schepers et al., 2011). We combined reported collisions and near miss incidents to get a total count of incidents at each intersection. We then used Poisson regression to model the number of incidents at each intersection against infrastructure characteristics measured during the site inspection (Table 3.3). We used backward step-wise regression, removing independent variables that were not associated with incidents

at a significance of  $p < 0.05$  (Crawley, 2005). To assess model error we calculate the mean average error (MAE) (Willmott and Matsuura, 2005).

### **3.5 Results**

#### **3.5.1 Comparing incident report attributes at multiuse trail-road intersections to road-road intersection results**

There were 37 incidents reported to BikeMaps.org at multiuse trail-road intersections and 232 incidents at road-road intersections in the CRD. Incidents reported at multiuse trail-road intersections and at road-road intersections did not differ significantly in terms of trip purpose, time of year, cyclist age, weather conditions, and daylight, but there was for topography ( $p=0.03$ ). In both types of intersections, 92% of incidents involved a motor vehicle. A higher proportion of incidents (38%) resulted in a collision at multiuse trail-road intersections compared to road-road intersections (27%). Incidents that resulted in an injury (no treatment, visited family doctor, or hospitalized) were more commonly reported at multiuse trail-road intersections compared to road-road intersections (35% versus 21%). More than two-thirds of cyclists were travelling straight through the intersection at the time of the crash for both groups. A summary of incident report attributes in each group is shown in Table 3.4.

#### **3.5.2 Assessing intersection characteristics associated with incident reporting at multiuse trail-road intersection results**

There were a total of 37 incidents reported at the 32 multiuse trail-road intersections studied (Table 3.5). The model identified three variables that influence incident reporting: cycling volumes, vehicle volumes, and speed reduction factors (Table

3.6). Model accuracy using Mean Average Error (MAE) was 41.3%. Positive or negative regression coefficients indicate the relationship between the variable and incident frequency. Cycling volumes and vehicle volumes both had positive coefficients, suggesting where there are higher volumes there are more incidents. The strength of the effect for cycling volumes was greater. For the categorical variable ‘vehicle speed reduction factors’, we compared intersections where vehicle speed reduction factors were present such as raised crossings, speed bumps, and curb bulges to intersections where these were absent. For categorical variables, we can compare the regression coefficient in one category to a baseline category (Zuur, A.F. et al., 2007). The exponential of the regression coefficient is interpreted as the relative risk (Schepers et al., 2011), and where it is less than one, there is a decreased risk of an incident occurring compared to the baseline category. The effect of vehicle speed reduction factors at an intersection decreased the chance of an incident being reported than when no vehicle speed reduction factors were present.

### **3.6 Discussion**

This study utilizes crowdsourced data on cycling incidents to evaluate safety at unsignalized multiuse trail-road intersections. Crowdsourced cycling incident reports, such as those available from BikeMaps.org, are a new data source for cycling safety research. Our results found that safety for cyclists in these locations are influenced by specific intersection characteristics.

In comparing incident report attributes at multiuse trail-road intersections to road-road intersections, we found interesting distinctions. Notably, a higher proportion of

incident reports at multiuse trail-road intersections were collisions (versus near misses), and reports at multiuse trail-road intersections had a higher proportion of incidents that resulted in an injury (35% versus 21%). Previous research has shown that injury severity on multiuse trails is higher than on other cycling facilities (Teschke et al., 2012), and our results (though non-significant) suggest this might also be true for multiuse trail-road intersections. We observed another difference between intersection types and the effect of turning movements. At multiuse trail-road intersections, only 3% of incidents involved turning movements and 70% of incidents indicated the cyclist was travelling straight through the intersection. At road-road intersections, turning movements have been identified to be a common cause of crashes (Dill, Monsere, & McNeil, 2012; Wang & Nihan, 2004); however, in our study at multiuse trail-road intersections, turning movements did not appear to cause incidents. At unsignalized priority intersections, cyclists or motor vehicles failing to yield are common causes of a crash (Schepers et al., 2011). Given that intersection priority between trail users and road users fluctuates at intersections along the Galloping Goose Trail, failure to yield by either a cyclist or motor vehicle could be potential causes of these incidents.

Three factors were associated with increased risk of a cycling incident at multiuse trail-road intersections: vehicle volumes, cycling volumes, and vehicle speed reduction factors. The effect of vehicle volumes showed that increased vehicle volumes increased incident reporting, which is consistent with previous research on cyclist-vehicle crash analysis (Turner et al., 2011). For cycling volumes, we also found that increased cycling volumes were associated with a higher number of reports. Our model estimate of 1.99 for cycling volumes indicated a positive association, however, was higher than in previous

studies. Model estimates between 0.3 and 0.9 for cycling volumes are commonly found in previous crash research, which reflects the inverse relationship between cycling volumes and safety due to the ‘safety in numbers’ effect (Elvik, 2009; Schepers et al., 2011). However, previous cycling crash studies that have found a ‘safety in numbers’ effect were most commonly conducted at road intersections in European cities and focused on crashes with motor vehicles (Elvik, 2009). Our study was in a North American city at multiuse trail-road intersections where a variety of users converge. We did not see the same safety in number phenomenon; however, not all incidents reported here involve motor vehicles. Other incidents involving cyclists, pedestrians, and other trail users are also reported, which could explain why our model estimate for cycling volumes was higher. Differences could also reflect the sample size used in this study, which was limited to a single regional trail. We found that locations with infrastructure to lower vehicles speeds (‘speed reduction factors’) such as raised crossings, speed bumps, and curb bulges were associated with decreased incident reporting. Our results were similar to other studies where reducing motor vehicle speeds at intersections reduces potential collisions and the severity of injury for cyclists (Boufous et al., 2012; Kim et al., 2007).

Lack of data is a common issue for cycling research, which explains the growing research field assessing how crowdsourced data can be used monitor safety and ridership (Griffin and Jiao, 2015; Jestico et al., 2016; Nelson et al., 2015). In our study- like in many settings – location-specific collision statistics such as police reports, insurance claims, or hospital records were not readily available. The Galloping Goose Trail crosses through nine municipalities, which made obtaining police report information from

collisions difficult. There were 10 insurance claim reports at the multiuse trail-road intersections in our study area; however, the details of these reports were limited and did not allow for an examination of potential safety factors. BikeMaps.org provides a forum for reporting incidents that do not have a traditional mechanism to do so and as such, allows for a variety of cycling environments to be studied. Reports also help provide more complete incident information including important demographics of cyclists. As well, incorporating near miss information will provide more information for assessing safety (Sanders, 2015), and in locations with limited data on crashes, such as more rural settings.

Our research provides a methodological approach to using crowdsourced cycling incident data to examine the safety of multiuse trail-road intersections, but there are limitations to consider. The nature of collecting data from citizens through crowdsourcing can create uncertainty about the quality of the data (Foody et al., 2013; Jackson et al., 2013). When possible, we suggest incorporating crowdsourced data with official reports to get the most comprehensive perspective on riding and safety patterns (Jestico et al., 2016). The focus of this research was on multiuse trail-road intersections in a North American setting where there are high volumes of cyclists throughout the year; however, in other locations multiuse trails may have much lower use and different design approaches. We welcome repetition of this work in other settings, especially with the expansion of BikeMaps.org and other crowdsourced technologies.

### **3.7 Conclusions**

Crowdsourced cycling incident data can provide valuable information when traditional cycling crash records are limited. Multiuse trails are popular locations for novice cyclists, however at intersections, potential conflicts with other road users present safety concerns. Our study used crowdsourced cycling incident data to identify variables found to affect safety at multiuse trail-road intersections in a midsized North American city. Integrating the growing dataset from crowdsourcing applications allows for repeated surveys to be conducted through time to capture more information as incidents occur. The information content can provide insight to help prioritize safety improvements at these intersections. Novel contributions to cycling incident research from crowdsourcing applications allows for a broader understanding of cyclist safety that can be monitored through time.

### **3.8 Acknowledgements**

This research was supported by the Social Science and Humanities Research Council of Canada (SSHRC) grant number 766-2015-0315. We would like to the Bunt & Associates for their support throughout all stages of the project. We would also like to thank the Capital Regional District, and numerous volunteers for their assistance in data collection and technical support.

Table 3.1 Select BikeMaps.org incident attributes that citizens are asked to detail when they map a cycling incident.

*When was the incident	What direction were you heading?
*What type of incident was it?	How were you moving?
*Were you injured?	What is your year and month of birth?
What sort of object did you collide or nearly collide with?	Please select your gender
What was the purpose of your trip?	Do you bike at least once a week?
What were the road conditions?	Were you wearing a helmet?
How were the sight lines?	Were you intoxicated?
Were there cars parked on the roadside?	What was the terrain like?
Were you using bike lights?	Where were you riding your bike?

\*Mandatory field

Table 3.2 Age and gender breakdown of incident reports at multiuse trail-road and road-road intersections

Age	Multiuse trail-road (n=37)		Road-road (n=232)	
	Male	Female	Male	Female
0 – 30	1 (3%)	1 (3%)	23 (10%)	15 (6%)
31 – 40	4 (11%)	3 (8%)	29 (13%)	11 (5%)
41 – 50	3 (8%)	2 (5%)	19 (8%)	7 (3%)
51 – 60	3 (8%)	2 (5%)	16 (7%)	4 (2%)
60+	0 (0%)	0 (0%)	1 (0%)	4 (2%)
Age not specified	1 (3%)	0 (0%)	6 (3%)	3 (1%)
Gender and age not specified	17 (46%)		94 (40%)	

Table 3.3 Data collected at each multiuse trail-road intersection

Description	Type	Procedure	Relevance
Cyclist, motor vehicle	Counts	Manual intersection counts done in October-November, 2015 between 4:15-5:15pm.	Exposure data important variable for determining crash frequency (Nelson et al., 2015) .
Sight Distances for cyclists and vehicles	Continuous	At the intersection, measured how far users can see down the trail or road.	Sufficient sight distance necessary to for motorists to determine if safe to proceed through an intersection (Harwood et al., 2000).
Road width and trail width	Continuous	Using a distance wheel, measured the width of road from curb to curb and paved section of the trail.	Wide roadways influence cyclist safety (Allen-Munley and Daniel, 2006).
Sign Inventory (Pictures)	Present/ Absent	All signs at intersection and in advance on all intersection legs	Route signage in the area for cyclists and vehicles.
Intersection Control	Bike Vehicle	Vehicles required to stop or cyclists required to stop or yield	Failure to yield common cause of crashes (Schepers et al., 2011).
Vehicle Speed reduction measures	Yes No	Presence of vehicle speed reduction measures including raised crossings, speed bumps, or curb bulges,	Vehicle speeds influence on injury severity (Cinnamon et al., 2011)
Road Grade and Trail Grade	Continuous	Measured using a rangefinder from the intersection down each leg and measured in degrees	Road grade can cause more severe injuries for cyclists (Reynolds et al., 2009).

Table 3.4 Incident report attributes at multiuse trail-road and road-road intersections in the CRD.

	Multiuse trail-road intersections		Road-road intersections		p-value
	Number of Incidents	Percent (%)	Number of Incidents	Percent (%)	
<i>Incidents</i>	37		232		
Near Miss	23	62%	169	73%	0.18
Collision	14	38%	63	27%	
<i>Injury</i>					
None	24	65%	184	79%	0.13
Injury no treatment	3	8%	14	6%	
Injury treatment required (visited family doctor, or hospitalized)	10	27%	34	15%	
<i>Incident involving</i>					
Vehicle	34	92%	213	92%	0.38
Cyclist	2	5%	5	2%	
Other	1	3%	14	6%	
<i>Movement</i>					
Heading Straight	26	70%	164	71%	0.06
Turning right/left	1	3%	32	14%	
Unknown	10	27%	36	15%	
<i>Terrain</i>					
Flat	24	65%	139	60%	**0.03
Downhill	0	0%	19	8%	
Uphill	1	3%	31	13%	
Unknown	12	32%	43	19%	

\*Statistically significant using Chi-Squared test.

Table 3.5 Summary data for the number of incidents reported at multiuse trail-road intersections

Number of intersections	Number of incidents (2005-2015)	Intersections with # incidents									
		# of incidents	0	1	2	3	4	5	6	7	8
32	37	# of intersections	22	3	1	2	1	0	1	0	2

Table 3.6 Poisson GLM regression results for infrastructure characteristics that influence incident reporting at multiuse trail-road intersections.

Variable	Number of intersections	Regression coefficient	Exponential of the regression coefficient	P-value
Intercept	32	-13.01		<0.01
Volumes				
Cyclist volume	32	1.99		<0.01
Vehicle volume	32	0.62		<0.01
Speed reduction factors				
No (reference)	21	0 (reference)	1 (reference)	
Yes	11	-1.25	0.27	0.01

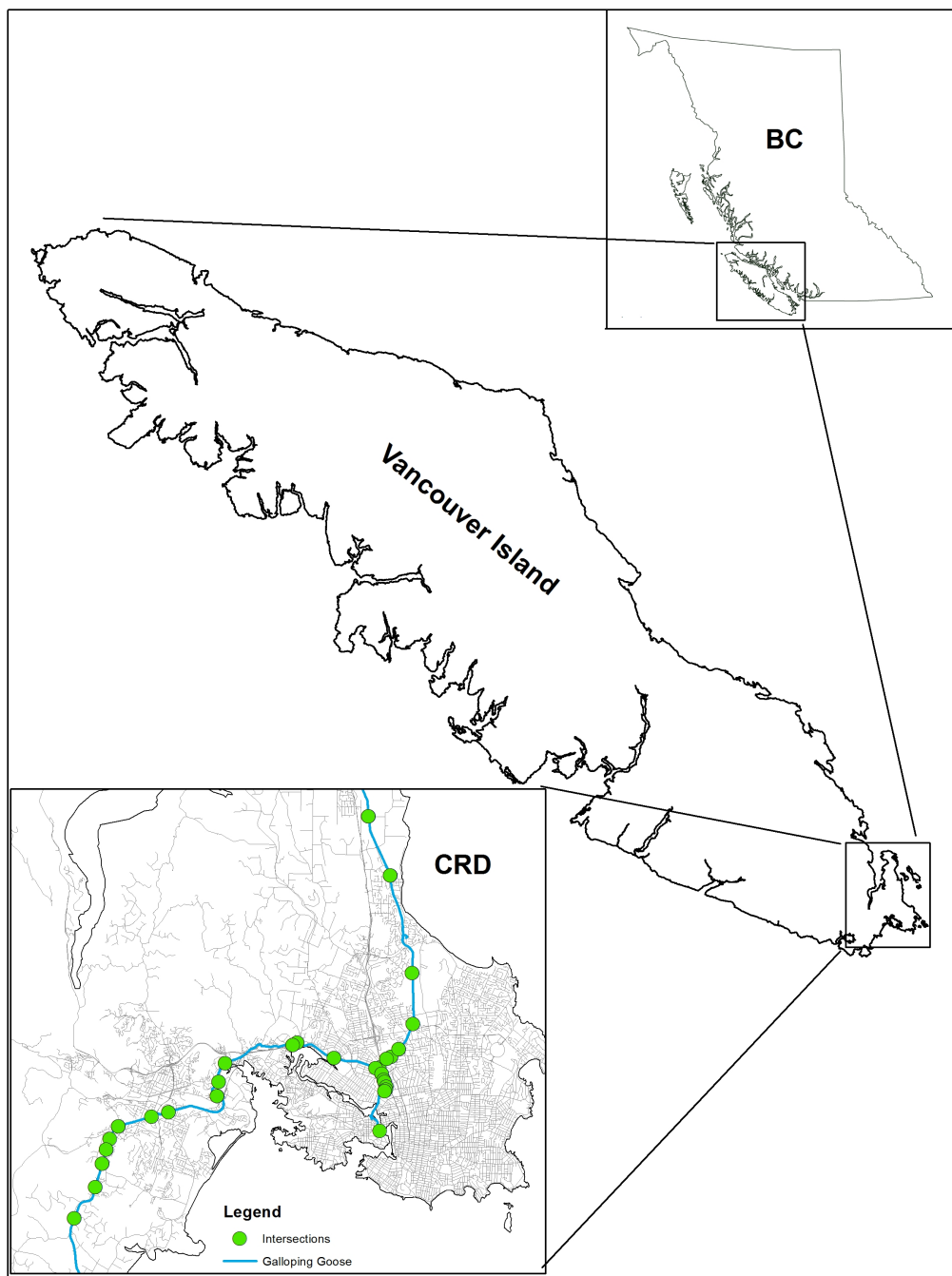


Figure 3.1 Study area in the Capital Regional District (CRD), BC, Canada



Galloping Goose/Kelvin Road intersection      Galloping Goose/Camden Avenue

Figure 3.2 Sample intersections along the Galloping Goose Trail intersections

## References

- Aultman-Hall, L., & Kaltenecker, M. G. (1999). Toronto bicycle commuter safety rates. *Accident Analysis and Prevention*, 31(6), 675–86.
- Boufous, S., De Rome, L., Senserrick, T., & Ivers, R. (2012). Risk factors for severe injury in cyclists involved in traffic crashes in Victoria, Australia. *Accident Analysis and Prevention*, 49, 404–409. doi:10.1016/j.aap.2012.03.011
- Buehler, R., & Dill, J. (2015). Bikeway Networks: A Review of Effects on Cycling. *Transport Reviews*, 1647(March), 1–19. doi:10.1080/01441647.2015.1069908
- Crawley, M. (2005). *Statistics - An Introduction using R* (1st editio.). West Sussex, England: John Wiley & Sons, Inc.
- De Geus, B., Vandenbulcke, G., Int Panis, L., Thomas, I., Degraeuwe, B., Cumps, E., ... Meeusen, R. (2012). A prospective cohort study on minor accidents involving commuter cyclists in Belgium. *Accident Analysis and Prevention*, 45, 683–693. doi:10.1016/j.aap.2011.09.045
- De Rome, L., Boufous, S., Georgeson, T., Senserrick, T., Richardson, D., & Ivers, R. (2014). Bicycle crashes in different riding environments in the Australian capital territory. *Traffic Injury Prevention*, 15(1), 81–8. doi:10.1080/15389588.2013.781591
- Dill, J., Monsere, C. M., & McNeil, N. (2012). Evaluation of bike boxes at signalized intersections. *Accident Analysis and Prevention*, 44(1), 126–134. doi:10.1016/j.aap.2010.10.030
- Capital Regional District (2011). *Regional Pedestrian & Cycling Masterplan*. Retrieved from <https://www.crd.bc.ca/project/pedestrian-cycling-master-plan>
- Capital Regional District (2014). *Estimates of Population Growth , Capital Region*. Retrieved from <https://www.crd.bc.ca/about/data/regional-information/fact-sheets/population>
- Dozza, M., & Werneke, J. (2014). Introducing naturalistic cycling data: What factors influence bicyclists' safety in the real world? *Transportation Research Part F: Traffic Psychology and Behaviour*, 24, 83–91. doi:10.1016/j.trf.2014.04.001
- Elvik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis and Prevention*, 41(4), 849–855. doi:10.1016/j.aap.2009.04.009

- Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., & Boyd, D. S. (2013). Assessing the Accuracy of Volunteered Geographic Information arising from Multiple Contributors to an Internet Based Collaborative Project. *Transactions in GIS*, 17(6), 847–860. doi:10.1111/tgis.12033
- Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211–221. doi:10.1007/s10708-007-9111-y
- Griffin, G. P., & Jiao, J. (2015). Where does Bicycling for Health Happen? Analysing Volunteered Geographic Information through Place and Plexus. *Journal of Transport & Health*, 2(2), 238–247. doi:10.1016/j.jth.2014.12.001
- Handy, S., Van Wee, B., & Kroesen, M. (2014). Promoting Cycling for Transport: Research Needs and Challenges. *Transport Reviews*, 34(1), 4–24. doi:10.1080/01441647.2013.860204
- Harris, M. A., Reynolds, C. C. O., Winters, M., Crompton, P. a, Shen, H., Chipman, M. L., ... Teschke, K. (2013). Comparing the effects of infrastructure on bicycling injury at intersections and non-intersections using a case-crossover design. *Injury Prevention : Journal of the International Society for Child and Adolescent Injury Prevention*, 19(5), 303–10. doi:10.1136/injuryprev-2012-040561
- Heinen, E., van Wee, B., & Maat, K. (2010). Commuting by Bicycle: An Overview of the Literature. *Transport Reviews*, 30(1), 59–96. doi:10.1080/01441640903187001
- Jackson, S., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., & Stefanidis, A. (2013). Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information. *ISPRS International Journal of Geo-Information*, 2(2), 507–530. doi:10.3390/ijgi2020507
- Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*, 52, 90–97. doi:10.1016/j.jtrangeo.2016.03.006
- Kim, J.-K., Kim, S., Ulfarsson, G. F., & Porrello, L. a. (2007). Bicyclist injury severities in bicycle-motor vehicle accidents. *Accident Analysis and Prevention*, 39(2), 238–51. doi:10.1016/j.aap.2006.07.002
- McGrew, J. Chapman; Monroe, C. (2000). *An Introduction to Statistical Problem Solving in Geography* (Second Edi.). USA: McGraw-Hill.
- Miranda-Moreno, L. F., Strauss, J., & Morency, P. (2011). Disaggregate Exposure Measures and Injury Frequency Models of Cyclist Safety at Signalized Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 2236, 74–82. doi:10.3141/2236-09

- Moritz, W. (1998). Adult Bicyclists in the United States: Characteristics and Riding Experience in 1996. *Transportation Research Record*, 1636(98), 1–7. doi:10.3141/1636-01
- Nelson, T. A., Denouden, T., Jestico, B., Laberee, K., & Winters, M. (2015). BikeMaps.org: A Global Tool for Collision and Near Miss Mapping. *Frontiers in Public Health*, 3(53), 1–8. doi: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. *Environmental Health : A Global Access Science Source*, 8, 47. doi:10.1186/1476-069X-8-47
- Sanders, R. L. (2015). Perceived traffic risk for cyclists: The impact of near miss and collision experiences. *Accident Analysis & Prevention*, 75, 26–34. doi:10.1016/j.aap.2014.11.004
- Schepers, J. P., Kroeze, P. a, Sweers, W., & Wüst, J. C. (2011). Road factors and bicycle-motor vehicle crashes at unsignalized priority intersections. *Accident Analysis and Prevention*, 43(3), 853–61. doi:10.1016/j.aap.2010.11.005
- Teschke, K., Harris, M. A., Reynolds, C. C., Winters, M., Babul, S., & Chipman, M. (2012). Route Infrastructure and the Risk of Injuries to Bicyclists : A Case-Crossover Study. *American Journal of Public Health*, 102(12), 2336-2343. doi:10.2105/AJPH.2012.300762
- 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, 420. doi:10.1186/1471-2458-13-420
- Turner, S., Wood, G., Hughes, T., & Singh, R. (2011). Safety Performance Functions for Bicycle Crashes in New Zealand and Australia. *Transportation Research Record: Journal of the Transportation Research Board*, 2236(2236), 66–73. doi:10.3141/2236-08
- Wang, Y., & Nihan, N. L. (2004). Estimating the risk of collisions between bicycles and motor vehicles at signalized intersections. *Accident Analysis and Prevention*, 36(3), 313–21. doi:10.1016/S0001-4575(03)00009-5
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82. doi:10.3354/cr030079
- Winters, M., Babul, S., Becker, H. J. E. H. J., Brubacher, J. R., Chipman, M., Cripton, P., ... Reynolds, C. C. O. (2012). Safe Cycling : How Do Risk Perceptions Compare With Observed Risk?. *Canadian Journal of Public Health*, 103(3), S42-S47.

Winters, M., & Teschke, K. (2010). Route preferences among adults in the near market for bicycling: Findings of the cycling in cities study. *American Journal of Health Promotion*, 25(1), 40–47. doi:10.4278/ajhp.081006-QUAN-236

Zaki, M. H., Sayed, T., & Cheung, A. (2013). Automated Collection of Cyclist Data Using Computer Vision Techniques. *Transportation Research Record: Journal of the Transportation Research Board*, 2387, 10–19. doi:10.3141/2387-02

Zuur, A.F., Ieno, E.N., Smith, G.M., (2007). *Statistics for Biology and Health: Analysing Ecological Data*. Springer-Verlag, New York, USA.

## 4.0 CONCLUSIONS

### 4.1 Discussion and conclusions

In order to encourage new ridership, it is critical to have sufficient data to understand the factors that influence safety and knowledge of where cycling carries more risk. Safety is a primary concern for potential cyclists and remains a barrier to increase ridership (Poulos et al., 2012). Citizens are generally interested in cycling, but voice concerns over safety (Dill and McNeil, 2012), and potential crashes with motor vehicles (Parkin et al., 2007). In order to properly assess safety and risk it is critical to have reliable ridership data to account for exposure, or the number of cyclists that are ‘at risk’ in a particular area (Nelson et al., 2015; Reynolds et al., 2009). Existing methods for monitoring ridership are often limited in scale and difficult to collect over time (Gosse and Clarens, 2014; Nordback et al., 2013). Monitoring safety continues to be a challenge as cycling incidents are underreported (Tin Tin et al., 2013) and limits our ability to adequately assess the factors that influence safety. Minor and less severe events are often not reported through traditional data sources such as police reports, insurance claims, or hospital records but are important for understanding safety and risk. Additionally, near miss information is not being reported but can have a profound effect on the perception of safety and provide valuable information about potential problem areas for cyclists (Sanders, 2015). Communities recognize the safety concerns around cycling and are investing in cycling infrastructure in an effort to encourage new ridership. However, in order to properly plan, monitor, and prioritize cycling initiatives, the need for reliable data to assess ridership and monitor safety is paramount (Nelson et al., 2015). Understanding how ridership changes as a result of new cycling infrastructure can

provide compelling support to the benefits of constructing cycling facilities. Existing methodologies to monitor ridership and safety can be improved to capture more information using new data sources from crowdsourcing applications to support cycling research and communities seeking to increase ridership.

In Chapter 2, we demonstrated a methodology to overcome existing challenges for collecting ridership volumes by incorporating crowdsourced fitness app data to predict cycling ridership. Using manual intersection counts, we compared the number of crowdsourced cyclists counted during AM and PM peak periods and our results indicated there was a linear relationship between the number of crowdsourced cyclists and overall cycling volumes. Using Poisson regression, we were able to model the number of manually counted cyclists at intersections against various covariates including crowdsourced cycling volumes. We created continuous prediction maps of categorical ridership volumes across Victoria using crowdsourced data and other GIS covariates that influence cycling volumes. There are many concerns about the use of crowdsourced data for planning and concerns range from the lack of representativeness of sampling (i.e., oversampling of a few demographics) to incompleteness of data. However, our work indicates that when crowdsourced data on ridership are available to build statistical models it may be possible to utilize the best components of each type of data, leveraging the resolution of crowdsourced data against the completeness of traditional counts. Our results integrating crowdsourced fitness app data with traditional counts found that in mid-size North American cities, cyclists that use crowdsourced fitness apps may choose similar routes as commuter cyclists. As well, GIS offers continuous map data on covariates that influenced ridership such as: on street parking (Stinson and Bhat, 2003),

traffic speeds (Hood et al., 2011; Landis et al., 1997), and slope (Broach et al., 2012). These data enable prediction of cycling volumes at unsampled locations. By creating prediction maps of low, medium, and high ridership, we can estimate ridership volumes across time and space in locations where no data are presently available and address pressing limitations to current data streams.

In Chapter 3, we highlighted the potential to use crowdsourced cycling incident data to assess safety at intersections between multiuse trails and roadways. Our study found that a higher proportion of collisions and injuries were reported at multiuse trail-road intersections compared to road-road intersections, suggesting that injury risk could be higher. Like in many settings, there were limited cycling incident data from traditional sources such as police reports, insurance claims, or hospital records. By using crowdsourced data, we were able to assess safety in locations that were otherwise not being tracked. By examining infrastructure at multiuse trail and roadway intersections, we were able to assess factors that are associated with incidents including cycling volumes, vehicle volumes, and vehicle speed reduction factors. We overcome a major challenge assessing safety at multiuse trail-road intersections by supplementing limited traditional data sources with new information from crowdsourced data. Citizens were able to directly provide details of a cycling incident that provided valuable information on safety at these intersections. Crowdsourced cycling incident data provide a novel contribution to cycling safety research that allows for a broader understanding of safety at multiuse trail-road intersections that can be monitored through time.

## 4.2 Research contributions

Our analysis and methodology contributes to a growing research field using crowdsourced data. Showcasing the value of crowdsourced data by providing direct examples of potential applications are powerful ways to convey the legitimacy of using these types of data. The popularity of smartphone apps continues to increase where citizens can provide a wealth of information for a variety of different applications (Kanhere, 2013). Despite the amount of information contained in crowdsourced data, there is a common hesitancy to using data generated from ‘the crowd’ for concerns about the overall accuracy of data provided (Barbier et al., 2012; Foody et al., 2013). Concerns over the accuracy of crowdsourced data are difficult to overcome unless compared to reference data sets where available (Jackson et al., 2013). However, it can be argued that local citizens that provide data through crowdsourcing may in fact provide higher quality information (Kamel Boulos et al., 2011). Local residents have an immediate connection and knowledge of their own community (Kamel Boulos et al., 2011) and those genuinely interested and supportive of a crowdsourced data project may provide high quality information. Our research showcases how to use the high spatial and temporal detail from crowdsourced data to supplement traditional data sources.

A key contribution of this thesis is providing examples of how crowdsourced cycling information can be used to predict ridership, and assess safety. Lack of data is a common issue for cycling research and our methodology allows crowdsourced data to help overcome these limitations. We provide a methodology to capture ridership volumes throughout the year using crowdsourced data and GIS covariates that communities can use to monitor ridership across space and time. Using information from crowdsourcing

tools can also be used to help plan new cycling infrastructure and provide a forum for citizens to be a part of the planning process (Le Dantec et al., 2015; Seltzer and Mahmoudi, 2013). In addition to ridership, we were also able to evaluate safety at multiuse trail and road intersections when no other cycling safety information was available. Using crowdsourced data we provide key safety information for communities when constructing multiuse trails that intersect with roadways and a methodology for monitoring ridership through time.

While not the focus of the manuscripts outlined in this thesis, our work contributes to future research with lessons learned from launching BikeMaps.org. Crowdsourcing at its basic level relies on citizens to contribute and volunteer data. In order for crowdsourcing applications to achieve success, it is imperative to reach a large enough audience in order to engage with citizens to provide data. During the initial launch of BikeMaps.org, we reached out to a number of local cycling organizations and explained the project goal and objectives and, through word of mouth, drew more visitors to the website. In addition to these grassroots organizations supporting BikeMaps.org, a critical component to the initial success was being active on social media using Twitter and Facebook. Engaging with citizens and stakeholders through social media is a powerful way to reach a large audience (Lovejoy et al., 2012) and by being active and answering questions about the project, our website views and incident reporting steadily grew. We also engaged with the community at a number of local community events in the CRD including water bottle and seat cover promotional giveaway campaigns. Water bottles and seat covers with the BikeMaps.org logo were placed on parked bicycles in Victoria and Vancouver to create awareness about the project and draw visitors. As more

visitors viewed the website, more incident reports were added and by the end of our first year active, we had reporting in 20 countries across the world. The outreach component of this thesis not only supported the data collected for Chapter 3, but also for future graduate students who will work on different aspects of cycling research from data collected on BikeMaps.org.

### **4.3 Research opportunities**

Our methodology provides a novel approach to incorporate crowdsourced data in cycling research, but further analysis would be beneficial. Our analysis contributes to a key research gap understanding how crowdsourced cycling volumes compare with overall cycling populations (Griffin and Jiao, 2015). However, repeated analysis using more count stations or comparing urban and rural settings would provide a more comprehensive understanding of how representative these datasets are in different locations (Jestico et al., 2016). It would be beneficial to understand how city scale and population size impact crowdsourcing and if there are limitations to appropriate use in cities of different sizes. Repeated analysis would also be beneficial to determine how robust crowdsourcing is over time or how data provided through crowdsourcing changes as the length of a project continues. There is little research that examines the frequency of data provided and overall sustainability through crowdsourcing over time or if there is a ‘tapering off’ effect where citizens may initially provide lots of information but become less interested over time.

Broadening research using crowdsourced data could compare results at a city level. Given the high spatial detail of crowdsourced data, there are opportunities to

conduct research at a larger scale, such as in multiple cities in a geographic region, or cities in different countries. For example, future research could compare multiuse trail and roadway intersections in multiple cities to identify if our results are similar in different cities or regions. Expanding research to all intersections (rather than multi-use trail intersections) and integrating crowdsourced incident data with traditional incident data would provide a more comprehensive representation of cycling incidents that can be used to monitor safety.

A major benefit of crowdsourced data is the ability to conduct before and after analysis for new cycling infrastructure with limited expenditures and resources. Comparing ridership and safety before new infrastructure is created can provide baseline information to compare with after construction. Changes in ridership and safety can help communities support ongoing cycling initiatives. In Victoria, a network of cycle tracks will be constructed in the next three years (City of Victoria, 2016) and our analysis using crowdsourced fitness app data helps provide baseline ridership data. Conducting a similar analysis after construction can help monitor ridership changes along these corridors that can support future projects and overall goals of increasing ridership in the CRD (CRD, 2011).

In the future, as crowdsourcing technology continues to grow (Jackson et al., 2013), there are potential opportunities to integrate ridership and safety simultaneously. Capturing ridership data along with safety information would provide an opportunity to create safe routes that could be conditioned on safety. Citizens could provide a start and end location and the application could provide the safest route that reduces cycling in areas with high risk. In short, crowdsourcing applications can fill an important gap in

ridership and safety information that provide a critical component to improving safety in an overall attempt to encourage new ridership.

## References

- Barbier, G., Zafarani, R., Gao, H., Fung, G., & Liu, H. (2012). Maximizing benefits from crowdsourced data. *Computational and Mathematical Organization Theory*, 18(3), 257–279. doi:10.1007/s10588-012-9121-2
- Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, 46(10), 1730–1740. doi:10.1016/j.tra.2012.07.005
- City of Victoria. (2016). #Biketoria. Retrieved May 17, 2016, from <http://www.victoria.ca/EN/main/community/cycling.html>
- Dill, J., & McNeil, N. (2012). Four Types of Cyclists? *Transportation Research Record*, 2387(2387), 129–138. doi:10.3141/2387-15
- District, C. R. (2011). Regional Pedestrian & Cycling Masterplan. Retrieved from <https://www.crd.bc.ca/project/pedestrian-cycling-master-plan>
- Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., & Boyd, D. S. (2013). Assessing the Accuracy of Volunteered Geographic Information arising from Multiple Contributors to an Internet Based Collaborative Project. *Transactions in GIS*, 17(6), 847–860. doi:10.1111/tgis.12033
- Gosse, C. A., & Clarens, A. (2014). Estimating Spatially and Temporally Continuous Bicycle Volumes Using Sparse Data. *Transportation Research Record: Journal of the Transportation Research Board.*, 2443, 115–122. doi:10.3141/2443-13
- Griffin, G. P., & Jiao, J. (2015). Where does Bicycling for Health Happen? Analysing Volunteered Geographic Information through Place and Plexus. *Journal of Transport & Health*, 2(2), 238–247. doi:10.1016/j.jth.2014.12.001
- Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters: The International Journal of Transportation Research*, 3, 63–75. doi:10.3328/TL.2011.03.01.63-75
- Jackson, S., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., & Stefanidis, A. (2013). Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information. *ISPRS International Journal of Geo-Information*, 2(2), 507–530. doi:10.3390/ijgi2020507
- Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*, 52, 90–97. doi:10.1016/j.jtrangeo.2016.03.006

- Kamel Boulos, M. N., Resch, B., Crowley, D. N., Breslin, J. G., Sohn, G., Burtner, R., ...  
Chuang, K.-Y. (2011). Crowdsourcing, citizen sensing and sensor web technologies  
for public and environmental health surveillance and crisis management: trends,  
OGC standards and application examples. *International Journal of Health  
Geographics*, 10, 67. doi:10.1186/1476-072X-10-67
- Kanhere, S. S. (2013). Participatory sensing: Crowdsourcing data from mobile  
smartphones in urban spaces. *Lecture Notes in Computer Science (Including  
Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in  
Bioinformatics)*, 7753 LNCS, 19–26. doi:10.1007/978-3-642-36071-8-2
- Landis, B., Vattikuti, V., & Brannick, M. (1997). Real-Time Human Perceptions: Toward  
a Bicycle Level of Service. *Transportation Research Record: Journal of the  
Transportation Research Board*, 1578(1), 119–126. doi:10.3141/1578-15
- Le Dantec, C. A., Asad, M., Misra, A., & Watkins, K. E. (2015). Planning with  
Crowdsourced Data : Rhetoric and Representation in Transportation Planning.  
*CSCW '15 Proceedings of the 18th ACM Conference on Computer Supported  
Cooperative Work & Social Computing*, 1717–1727.  
doi:http://dx.doi.org/10.1145/2675133.2675212
- Lovejoy, K., Waters, R. D., & Saxton, G. D. (2012). Engaging stakeholders through  
Twitter: How nonprofit organizations are getting more out of 140 characters or less.  
*Public Relations Review*, 38(2), 313–318. doi:10.1016/j.pubrev.2012.01.005
- Nelson, T. A., Denouden, T., Jestic, B., Laberee, K., & Winters, M. (2015).  
BikeMaps.org: A Global Tool for Collision and Near Miss Mapping. *Frontiers in  
Public Health*, 3(53), 1–8. doi:10.3389/fpubh.2015.00053
- Nordback, K., Marshall, W. E., Janson, B. N., & Stolz, E. (2013). Estimating Annual  
Average Daily Bicyclists. *Transportation Research Record: Journal of the  
Transportation Research Board.*, (2339), 90–97. doi:10.3141/2339-10
- Parkin, J., Wardman, M., & Page, M. (2007). Models of perceived cycling risk and route  
acceptability. *Accident; Analysis and Prevention*, 39(2), 364–71.  
doi:10.1016/j.aap.2006.08.007
- Poulos, R. G., Hatfield, J., Rissel, C., Grzebieta, R., & McIntosh, A. S. (2012). Exposure-  
based cycling crash, near miss and injury rates: the Safer Cycling Prospective  
Cohort Study protocol. *Injury Prevention : Journal of the International Society for  
Child and Adolescent Injury Prevention*, 18(1), e1. doi:10.1136/injuryprev-2011-  
040160
- Reynolds, C. C. O., Harris, M. A., Teschke, K., Crompton, P. a., & Winters, M. (2009). The  
impact of transportation infrastructure on bicycling injuries and crashes: a review of  
the literature. *Environmental Health : A Global Access Science Source*, 8, 47.  
doi:10.1186/1476-069X-8-47

- Sanders, R. L. (2015). Perceived traffic risk for cyclists: The impact of near miss and collision experiences. *Accident Analysis & Prevention*, 75, 26–34. doi:10.1016/j.aap.2014.11.004
- Seltzer, E., & Mahmoudi, D. (2013). Citizen Participation, Open Innovation, and Crowdsourcing: Challenges and Opportunities for Planning. *Journal of Planning Literature*, 28(1), 3–18. doi:10.1177/0885412212469112
- Stinson, M. A., & Bhat, C. R. (2003). Commuter Bicyclist Route Choice: Analysis Using a Stated Preference Survey. *Transportation Research Record: Journal of the Transportation Research Board*, 1828, 107–115.
- 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, 420. doi:10.1186/1471-2458-13-420