A Comparative Study of the Characteristics of Collisions Involving Bicycles on Frequently and Infrequently Used Bicycle Routes

Joshua James Coniglio, Jianhong (Cecilia) Xia and Mark Ryan
Department of Spatial Sciences
Curtin University
Perth, Western Australia
email:joshua.coniglio@hotmail.com, c.xia@curtin.edu.au and mark_ryan.au@outlook.com

***Abstract*— Cycling is a popular and sustainable transport mode. However, cyclists make up three percent of all road fatalities and fifteen percent of all road hospitalisations in Australia. Limited studies have been conducted to investigate the characteristics of Bicycle Motor Vehicle Crashes (BMVC) from a spatial perspective. This paper aims to compare the characteristics of BMVC on frequently and infrequently used bicycle routes within the Perth metropolitan region. It is broken down into two parts. The first part was to identify the frequently and infrequently used recreational bicycle paths based on Strava heat maps. The second task uses market segmentation with the Expectation-Maximisation (EM) algorithm to identify the major characteristics involved in BMVC on frequently and infrequently used routes. Through the findings presented in this study, the overall safety of these frequently used bicycle routes was determined.**

Keywords- freqently used bicycle route; bicycle crash characteristics; bicycle motor vehicle crashes; spatial distribution.

I. INTRODUCTION

Over the last decade, cycling has become increasingly popular not only as an environmentally friendly form of transportation, but as a healthy recreational activity undertaken individually or as part of a group. Cycling also poses more risk to severe injury in the event of a crash because the human body is unprotected against any hazardous road environments [1]. Various studies conducted by researchers into the risk factors involved in cycling related incidents along with the road and crash characteristics that contributed to the bicycle crashes. However, there is limited research comparing the characteristics of motorists, cyclists and vehicles on frequently used and unfrequently used routes.

The main objective of this study is to identify the most frequently used bicycle routes within the Perth Metropolitan area and relate these routes to BMVC to evaluate the safety of these cycling routes. The purpose behind this study is to reduce the number of incidents involving cyclists in Perth by analyzing Perth's current cycling infrastructure and major crash characteristics involved in order to provide recommendations for improvement. The rest of this paper is organized as follows. Section II describes the related studies according to the crash locations. Section III explains the methods used. Section IV addresses the results. The discussion and conclusions close the paper in Section V.

II. LITERATURE REVIEW

This review primarily focuses on identifying hazardous environments for cyclists by examining the crash locations with the highest percentage of crashes and the road and driver characteristics, which may influence these crashes. The four main crash sites include crashes in traffic, on shared paths, off-road crashes and in cycle lanes.

A. In Traffic Crash Locations

According to Meuleners, Lee, and Haworth [2], 92.5 percent of recorded cyclist crashes occurred within the Perth Metropolitan area, whilst the remaining 7.5 percent were located in rural areas [2]. This suggests that urban areas have more hazardous road environments that can threaten the safety of cyclists and increase their likelihood of suffering a severe injury. Severe injuries refer to cases where the victim is hospitalized or killed following a crash. One of the most severe crash sites are in traffic crashes, where a cyclist is travelling on the road but not within a bicycle lane. A study in the Australian Capital Territory by Rome et al. [4] in 2014 identified that 31.9 percent of cycling crashes occurred in traffic and of those, 35.4 percent involved a motor vehicle [4]. Crashes between cyclists and motor vehicles, particularly cars, pose the greatest threat to the safety of cyclists. This is supported by Stevenson et al. [5] whose research into the safety of cyclists within an Australian urban road environment identifies that a cyclist has a 3.6 greater chance of being severely injured in a crash involving a motor vehicle than any other crash type [5].

B. Crashes on Shared Paths

The next type of crash site that contributes to high crash statistics are shared paths, which account for 36.1 percent of crashes according to the results from the analysis conducted by Rome et al. [4]. Shared paths are off-road routes that are accessible by both cyclists and pedestrians. These paths however are often mistaken as bicycle only paths by cyclists thus justifying why cyclists were travelling at high speeds when they crash. Of the crashes that occurred on a shared path, 16.4 percent involved pedestrians, whilst 23.3 percent involved other cyclists. The cyclist either crashed into another road user or crashed trying to avoid a collision [4].

 ISBN: 978-1-61208-539-5

Consequently, shared paths are associated with crashes that result in serious injuries suffered by cyclists. An important safety measure that could be implemented to reduce the percentage of shared path crashes could be to recognize shared paths by installing signs near the paths. This measure would be a cost effective attempt to communicate this information to cyclists and may encourage them to slow down.

C. Off-Road Crashes

Off-road paths is a broad term used to describe several locations that contribute to over half of the cycle related crashes that occur within Western Australia. Such locations include "sidewalks, driveways, yards, cycle paths, bike trails and parking areas" [2, p 1223]. Despite 58 percent of the crashes being located off-road, cyclists injured off-road were less likely to go to hospital to receive care than those injured on the road. As a result, many bicycle related crashes are not reported to police if the injuries sustained are only minor and do not require medical treatment [2]. From these off-road crashes, 82 percent and therefore the majority occurred on footpaths, which ran parallel to a road. Footpaths can be classified as off-road paths that accounted for 16.8 percent of all bicycle crashes within the Australian Capital Territory [4]. This is a relatively low percentage though compared to the Western Australia study, reinforcing the idea that many off-road crashes are not reported and therefore cannot be included in crash statistics. A study by Wegman, Zhang, and Dijkstra [6] suggests that a vital issue in relation to cycling crash data is the underreporting of such crashes to the appropriate authorities [6].

D. On-Road Cycle Lane Crashes

The final type of crash site is on-road cycle lanes these are small lanes located on and run parallel to the road, and are dedicated to cycle traffic. They also assist in separating cyclists from motor vehicles in an attempt to improve road safety for cyclists [3]. Only 7.9 percent of crashes occur within on-road cycle lanes, which is the lowest percentage of crashes compared to any other crash site [4]. As a result, it can be debated that on-road cycle lanes are the safest way for a cyclist to travel. To back up this theory, multiple arguments can be made; firstly, on-road cycle lanes have the lowest number of crashes, secondly, it is illegal for a motor vehicle to cross over a solid line into the bicycle lane unless the line is broken. Lastly, there is a greater chance of an incident being reported to the police because the lane is located on a public road and the majority of these incidents often involve a motor vehicle [1].

III. METHODS

This section presents data collection methods, the study areas and data analysis methods.

A. The Study Area and Bicycle and Motor Vehicle Crash Data

The area of focus for this study is the Perth metropolitan area located in Western Australia (WA). The Western Australian road network was acquired from the Department of Transport (DoT) and was clipped to the study area. DoT also provided the shared paths data that consisted of shared paths, recreational shared paths and the principle shared paths. Lastly, the WA boundary was provided by Main Roads (See Figure 1).

For this study, Main Roads provided their BMVC data from 2005 to 2014. This data contained 12057 records, of which 5755 were bicycle crashes. This total of bicycle crashes was lower than expected considering the ten-year time period, suggesting that many more crashes occurred but were not reported and therefore not included in the data. This needs to be taken into consideration when using the data throughout the study. The other records were all the motorized vehicles involved in bicycle crashes. These records are therefore referred to as vehicular crashes. It is possible for multiple vehicles and one bicycle to be involved in a crash, thus the reason why there are more vehicle records than bicycles.

B. Frequently Used Bicycle Routes

Multiple acquisition approaches were considered in order to obtain data in regards to the frequently used bicycle routes. As the main focus of this study, it was crucial that this data was accurate and current to ensure that the results and any recommendations based on the results were reliable. The first approach considered was using a GPS device attached to the bicycles of multiple frequent cyclists so that different routes could be acquired. The disadvantage of this approach however was that the data could be biased based towards the route preferences of those cyclists and therefore was not a good representation of all cyclists in Perth.

The second approach was the use of existing data collected by Run and Cycling Tracking on the Social Network for Athletes (Strava). Strava is a mobile application that utilises mobile Global Positioning System (GPS) to record the routes of runners and cyclists worldwide. Using this data, Strava developed a global heat-map that displays the most highly used paths by runners and cyclists. An inquiry was sent to Strava requesting their cycling data for the Perth metropolitan area to be used for this study but no response was ever received. As a result, a third approach was ultimately taken.

The frequently used bicycle routes data was obtained by digitising the data from a road network and the shared paths data based on Strava's global heat-map. Care, time and precision was required to ensure the data was accurate in accordance with this heat-map. The shared paths data was used to obtain any paths that were not located along roads such recreational paths around the Swan River. The resulting digitized routes are therefore the most frequently used recreational cycling routes

For simplicity, these frequently used paths are referred as the popular routes, which are highlighted in the Strava heat-map and all other paths are known as the unpopular routes.

 ISBN: 978-1-61208-539-5

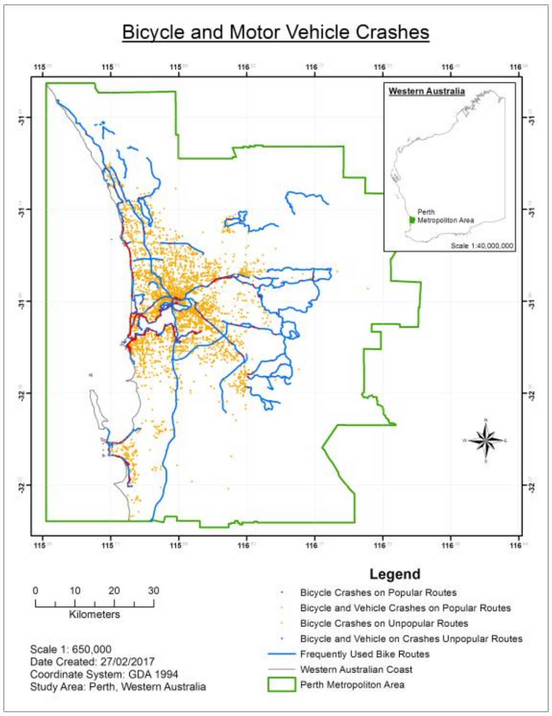

Figure 1. The study area

C. *Identification of Major Characteristics of Drivers and Vehicles of BMVC.*

A data mining approach was taken for the second part of the analysis to identify the dominant characteristics of each variable involved in BMVC on the identified popular and unpopular routes. These characteristics will in turn identify vulnerable populations that are more prone to severe crashes. Data mining uses existing information to generate new information for a particular purpose. The method chosen to implement this approach is known as market segmentation with the use of the EM algorithm. Market segmentation is the process of separating a market into smaller groups known as market segments based on the similar characteristics from the market [7]. The EM algorithm is a clustering based algorithm that identifies the dominant characteristics of objects by implementing maximum likelihood principles. For model-based clustering, this algorithm is also known as mixture modelling and can be used for both categorical and numerical data.

The crash data was broken down into four subsets so that cyclist crash characteristics on frequently used routes could be easily compared against those on infrequently used routes. Vehicle crashes were also of interest because each case that involved at least one vehicle in the crash data also involved a cyclist. As a result, characteristics of vehicles could have been contributing factors that caused the crashes to occur. For example, a drunk driver could hit a cyclist so the driver's alcohol level would be an important factor that caused the crash. The original crash data was first broken down into bicycle only crashes and vehicle crashes. Using the frequently used bicycle routes (line geometry) and the crash sites (point geometry) within a Geographic Information System (GIS), the second task was to determine an appropriate distance (in metres) to classify if a bicycle crash had occurred on a frequently used route. Three distances were tested and the most appropriate was chosen.

Firstly, a distance of 0 metres was used meaning the bicycle crashes had to intersect with the line geometry of the popular route. However, only 2 crash sites out of 5755 records actually intersected with the popular routes line because the majority of crashes were slightly offset from the line. A distance of 10 metres was then implemented but crashes that occurred on roads parallel to the popular routes (i.e., in the other lane) were also selected. Ultimately, a distance of 5 metres was chosen because it selected all crashes on popular routes without selecting crashes in the opposing lanes near the popular routes. In other words, crashes within a 5 metre proximity of the popular routes were selected and classified as having occurred on the popular route. Crashes that occurred on unpopular routes are therefore the inverse of the popular routes selection. The same process was carried out for the vehicular crashes. Table I summarises the number of records for the four subsets. The last task was to clean the data by converting numerical data to categorical data because the data provided by Main Roads used numbers to represent categories.

TABLE I. NUMBER OF RECORDS PER CRASH SUBSET

	Bicycle Crashes	**Vehicular Crashes**	**Total**
Popular Routes	591	597	1188
Unpopular Routes	5164	5705	10869
Total	5755	6302	12057

According to Xia et al. [7], there are three main steps involved to carry out this analysis. The first is to identify the spatial patterns involved. For this study there are only two patterns of interest, which are the crashes on popular routes compared against the crashes on the unpopular routes. The second step was to apply market segmentation with the EM algorithm to the variables for each crash subset using software called Weka. This generates between one to seven market segments in a table based on the characteristics inputted. In total, 10 tables were created so each crash subset had a table for each variable. Tables for the vehicle variable were not required for the bicycle crash subsets because the vehicle is a bike, thus there are no data in regards to the bike itself. The last step was to distinguish the major characteristics by interpreting the results from step two.

IV. RESULTS

Using driver and vehicle variables, two tables were created from the results of the ten crash subset tables to identify the most statistically significant characteristics involved in bicycle and vehicular crashes. In each table, the dominant market segment is bolded for both cyclist and motorized categories for each movement pattern. The value in brackets indicates the percentage of the market that belongs to that cluster. The dominant value under each characteristic in each market segment (i.e., each row) is also shaded. Characteristics with empty cells were not involved in the analysis of that crash subset because the characteristic was not deemed as relevant. Tables II to III show the results for each variable.

For crashes on the popular routes, in the first row of Table II for example, 68% of bicycle crashes on popular routes involved a male between the ages of 40 to 49 who were wearing protective gear and only minor property damage was sustained as a result of the crash. This is therefore the largest and most significant market segment for that crash subset. The other two segments were dominated by young male cyclists aged between 20-29, the majority of which were hospitalized or required medical attention. While for crashes on the unpopular routes, middle-aged male cyclists aged between 30-39 in segment 6 dominated and only minor property damage was sustained as a result of the crash. Young male cyclists aged between 10-19 in segment 5 dominated and the majority of them did not wear protection. We did not find any noteworthy characteristics for motorists, except minor injuries were dominant for all segments of the motorists.

In Table III, 73% of the vehicular crashes on popular routes involved a car with a four-cylinder engine manufactured between 2000 and 2009 while the remaining 27% of the market had a six-cylinder engine. Segment 2 contains 73% of the vehicular crashes on popular routes and therefore is the dominant group. For vehicular crashes on unpopular routes, segment 1 is the most dominant with 60% of the market falling within this segment. The results from this segment are the same to that of segment two from the popular routes movement pattern. However, segments 4 and 5 are worth noting because the dominant vehicle in segment 4 is a station wagon rather than a car and segment 5 is the same as segment one for the popular routes. Together, segments four and five equal to just over a quarter of the market.

V. DISCUSSION AND CONCLUSIONS

The major characteristics of cyclists, drivers and motor vehicles were discovered. One vulnerable population were teenagers involved in severe bicycle crashes on the unpopular routes without wearing helmets. The age group of male drivers varied from young adults to middle aged who wore protective equipment such as seatbelts and were usually involved in incidences resulting in minor property damage. In incidences involving such drivers however, inattention and fatigue were the defining characteristics that would have caused the crash. Fatigue not only affects a drivers' reaction time but leads to inattention. This threatens the safety of cyclists regardless if they are within an on-road cycle lane because cyclists on the road rely on other road users to overtake them safely. Surprisingly, the injuries sustained by crashes involving fatigue and inattention are minor but could have been much worse, especially in the case of a vehicle colliding with a cyclist.

Of the major vehicle characteristics, modern cars with four cylinder engines were mostly involved in crashes and could be related to the high number of crashes caused by young adults not paying attention while driving. Consequently, there is strong causation that the young drivers were using their phones whilst driving and therefore not paying attention to the road as it is easier to use a mobile device while driving a small automatic vehicle.

In conclusion, the most frequently used recreational bicycle routes were identified using Strava's global heat-map. The major characteristics involved in bicycle-related crashes on and off the popular routes were also identified. The market segmentation method with the use of the Expectation-Maximisation algorithm successfully divided the crash data for each variable into groups that share common characteristics. These groups were examined in depth in comparison to each other and several justifiable theories were developed as a result.

Overall, the popular recreational bicycle routes were found to be safe despite the high traffic flow and urgent need of maintenance. Not many crashes were found to have occurred on the popular routes compared to the unpopular routes, suggesting they are much safer and therefore should be utilised by regular recreational cyclists. This in turn, may reduce the number of cyclists involved in dangerous road incidents by diminishing the risk. The major limitation of the study is the methodology in defining popular and unpopular routes. Strava is mainly for recreation purpose. Many bicycle trips, with other purposes, such as work, haven't been considered. In the future, other social media tool, such as Bikemap, can be used for understanding bike travel behavior in Perth in a more holistic way.

TABLE II. DRIVERS' CHARACTERISTICS

Market Segments of Significant Crash Patterns for Driver Characteristics

Movement Patterns	Cyclist vs Motorist	Market Segments	Sex		Age								
			M	F	1 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80 to 89
Popular Routes	Cyclist	**1 (0.68)**	365.1	40.2	30.3	4.6	29.3	63.5	216.0	48.5	15.4	2.9	1.8
		2 (0.08)	38.6	12.8	2.1	17.4	19.3	2.4	9.0	2.0	2.2	2.8	1.1
		3 (0.23)	88.3	52.0	21.6	5.0	29.5	28.1	21.9	23.4	10.4	6.3	1.1
	Motorist	**1 (1)**	389.0	210.0	40.0	34.0	282.0	61.0	58.0	57.0	46.0	23.0	7.0
Unpopular Routes	Cyclist	1 (0.11)	483.4	80.6	84.0	46.7	76.9	120.0	128.2	69.2	31.8	13.0	1.3
		2 (0.25)	1238.2	64.2	140.7	119.8	274.9	96.6	389.6	199.3	64.0	21.3	3.2
		3 (0.09)	71.9	416.8	44.3	48.2	144.4	116.0	72.9	46.9	18.7	3.1	1.1
		4 (0.07)	535.2	19.3	36.1	83.4	35.5	272.9	36.5	60.8	19.5	8.0	8.8
		5 (0.07)	331.4	42.9	52.2	151.3	87.8	34.6	33.9	11.3	5.3	2.0	2.9
		6 (0.37)	1684.8	207.2	48.7	111.6	70.5	1595.9	32.0	24.5	9.7	4.6	1.7
	Motorist	**1 (0.99)**	3427.9	2230.0	545.0	309.0	799.0	678.0	2319.9	526.0	303.0	138.0	44.0
		2 (0.01)	50.1	1.0	1.0	1.0	1.0	1.0	50.1	1.0	1.0	1.0	1.0

Movement Patterns	Cyclist vs Motorist	Market Segments	Protection		Severity					Alcohol	Inattention		Fatigue		Purpose of Travel	
			Worn	Not Worn	Fatal	Hospitolised	Medical	PDO Major	PDO Minor	Level	Yes	No	Yes	No	Private	Business
Popular Routes	Cyclist	**1 (0.68)**	387.7	17.6	1.2	77.6	113.0	41.8	174.6							
		2 (0.08)	28.8	22.6	1.5	26.9	8.3	2.2	15.5							
		3 (0.23)	129.5	10.8	1.4	26.5	97.7	5.0	12.8							
	Motorist	**1 (1)**	587.0	12.0	2.0	111.0	214.0	54.0	221.0	0.0	583.0	16.0	597.0	2.0	564.0	35.0
Unpopular Routes	Cyclist	1 (0.11)	545.9	18.1	3.1	408.5	70.2	19.4	65.8							
		2 (0.25)	1294.1	8.4	2.2	80.1	583.4	89.6	550.2							
		3 (0.09)	466.5	22.3	2.6	82.6	252.8	8.3	145.4							
		4 (0.07)	471.6	82.9	6.8	45.7	438.9	28.2	37.9							
		5 (0.07)	32.9	341.4	14.1	172.5	122.1	11.6	57.0							
		6 (0.37)	1611.0	281.0	2.2	80.7	131.5	285.9	1394.7							
	Motorist	**1 (0.99)**	5593.9	64.0	22.0	866.0	1695.0	505.0	2573.0	0.0	5519.9	138.0	5650.9	7.0	5280.9	376.0
		2 (0.01)	50.1	1.0	4.0	11.0	18.0	3.0	18.0	0.9	48.1	3.0	50.1	1.0	38.1	1.0

TABLE III. VEHICLES' CHARACTERISTICS

Market Segments of Significant Crash Patterns for Vehicle Characteristics

Movement Patterns	Bicycle vs Vehicle	Market Segments	Year Manufacture						Number Cylinder			U Type			
			1960 to 1969	1970 to 1979	1980 to 1989	1990 to 1999	2000 to 2009	2010 to 2014	4	6	8	Car	Utility	4WD	Station Wagon
Popular Routes	Vehicle	1 (0.27)	3.0	2.0	9.0	38.0	107.0	9.0	1.0	128.0	21.0	98.0	8.0	15.0	22.0
		2 (0.73)	1.0	2.0	8.0	60.0	307.0	63.0	436.0	1.0	1.0	300.0	41.0	11.0	53.0
Unpopular Routes	Vehicle	**1 (0.6)**	1.8	3.0	120.4	543.3	2734.5	44.0	3380.4	29.8	12.3	2766.6	233.0	81.3	41.4
		2 (0.07)	1.7	3.9	28.4	306.7	18.6	28.7	8.0	345.3	23.7	275.2	20.6	41.8	14.7
		3 (0)	1.0	1.1	2.9	1.0	17.1	11.1	2.0	1.3	1.2	1.5	1.3	1.5	1.1
		4 (0.12)	2.1	1.2	7.9	77.1	519.6	100.7	418.9	254.7	14.4	25.7	18.0	24.8	597.0
		5 (0.14)	2.4	6.7	13.7	33.7	708.4	26.1	24.9	665.9	78.3	559.4	99.0	56.6	26.2
		6 (0)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		7 (0.06)	2.0	2.2	15.7	16.2	18.8	317.3	330.8	15.0	7.0	277.6	30.2	13.0	35.6

REFERENCES

[1] S. Boufous, L. D. Rome, T. Senerrick, and S. Ivers, "Risk factors for severe injury in cyclists involved in traffic crashes in Victoria, Australia," Accident Analysis and Prevention vol. 49, pp. 404-409, 2012, doi: 10.1016/j.aap.2012.03.011.

[2] L. B. Meuleners, A. H. Lee, and C. Haworth, "Road environment, crash type and hospitalisation of bicyclists and motorcyclists presented to emergency departments in Western Australia," Accident Analysis and Prevention vol. 39, pp.1222-1225, 2007, doi: 10.1016/j.aap.2007.03.006

[3] J. Parkin, and C. Meyers, "The effect of cycle lanes on the proximity between motor traffic and cycle traffic," Accident Analysis and Prevention, vol. 42, pp. 159-165, 2010, doi: 10.1016/j.aap.2009.07.018.

[4] L. D. Rome, S. Boufous, T. Georgeson, T. Senserrick, D. Richardson, and R. Ivers, "Bicycle Crashes in Different Riding Environments in the Australian Capital Territory," Traffic Injury Prevention, vol. 15, pp. 81-88, 2014, doi:10.1080/15389588.2013.781591.

[5] M. Stevenson, M. Johnson, J. Oxley, L. Meuleners, B. Gabbe, and G. Rose, "Safer cycling in the urban road environment: study approach and protocols guiding an Australian study," Journal of the International Society for Child and Adolescent Injury Prevention, vol. 21, pp. 1-5, 2014, doi:10.1136/injuryprev-2014- 041287.

[6] F. Wegman, F. Zhang, and A. Dijkstra, "How to make more cycling good for road safety?" Accident Analysis and Prevention, vol. 44, pp. 19-29, 2012, doi:10.1016/j.aap.2010.11.010.

[7] J. Xia, F. H. Evans, K. Spilsbury, V. Ciesielski, C. Arrowsmith, and G. Wright, "Market segments based on the dominant movement patterns of tourists," Tourism Management, vol.31, pp. 464-469, 2010, doi:10.1016/j.tourman.2009.04.013