



OPEN ACCESS

Safety in numbers for cyclists beyond national-level and city-level data: a study on the non-linearity of risk within the city of Hong Kong

Shenjun Yao,¹ Becky P Y Loo²

¹Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University, Shanghai, China
²Department of Geography, The University of Hong Kong, Hong Kong, Hong Kong

Correspondence to
Professor Becky P Y Loo,
Department of Geography,
The University of Hong Kong,
Jockey Club Tower, Pokfulam,
Hong Kong, Hong Kong;
bpyloo@hku.hk

Received 18 January 2016
Revised 23 May 2016
Accepted 24 May 2016

ABSTRACT

Objective This paper examines the relationship between bicycle collisions and the amount of cycling at the local level. Most previous research has focused on national and city comparisons, little is known about differences within a city (the mesoscale).

Methods This study mainly used three types of data sets relating to bicycle collisions, use of bicycles and local neighbourhood characteristics in Hong Kong. In particular, bicycle usage, measured as bicycle-kilometres travelled, was estimated from travel surveys following the activity-based approach. Negative binomial regression models were established to model the relationship between the amount of cycling and the occurrence of bicycle collisions at the spatial scale of the Tertiary Planning Unit, which is the smallest planning unit of the city.

Results The numbers of bicycle collisions went up with the increasing use of bicycles, but the increase in the number of collisions in a given community was less than a linear proportion of the bicycle flow. When other local neighbourhood variables are controlled, the amount of cycling is a statistically significant variable in accounting for the number of collisions.

Conclusions Even in a highly motorised city where bicycles are a minor transport mode, cyclists are less likely to be involved in road collisions in communities with higher cycling volume. Since cycling activities are likely to vary within a city, a more local-based approach in promoting cycling is needed. In particular, the higher safety risks in neighbourhoods of low bicycle usage, especially at an initial stage of promoting cycling, need to be addressed properly.

INTRODUCTION

The increasing concern about global warming and environmental degradation has highlighted the importance of environmentally sustainable transport.¹ Since cycling has great potential in reducing carbon emissions and noise levels when compared with travelling by car, it is widely considered as one of the most environmentally sustainable transport modes.²⁻³ As an active mode of transport, cycling is also regarded as an excellent way to benefit public health as a form of vigorous physical activities.⁴ Nonetheless, cyclists are more vulnerable to traffic injuries than other road users such as car drivers because they are less likely to be protected by their vehicles. As the benefits of cycling can be outweighed by problems of safety and the lack of an adequate infrastructure,⁵ efforts have been paid to making bicycle use safer.

In dealing with the cycling safety issue, one needs to figure out what ‘safety’ means before performing any analysis. Does the fact that there are no cyclists injured on a road necessarily mean that the road is safe for cycling? The reason for no cycling casualties can simply be that no one is cycling there. In this light, cycling safety is better measured as risk, where cycling exposure or the volume of cycling on the road is properly considered. In recent decades, a number of studies suggested a phenomenon called ‘Safety in Numbers (SIN)’ for cyclists.⁶⁻¹³ For instance, Jacobsen⁷ concluded that a motorist was less likely to collide with a cyclist or pedestrian if more people cycled and walked. Focusing on single-bicycle crashes, Schepers⁸ found that the risk of severe injury was also lower in municipalities with a high amount of cycling. Vandembulcke *et al*⁹ confirmed that high proportions of commuter cyclists are correlated with low risks of becoming a casualty at a regional level within Belgium.

Most previous studies on SIN for cyclists, however, performed the analysis only at the macro-levels such as cities, regions and countries,⁶⁻¹⁰ or microlevels such as individual intersections and road segments.¹¹⁻¹³ Few research efforts have been dedicated to the investigation at a zonal or neighbourhood level within a city. To help road safety administrations to gain a better understanding of the safety effects of increasing cycling at different spatial scales, it is necessary to focus on bicycle collisions within a city at a neighbourhood level and examine the relationship between the amount of cycling and the likelihood of bicycle collisions.

Regardless of the different approaches used in analysing bicycle collision risks, exposure is an essential factor that should be taken into account in risk modelling.¹⁴ In the literature, there is no consensus on the best way to measure bicycle exposures. Prevailing exposure measures include time spent on cycling, number of bicycles passing the crossings, number of cycling trips taken and distance travelled calculated from the number of passages or from the mapping of cycling routes by survey respondents based on questionnaires.¹³⁻¹⁵⁻¹⁸ Recently, Global Positioning System (GPS) data have enabled researchers to gain accurate and precise cycling exposure data, but they are very costly and hence have hardly been used for obtaining cycling flow in the whole territory of a city. Although advanced information technology, notably the ‘big data’, also has great potential for deriving cycling use within a city from various types of data, it is currently not capable of

To cite: Yao S, Loo BPY. *Inj Prev* Published Online First: [please include Day Month Year] doi:10.1136/injuryprev-2016-041964

providing a territory-wide picture of bicycle flow due to limited empirical data availability. In the absence of consistent cycling exposure data within a city, cycling safety at the local level cannot be modelled appropriately. In this study, an activity-based approach is introduced to measure cycling exposure at the local level. This method relies on an individual's travel and/or activity diary,^{19–21} which allows us to map daily trajectory of each person. Based on the travel information, routes for a specific transport mode are extracted for measuring exposure. The method was first proposed for the measurement of pedestrian exposures^{19–21} and has, to the authors' best knowledge, not been applied to the study of SIN or the cycling safety research. Once again, SIN has been systematically examined at international, national, and regional levels, but it has never been investigated across all communities within the same city. Will SIN for cyclists apply at the local level across the whole territory of a city? Answer to this question is particularly important for highly motorised cities, where bicycles are still a minor transport mode and that the safety risk of cyclists is notably higher both in eyes of other road users and the road safety administrations. Therefore, this research is of theoretical significance in studying SIN at the mesoscale and of practical significance in informing policy makers by emphasising the importance of local context.

In relation, the objectives of this study are (1) to examine SIN for cyclists within a city at the local level and (2) to illustrate the method and value of extracting cycling exposure data from travel surveys following an activity-based approach at an individual level. Hong Kong is chosen as the study area, where 10.4% of the households had bicycles available for use.²² Bicycle activities, however, vary substantially over the city. The proportion of bicycle usage is much higher in the new towns and New Territories than the dense urban area.^{22–24} As the local environment such as topography and cycling facilities can vary significantly within a city, it is worth exploring SIN across different neighbourhoods.

The following section will introduce the research methodology. Following the data description, the way in which cycling exposure is calculated will be presented in details. Next, negative binomial regression models for modelling bicycle collisions (ie, collisions involving bicycles) will be introduced. The model results will then be presented and discussed, followed by conclusions and further research directions.

METHODOLOGY

Data sources

The bicycle collision data are derived from the Traffic Road Accident Database System collected by the Hong Kong Police every year. Only traffic collisions with injury were recorded in the database. In Hong Kong, the degree of collision underreporting was comparable with other developed countries.²⁵ Property damage-only bicycle crashes are not recorded in the database. Many of them are single-bicycle crashes, that is, not involving any other bicycle or vehicles. An international overview (including Hong Kong) of the importance of single-bicycle crashes has been given by Schepers *et al.*²⁶ With geographical coordinates recorded in the database, the collisions are first plotted onto a map and then geovalidated by a geovalidation procedure.²⁷ As rare events, traffic collisions happening at a certain location may be due to randomness. To ensure a more representative bicycle collision samples, this study pools data for the study periods of 3 years rather than 1 year. To examine whether SIN for cyclists can be identified consistently over time, bicycle collisions happening in two periods (the period from 2001 to 2003 and the period from 2010 to 2012) are chosen for the analysis.

The principal database for constructing the activity-based cycling exposure factors is the Travel Characteristics Survey collected by the Transport Department of Hong Kong. It provides information on the characteristics of daily activity–travel patterns of respondents during weekdays. The respondents were asked to recall all types of activities they took and transport modes by which they travelled to fulfil activities on the preceding non-public-holiday weekday. Based on the vehicular counts on the roads, motorised trips have been corrected for the under-reporting due to the recall gap.^{22–23} However, as there are no territory-wide walking or bicycle counts, these travel survey data cannot be corrected for the under-reporting or recall gap in the same manner as the motorised traffic. The study relies on two Travel Characteristics Survey data sets collected in 2002 and 2011. The former was conducted between 10 September and 21 December 2002 with a total of 30 005 households (1.4% of the population) and the latter was undertaken between 5 September 2011 and 18 January 2012 with totally 35 401 households (1.5% of the population) successfully enumerated.^{22–23} They represent the latest travel characteristics data available in the city. Moreover, Hong Kong is having subtropical climate that the average monthly temperatures from September to December were 17°C–28°C. Hence, the weather of the survey periods is suitable for cycling.

In this study, bicycle collisions are analysed at the Tertiary Planning Unit (TPU) level, which is the smallest unit for planning purpose in Hong Kong. This research uses TPU maps provided by the Hong Kong Planning Department in 2001 and 2011. The descriptive statistics on the size of TPUs are shown in table 1. Figure 1 delineates the boundary and population density of the TPU system in a year. Bicycle collisions are analysed at the TPU level because the TPU system is also used in the city's population census conducted every 10 years. Using TPU as our spatial unit of analysis, it is possible to consider other socio-economic influence in our model.

Measurement of cycling exposure

In general, activity-based measures can be categorised into two types, namely space–time path (STP) and potential path tree methods.^{19–20–28} Although it is found that both approaches are useful exposure estimates, the former is superior because it measures exposure in a more efficient manner.¹⁹ In this light, this study estimates the cycling exposure measured as bicycle distance travelled by following the STP method.

The STP approach relies on the shortest path between an origin and a destination of a trip recorded in a travel and/or activity diary. Figure 2 is a flow chart illustrating the major procedures of deriving cycling STPs from the travel survey databases. Cycling-only trips and cycling trip legs are first identified from the database. The detailed locational information on origins (O) and destinations (D) is reported at the level of street

Table 1 Descriptive statistics on tertiary planning unit size in terms of area and population

	N	Mean	Median	SD	Minimum	Maximum
2001						
Area (km ²)	282	3.97	1.83	5.48	0.06	32.0
Population	282	23 780	7520	36 650	0	203 620
2011						
Area (km ²)	289	3.84	1.81	5.35	0.06	32.0
Population	289	24 470	8980	38 100	0	287 900

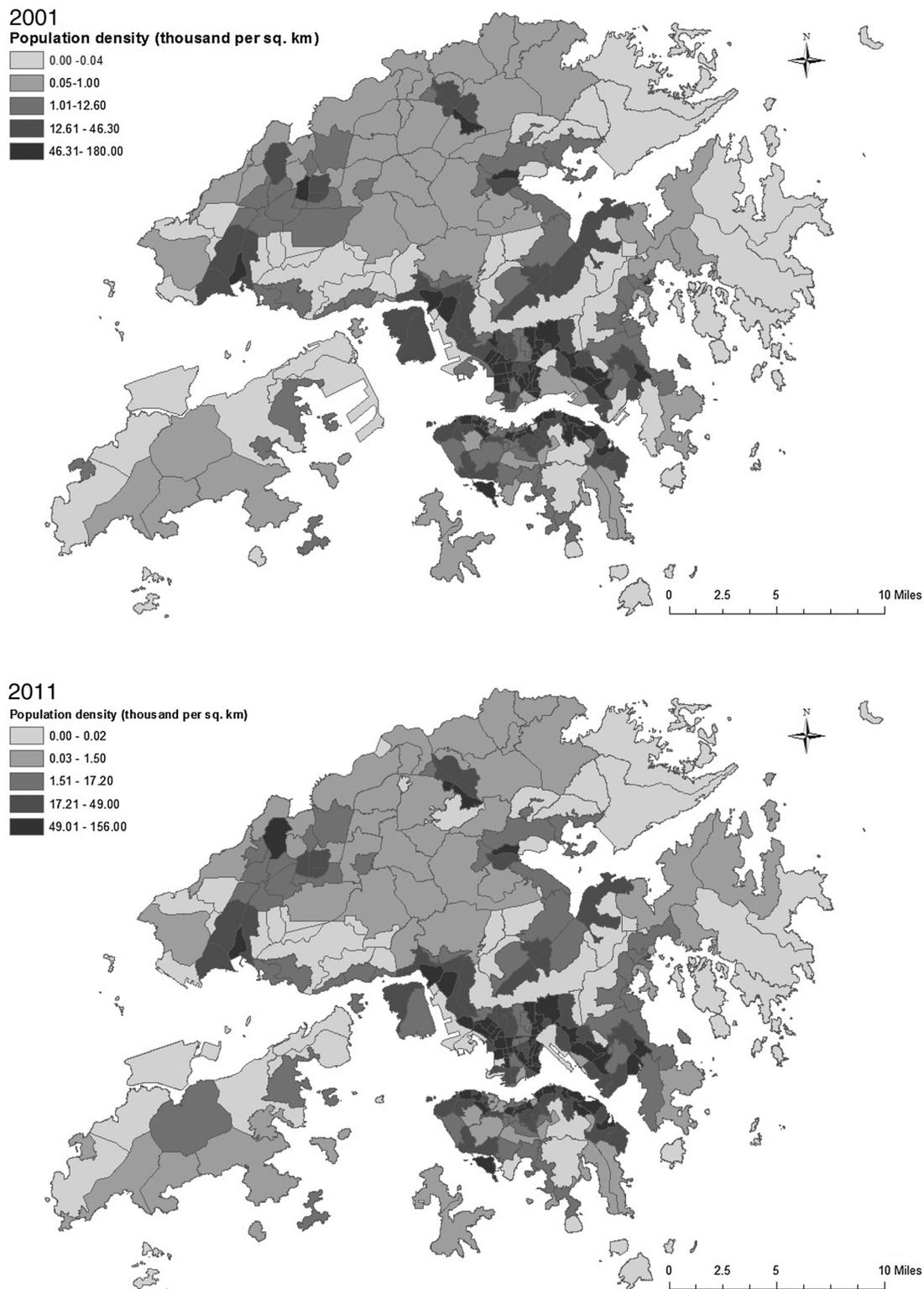


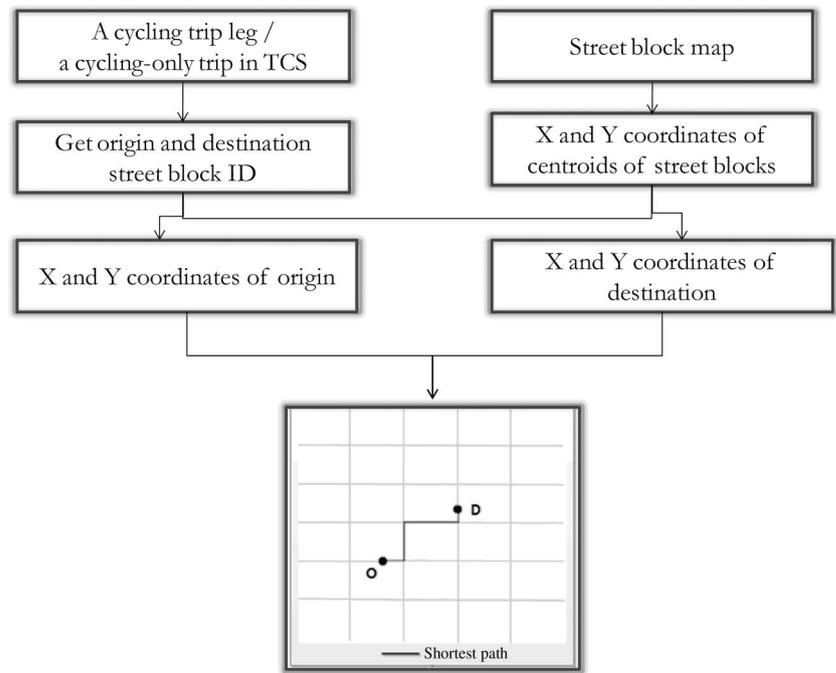
Figure 1 Population density in Hong Kong by TPU, 2001 and 2011.

blocks, of which the centroids are treated as locations where a trip/trip leg started or ended. In Hong Kong, there are about 5000 street blocks in the whole territory, with an average area of 0.22 km². With the geographic coordinates of the centroids of street blocks, the origin and destination for each trip/trip leg are plotted on the map. With the OD pairs, cycling STPs are calculated by performing the shortest path algorithm in ArcGIS 10. The STPs are then overlaid with the TPU maps to extract

the part of routes within the boundary of each TPU. By aggregating routes, the cycling exposure is generated by TPU.

Due both to sampling and non-sampling errors, the Travel Characteristics Surveys (like most other surveys) could not achieve proportionate sampling of different strata of the target population. The household expansion factors are statistical weightings used to correct for the bias of oversampling and undersampling of different subgroups.^{22 23 29} By taking the

Figure 2 Flow chart for generating bicycle exposure.



household expansion factors into account, the cycling exposure estimated from the Travel Characteristics Surveys 2002 and 2011 was calculated as 140 790 and 127 140 bicycle-km, respectively. Nonetheless, under-reporting due to recall gap or other factors like social desirability cannot be corrected by the household expansion factors.²⁵ Should observed bicycle counts be available, further validation needs to be conducted.

Model

Collision prediction models are used for collision analysis at the TPU level. The association between cycling exposure at the TPU level and the number of collisions is estimated by controlling for a set of factors such as motor vehicle flow, land use and demographic characteristics. In this research, a negative binomial regression model is built in a similar way of the model form by Schepers⁸ as:

$$BCollison_i = \exp(\alpha)BDist^{\beta_1} \exp(\beta_2 Z) \quad (1)$$

where $BCollison$ is the number of collisions at TPU i ; Z is a set of TPU-based characteristics as control variables, β_2 is the corresponding set of parameters and $BDist$ is cycling exposure measured as bicycle distance travelled. In the equation above, when $0 < \beta_1 < 1$, it suggests that the increase in bicycle collisions with increasing bicycle flow is less than linear.

Variables

The traffic collision databases from 2001 to 2003 and those from 2010 to 2012 were pooled into two 3-year period data sets, respectively. As the travel survey was undertaken on weekdays, bicycle collisions happening on weekends and public holidays on weekdays were excluded.

Apart from bicycle exposure, previous research also found demographic and built environmental factors important in affecting cycling safety, which have been comprehensively reviewed in the work by Vanparijs *et al.*¹⁴ In this research, some demographic and built environmental factors will be introduced

as control variables into the bicycle collision models. To select appropriate variables among various indicators, the variance inflation factor (VIF) was computed to check for interdependency. All the VIF values were in the range from 1 to 3.7 for each period, indicating small collinearity.³⁰ Table 2 presents the variable names and some brief descriptions as well as the data sources, and table 3 shows the descriptive statistics of each variable for both periods. The vehicle flow data are derived from recorded vehicle flow on counting stations installed by the Transport Department of Hong Kong. Every year, the Transport Department publishes the vehicular flow data in its Annual Traffic Census reports. Land use data are obtained by digitising land use paper maps provided by the Lands Department. The demographic and socioeconomic data are collected from various Hong Kong Census reports.

RESULTS

The initial model included all variables in table 3. Of the variables, only cycling exposure, vehicle flow, residential area, proportion of children aged 14 or below and median household income of the neighbourhoods were significant predictors of the number of bicycle collisions. The final models only included the significant variables, as reported in table 4.

The activity-based exposure variable is statistically significant at 0.05 level in accounting for the variability in the number of bicycle collisions in both periods. The coefficients of the variable were greater than 0 but less than 1, indicating that the number of bicycle collisions went up but the collision risk decreased with the increasing use of bicycles at a TPU. More precisely, if the volume of cycling doubled at a TPU, the number of bicycle collisions increased by 18% and 14% for the period from 2001 to 2003 and the period from 2010 to 2012, respectively, but the bicycle collision rates decreased by 41% and 43%, respectively. The evidence from this empirical study demonstrates that the SIN phenomenon for bicycle users is also found across TPUs within the city of Hong Kong.

The two models also indicate that five other variables are statistically significant at 0.05 level in explaining for the variability

Table 2 Name, description and data source of variables

Variable	Description	Data sources
Dependent variable		
BCollision	Number of collisions involving bicycles happening on weekdays	Police collision databases in 2001, 2002, 2003, 2010, 2011 and 2012 by Hong Kong Police
Built environment independent variables		
BDist	Cycling flow (bicycle distance travelled in bicycle-kilometre)	Travel Characteristics Surveys 2002 and 2011 by Transport Department
VehDist	Vehicle flow (vehicle distance travelled in millions vehicle-kilometre)	Traffic Census 2002 and 2011 by Transport Department
BTrack	Length of bicycle tracks (km)	Transport Department
ComArea	Commercial area (km ²)	Land use paper maps 2002 and 2011 by Lands Department
ResArea	Residential area (km ²)	Land use paper maps 2002 and 2011 by Lands Department
Road	Length of roads (km)	Node-link road systems 2004 and 2011 by Lands Department
JunDensity	Junction density (number of junctions per km)	Node-link road systems 2004 and 2011 by Lands Department
Demographic and socioeconomic independent variables		
Population	Population (in thousands)	Population census 2001 and 2011 by Census and Statistics Department
pAge0_14	Proportion of people aged between 0 and 14	Population census 2001 and 2011 by Census and Statistics Department
pAge15_24	Proportion of people aged between 15 and 24	Population census 2001 and 2011 by Census and Statistics Department
pAge25_44	Proportion of people aged between 25 and 44	Population census 2001 and 2011 by Census and Statistics Department
pAge45_64	Proportion of people aged between 45 and 64	Population census 2001 and 2011 by Census and Statistics Department
pMale	Proportion of men	Population census 2001 and 2011 by Census and Statistics Department
MeEmIncome	Median employment income (in thousands)	Population census 2001 and 2011 by Census and Statistics Department

Table 3 Descriptive statistics of variables

	N	Mean	Median	Std. deviation	Minimum	Maximum
<i>Period from 2001 to 2003</i>						
BCollision	282	11.34	2	28.54	0	174
BDist	282	523.44	1.05	1438.69	0.00	10 255.73
VehDist	282	0.11	0.07	0.13	0.00	0.81
BTrack	282	0.39	0.00	1.35	0.00	12.15
ComArea	282	0.01	0.001	0.02	0.00	0.18
ResArea	282	0.27	0.16	0.31	0.00	1.87
Road	282	15.33	11.71	13.02	0.00	80.13
JunDensity	282	6.30	6.47	2.83	0.00	21.16
Population	282	23.78	7.52	36.65	0.00	203.62
pAge0_14	282	12.75	13.40	3.35	0.0	21.4
pAge15_24	282	11.07	10.80	3.63	2.5	28.2
pAge25_44	282	35.01	34.60	5.45	21.9	57.9
pAge45_64	282	29.37	29.00	4.23	15.4	43.8
pMale	282	0.47	0.46	0.06	0.30	0.80
MeEmIncome	282	12.57	10.00	6.98	0.00	75.00
<i>Period from 2010 to 2012</i>						
BCollision	289	14.34	3	33.59	0	292
BDist	289	456.78	0.90	1146.06	0.00	7590.66
VehDist	29	0.10	0.06	0.12	0.00	0.76
BTrack	289	0.39	0.00	1.33	0.00	12.17
ComArea	289	0.01	0.001	0.02	0.00	0.18
ResArea	289	0.26	0.15	0.32	0.00	1.87
Road	289	15.49	11.87	13.66	0.00	79.42
JunDensity	289	6.01	6.01	2.75	0.00	21.16
Population	289	24.47	8.98	38.10	0.00	287.90
pAge0_14	289	12.28	12.30	3.00	0.0	20.8
pAge15_24	289	10.77	10.70	3.08	2.2	27.2
pAge25_44	289	34.00	33.00	5.32	21.0	57.3
pAge45_64	289	30.27	30.50	3.85	15.9	44.7
pMale	289	0.47	0.47	0.05	0.36	0.85
MeEmIncome	289	18.45	14.07	12.07	0.00	100.00

of safety of bicycle users across different TPUs. Notably, it was found that TPUs with higher motor vehicle flow, longer bicycle tracks, more residential land use and more children had greater numbers of bicycle collisions after controlling for bicycle exposure. Last but not least, the median employment income was negatively associated with the number of bicycle collisions, indicating that TPUs where there were more residents with lower socioeconomic status had larger number of bicycle collisions, even controlling for exposure.

DISCUSSION

Exposure is a very important variable for explaining systematic variations in collision frequency. This study is able to control both vehicle and bicycle flows. In particular, by reconstructing the cycling routes of respondents in Geographic Information System (GIS) based on travel survey data, this study is capable of controlling distance travelled by bicycles with a bicycle exposure variable (BDist), which is often difficult to obtain and hence often lacking in cycling safety research.

The exponent of the bicycle flow is about 0.2, much lower than that in most studies evaluating the non-linearity of the risks faced by cyclists (usually between 0.3 and 0.5).^{1 31} If only bicycle exposure is modelled (ie, not including the five control variables), the coefficient is equal to 0.37 and 0.33 in the two periods, respectively, which are comparable with other studies regardless of the spatial scale they used. This suggests that controlling other variables may help better evaluate the effect of bicycle exposure.

Many studies have discussed the reasons behind SIN.^{1 6 8 12 32} A recent study by Jacobsen *et al*³² discussed three possible reasons for explaining SIN, including safer street regulations, design and operation; changes in behaviour of cyclists; and changes in behaviour of motorists. They suggested that the most likely explanation for SIN was motorist behaviour. It is important to explicitly recognise that SIN is a phenomenon but it does not necessarily suggest a causal relationship. Once the statistical association is established, efforts should be made to

Table 4 Results from negative binomial models

	Coefficient	SD	z	P>z	95% CI	
Period from 2001 to 2003 (Model 0103) (LR χ^2 (6)=267.6 with p<0.001; log-likelihood=-671.86; AIC=1359.7; BIC=1388.85)						
BDist	0.24	0.03	7.71	<0.001	0.18	0.30
VehDist	2.55	0.69	3.69	<0.001	1.20	3.90
BTrack	0.24	0.08	2.88	0.004	0.08	0.41
ResArea	1.57	0.34	4.66	<0.001	0.91	2.22
pAge0_14	0.06	0.02	2.89	0.004	0.03	0.09
MeEmIncome	-0.05	0.02	-2.61	0.009	-0.08	-0.01
Period from 2010 to 2012 (Model 1012) (LR χ^2 (6)=281.3 with p<0.001; log-likelihood=-815.9; AIC=1647.9; BIC=1677.2)						
BDist	0.19	0.03	6.95	<0.001	0.14	0.25
VehDist	2.33	0.59	3.91	<0.001	1.16	3.49
BTrack	0.29	0.07	3.94	<0.001	0.15	0.44
ResArea	1.29	0.27	4.73	<0.001	0.75	1.82
pAge0_14	0.05	0.02	2.14	0.033	0.004	0.101
MeEmIncome	-0.02	0.006	-3.31	0.001	-0.03	-0.01

identify the reasons behind, which may not be the same in different contexts, for promoting cycling safety, especially in areas of low bicycle usage.

Next, the finding that traffic flow is significantly and positively associated with the occurrence of bicycle collision is in accordance with the previous studies such as Geyer *et al*¹³ and Wei and Lovegrove.³³ The main reason may be related to the complexity of transport systems and the mix of road users. The (positive) coefficient of bicycle tracks indicates that TPUs with cycle tracks have greater number of bicycle collisions even after controlling for bicycle exposure. This finding suggests that the safety risk of fragmented cycle tracks, especially where they interface with vehicular traffic, identified by Loo and Tsui,²⁴ has remained. Very often, cyclists who would like to ride on cycle tracks have to use public roads for linking multiple segments of cycle tracks,²⁴ which increases the chance of being involved in a bicycle-vehicle collision. Consistent with other studies,^{9 24} child cyclists (below 14 years old) are also more likely to be injured in a collision due probably to inadequate cycling skill. The reasons for the negative coefficient of median employment income can be twofold. First, people with higher socioeconomic status may have better knowledge about cycling safety and cycling behaviour. Second, socially deprived area may have fewer resources for installing enough transportation facilities, such as signage, for bicycle users.

It should be pointed out that bicycle collisions in this research were modelled by negative binomial regression, a conventional statistical model that does not account for spatial correlation (ie, values of neighbouring TPU's not being independent). One reason for the choice is that even though a number of advanced spatial statistical models have been developed, negative binomial regression has still been widely used in modelling traffic collisions. Although a failure to correct for spatial correlation probably yields an underestimation of the variance of coefficients, the results of the models have shown that the problem is not severe because the main coefficients were highly significant. Another issue is that the analysis in this study was conducted at the TPU level. However, TPUs are not homogeneous. It is hence worth further exploring SIN at a finer spatial scale like street blocks. In addition, as bicycle counts on roads in Hong

Kong are not available, this research employs an activity-based approach to measure cycling exposure based on the travel survey. The results have shown that the estimated bicycle exposure can well explain the occurrence of bicycle collisions in this empirical study. However, in the absence of actual bicycle flow data on roads, the activity-based approach cannot be further validated. Therefore, more research efforts should be dedicated to the validation of this approach in measuring cycling exposure when other bicycle data are available.

CONCLUSIONS

This research suggests that the often-claimed dilemma of discouraging cycling to prevent injury of cyclists and encouraging cycling for sustainable transport and health benefits may not necessarily hold. When specific and appropriate measures are in place,^{1 6 8 12 32} it is possible to achieve both the objectives of reducing cycling injury and maximising the health and environmental benefits of cycling in a sustainable transport system.

To examine whether SIN for cyclists applies across different neighbourhoods within the same city, this study conducted a local-level analysis on bicycle collisions in Hong Kong. The results indicate that SIN for cyclists applies to the local level as well within the same city. Even in a highly motorised city like Hong Kong, cyclists are less likely to be involved in road collisions in communities with higher cycling volume. Despite the health benefits from cycling, a model shift from driving to cycling may be greatly hindered by safety concerns. In the travel surveys of Hong Kong,^{22 23} it was found that about 20% of people cited 'danger to cycle on roads' as the main reason for not cycling. Cyclists are vulnerable road users and are more likely to be injured than motorists in a traffic collision. Hence, more research efforts should be dedicated to the interventions for protecting cyclists, especially in areas with lower bicycle usage.

What is already known on the subject?

- ▶ Safety in numbers for cyclists was found mainly using national-level and city-level data.
- ▶ The above was also found to apply to selected intersections and road segments.
- ▶ Data on cycling activities are scarce and not comprehensive in most administrations.

What this study adds?

- ▶ Safety in numbers for cyclists was also found to apply at the local level within the same city.
- ▶ In Hong Kong, cyclists are less likely to be involved in road collisions in communities with higher cycling volume.
- ▶ In areas where bicycle volume on roads is not systematically collected, cycling exposure can be estimated from travel surveys.

Acknowledgements The authors thank the Committee on Research and Conference Grants (CRCG) and the Knowledge Exchange (KE) Fund of the University of Hong Kong for funding 'Can Cycling Training Enhance Cycling Safety in Hong Kong? A Study from the Perspectives of Adolescent Cyclists and Their Parents' and 'Promoting Cycling Safety in Hong Kong: Evidences from Adolescent Cycling Training

Programme', respectively. The findings in this paper are directly related to the background research of the above projects.

Contributors BPYL identified the research gap, suggested the literature, provided the databases and critically reviewed the data quality and methodology. SY developed the methodology, performed the analysis and drafted the manuscript. BPYL provided comments on earlier drafts, revised the submission and made final approval of the version submitted.

Funding University of Hong Kong (CRCG104003758 and KE002100017).

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The data sets are owned and made accessible to the authors from the Hong Kong SAR Government.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Elvik R. The non-linearity of risk and the promotion of environmentally sustainable transport. *Accid Anal Prev* 2009;41:849–55.
- Higgins PA. Exercise-based transportation reduces oil dependence, carbon emissions and obesity. *Environmental Conservation* 2005;32:197–202.
- Rojas-Rueda D, de Nazelle A, Tainio M, et al. The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. *BMJ* 2011;343:d4521.
- Oja P, Titze S, Bauman A, et al. Health benefits of cycling: a systematic review. *Scand J Med Sci Sports* 2011;21:496–509.
- Mueller N, Rojas-Rueda D, Cole-Hunter T, et al. Health impact assessment of active transportation: a systematic review. *Prev Med* 2015;76:103–14.
- Robinson DL. Safety in numbers in Australia: more walkers and bicyclists, safer walking and bicycling. *Health Promot J Austr* 2005;16:47–51.
- Jacobsen PL. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Inj Prev* 2003;9:205–9.
- Schepers P. Does more cycling also reduce the risk of single-bicycle crashes? *Inj Prev* 2012;18:240–5.
- Vandenbulcke G, Thomas I, de Geus B, et al. Mapping bicycle use and the risk of accidents for commuters who cycle to work in Belgium. *Transp Policy* 2009;16:77–87.
- Schepers JP, Heinen E. How does a modal shift from short car trips to cycling affect road safety? *Accid Anal Prev* 2013;50:1118–27.
- Gärder P, Leden L, Pulkkinen U. Measuring the safety effect of raised bicycle crossings using a new research methodology. *Transportation Res Record* 1998;1636:64–70.
- Bonham J, Cathcart S, Petkov J, et al. Safety in numbers: a strategy for cycling? Paper presented at: The 29th Australasian Transport Research Forum 2006;2006 Sep 27–29 Brisbane, Australia.
- Geyer J, Raford N, Pham T, et al. Safety in numbers: data from Oakland, California. *Transportation Res Record* 2006;1982:150–4.
- Vanparijs J, Panis LI, Meeusen R, et al. Exposure measurement in bicycle safety analysis: a review of the literature. *Accid Anal Prev* 2015;84:9–19.
- Lusk AC, Furth PG, Morency P, et al. Risk of injury for bicycling on cycle tracks versus in the street. *Inj Prev* 2011;17:131–5.
- Aultman-Hall L, Hall FL. Ottawa-Carleton commuter cyclist on-and off-road incident rates. *Accid Anal Prev* 1998;30:29–43.
- Bacchieri G, Barros AJ, Dos Santos JV, et al. Cycling to work in Brazil: users profile, risk behaviors, and traffic accident occurrence. *Accid Anal Prev* 2010;42:1025–30.
- Blaizot S, Papon F, Haddak MM, et al. Injury incidence rates of cyclists compared to pedestrians, car occupants and powered two-wheeler riders, using a medical registry and mobility data, Rhône County, France. *Accid Anal Prev* 2013;58:35–45.
- Yao S, Loo BPY, Lam WWY. Measures of activity-based pedestrian exposure to the risk of vehicle-pedestrian collisions: space-time path vs. potential path tree methods. *Accid Anal Prev* 2015;75:320–32.
- Lam WWY, Yao S, Loo BPY. Pedestrian exposure measures: a time-space framework. *Travel Behav Soc* 2014;1:22–30.
- Lam WWY, Loo BPY, Yao S. Towards exposure-based time-space pedestrian crash analysis in facing the challenges of ageing societies in Asia. *Asian Geographer* 2013;30:105–25.
- Transport Department. *Travel Characteristics Survey 2011 Final Report*. Hong Kong: Transport Department, 2014.
- Transport Department. *Travel Characteristics Survey 2002 Final Report*. Hong Kong: Transport Department, 2003.
- Loo BPY, Tsui KL. Bicycle crash casualties in a highly motorized city. *Accid Anal Prev* 2010;42:1902–7.
- Loo BPY, Tsui KL. Factors affecting the likelihood of reporting road crashes resulting in medical treatment to the police. *Inj Prev* 2007;13:186–9.
- Schepers P, Agerholm N, Amoros E, et al. An international review of the hidden problem of single-bicycle crashes. *Inj Prev* 2015;21:e138–43.
- Loo BPY. Validating crash locations for quantitative spatial analysis: a GIS-based approach. *Accid Anal Prev* 2006;38:879–86.
- Yao S, Loo BPY, Yang BZ. Traffic collisions in space: four decades of advancement in applied GIS. *Annals of GIS* 2016;22:1–14.
- To DKB, Yau KT, Lam A. Travel characteristics survey: method of expanding household interview survey data. *Transportmetrica* 2005;1: 247–60.
- Belsley DA, Kuh E, Welsch RE. *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley & Sons, 2005.
- Elvik R, Bjørnskau T. Safety-in-numbers: a systematic review and meta-analysis of evidence. *Safety Sci*. Published Online First: 29 July 2015;
- Jacobsen PL, Ragland DR, Komanoff C. Safety in numbers for walkers and bicyclists: exploring the mechanisms. *Inj Prev* 2015;21:217–20.
- Wei F, Lovegrove G. An empirical tool to evaluate the safety of cyclists: community based, macro-level collision prediction models using negative binomial regression. *Accid Anal Prev* 2013;61:129–37.