Safety Science 159 (2023) 106015

Contents lists available at ScienceDirect

Safety Science

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

Examining the impact of exposure, built environment and socio-demographics on pedestrian safety: A case study of Greater London

Michael Forrest, Shahram Heydari *, Tom Cherrett

Transportation Research Group, Department of Civil, Maritime and Environmental Engineering, University of Southampton, Southampton, UK

A R T I C L E I N F O

Keywords:
Pedestrian safety
Crash frequency
Hierarchical (multilevel) data
Macro-level
Random parameters
Deprivation

A B S T R A C T

Walking as a mode of travel alleviates congestion, pollution and concerns over physical fitness. However, safety concerns might be a barrier for some people to walking more often, so finding ways to improve pedestrian safety is important. This paper conducts a ward-level study of Greater London such that factors which are associated with pedestrian safety can be identified and interventions to improve safety can be more appropriately targeted. A wide range of factors relating to exposure, land use, built and natural environment, and socio-demographics, including markers of deprivation were considered in our analysis. We employed a multilevel random parameters negative binomial regression model with a hierarchical structure which encompassed the boroughs and wards of Greater London allowing dependency in the data and unobserved heterogeneity to be addressed more fully. The results show that BAME population, number of schools, job density and alcohol expenditure are associated with an increase in pedestrian crashes as are crime rate and children in child benefit households, which are both markers of deprivation. Also, the average number of cars per household, the proportion of green space and the percentage of adults who commute by walking have a decreasing effect on pedestrian crash frequency. This study identifies important determinates of ward-level pedestrian safety and estimates the magnitude of their association with pedestrian safety and in doing so reveals important between borough (local authority) differences in the Greater London area in terms of pedestrian safety which were not previously known or well-understood.

1. Introduction

There is increasing encouragement for people to walk short and medium length trips in order to help alleviate congestion, pollution and physical health concerns. Because of this, ensuring the safety of pedestrians is very important. Although there are well documented benefits of walking, pedestrians are vulnerable road users as they do not have the protection from collisions which would be afforded by a vehicle. Looking at statistics from Great Britain in 2019 (DfT, 2020) pedestrians incur 1640 casualties per billion passenger miles travelled, where for car drivers this value is only 195. It is crucial to understand why this figure is so high and find ways to address it as safety can be a key barrier to the uptake of walking for short trips (Bozovic et al., 2021). As well as this, for governmental bodies to encourage walking as a travel mode without an attempt to mitigate these casualties is irresponsible.

One means of identifying areas for improvement regarding pedestrian safety is by identifying local characteristics which are associated with a higher crash frequency. Rather than studying individual crashes at a micro level, these crashes are aggregated to larger areas and the factors which are associated with an increased or decreased level of safety can be identified across these areas; i.e., at a macro level. In this way, resources to address higher pedestrian crash frequency can be more appropriately targeted. This approach has been utilised in multiple studies, but less commonly with active travel modes as the subject and even more rarely based in the UK.

1.1. Previous research

Microscopic, or crash-level, studies, are used to examine crashes at individual road elements such as junctions or segments and suggest improvements which might alleviate crashes at this level. Macroscopic, or area-level, studies are used instead when the intention is to investigate factors affecting safety across a larger study area. A London based example was carried out by Quddus (2008) who displayed the possibility to better account for spatial correlations and heterogeneity across wards with this approach. A multi-level modelling approach has been used by researchers to overcome the issues with dependencies in the data which come from modelling a hierarchical structure (Dupont et al.

* Corresponding author.
E-mail addresses: mgf1g18@soton.ac.uk (M. Forrest), s.heydari@soton.ac.uk (S. Heydari), t.j.cherrett@soton.ac.uk (T. Cherrett).

https://doi.org/10.1016/j.ssci.2022.106015
Received 22 June 2022; Received in revised form 5 October 2022; Accepted 9 November 2022
Available online 17 November 2022

2013; Heydari et al. 2016). Another issue which is found in macro-level data is unobserved heterogeneity which has been addressed in different ways including random parameter models (Mannering et al., 2016).

Exposure, socio-demographics, road network and traffic characteristics and land use, built and natural environment have been explored in previous research and found to be associated with crash frequency. Exposure covers any measure which describes how much a given travel mode is used, for example vehicle miles travelled in a borough. Pedestrian exposure can be difficult to measure as there are very rarely observed counts available in the same way as there might be for motor vehicles. Population and population density are common proxies (Cottrill and Thakuriah, 2010), though some residents take more trips than others and some areas may have a greater number of destination points such as workplaces or leisure activities. Revealed preference can instead be used (Sze et al., 2019) as well as trip assignment methods (Li et al., 2020).

Socio-demographic factors are those which concern the characteristics of people living in an area and researchers have demonstrated associations between these factors and crash frequency. This includes factors such as proportion of children (Bernhardt and Kockelman, 2021) or elderly people (Lee et al., 2017), ethnicities (Ding et al., 2020; Su et al., 2021) and education levels (Su et al., 2021). Also, economic measures such as income (Ding et al., 2020) and the average number of cars per household (Siddiqui et al., 2012). The latter has also been used as a proxy for pedestrian exposure (Cai et al., 2016). Deprivation is a term used to cover a range of measures which indicate that an area is economically deprived and, in the UK, these characteristics are categorised in the index of multiple deprivation (IMD) (see Graham and Stephens (2008) for detail on the individual indices). An association has been shown to exist between deprivation and pedestrian crash frequency (Graham et al., 2005; Graham and Stephens, 2008; Green et al., 2011). Crime rate is also noted to be associated with an increase in pedestrian crashes (Cottrill and Thakuriah 2010), however the reasons behind this relationship are not well defined.

Variables which are related to transport, but which are not exposure measures are here referred to as road network and traffic characteristics. Road network characteristics (Wang et al., 2016; Kraidi and Evdorides, 2020) and transit stops (Pulugurtha and Sambhara, 2011) have been shown to be associated with safety. The volume of traffic can have the effect of creating more conflicts with pedestrians, but also at high volumes may reduce traffic speed and associated crash risk (Su et al., 2021). Land use measures i.e., the percentage of an area which is used for a certain purpose such as industrial or residential, has been shown to have an association with crash frequency (Amoh-Gyimah et al., 2016; Hu et al., 2020). The built environment includes areas of interest which might cause an increased number of trips to or from the area such as the number of schools (Cottrill and Thakuriah, 2010). While the effect of the built environment on pedestrian safety is the subject of much micro level research (Corazza et al., 2020; Fonseca et al., 2022; Fossum and Ryeng, 2022), the effect at macro-level is less well reported, especially in the UK context. Natural environment refers to factors which are less related to person-made alterations to an area such as proportion of greenspace which has been shown to be associated with crash frequency (Ding et al., 2020).

Research does not always align on the effect of a factor; for example, number of schools was found to be positively associated with an increase in pedestrian and cyclist crashes by Cottrill and Thakuriah (2010), likely because of an increased amount of exposure in these areas (Ukkusuri et al., 2012), although this has not been the universal conclusion of research into the impact of schools. Better traffic control and driver awareness could account for a decrease in crashes around schools (Pljakića et al.; 2019; Su et al., 2021). More research into the impacts which schools have on safety would help to explain why seemingly opposing effects can be observed. Greater London has been used for investigations into how area-level factors can affect public health (Jayakody et al. 2006) and transport route choice (Raveau et al. 2014) but relating these factors to traffic safety is uncommon. Research by Li et al. (2017) and Ding et al. (2020) explore how local characteristics affect the distribution of cycle crash frequency across Greater London.

1.2. Current article

Surprisingly, research on pedestrian safety in the Greater London area, although striving to be the world's most walkable city (TfL, 2018), is relatively limited. For example, Zhu et al. (2022) investigated the impact of street layout on pedestrian crashes at a micro level. Steinbach et al. (2010) conduct an area-level investigation for Greater London; however, there is a specific focus on ethnic differences in child pedestrian injury rates. To the researchers' knowledge, a comprehensive investigation into the factors which are associated with zonal-level pedestrian crash frequency in Greater London represents a gap in knowledge. Greater London makes an interesting study location in this regard as each of the 32 (33 with City of London) boroughs are independent local authorities and as such are administrated by different councils. Furthermore, using a multi-level modelling technique to address the hierarchical structure of Greater London in this way represents an empirical novelty in the context of pedestrian safety. The current paper will contribute to the literature by undertaking a wide-ranging examination of the factors affecting pedestrian crash frequency in London at a macroscopic level using recent road safety data, which is a less well understood area of safety research in England compared to other parts of the world such as North America.

Many factors can have an effect on crash frequency and identifying what these are for a given study area as well as the magnitude of their effects is an important area of research. These factors can be grouped into a number of categories including exposure, sociodemographics, land use and built and natural environment, road network and traffic characteristics, walkability measures and traffic calming measures. Data on variables relating to the former four categories is included in the present analysis to ensure that a range of factors are considered. Data from Greater London will be used and will be granulated at ward level. Using the 630 wards nested within the 32 boroughs of Greater London, this paper will use a hierarchical (multi-level) modelling approach to identify zonal-level characteristics of London which are significant in explaining ward-level pedestrian crash frequency. To better address unobserved heterogeneity between boroughs and dependencies within them, a random parameters model will be used.

The paper has given a concise review of pertinent literature in section one, section two will give more detail about the dataset used and section three will describe how the analysis was undertaken. Section four will outline the results and provide a discussion of their implications for policy; and finally, conclusions, including a comparison with previous research and recommendations for safety improvements, will be presented in section five.

2. Data

2.1. Crash data

The crash data which are used in this study represent all crashes recorded in Greater London in the years 2016–2018 which were reported to the police and involved a pedestrian casualty. This information is extracted from STATS19 databases as recorded by UK police forces and collated by the Department for Transport. All crashes in England which are reported to the police are required to include details of the local authority in which the crash occurred, so in the case of London they are assigned to a borough. The ward is not recorded by the police; however, each crash is geo-tagged so using GIS software it is possible to identify crash counts for each ward. While it has been noted that it is likely that these geo-tags are not perfectly accurate (Imprialou et al., 2015) it has been deemed that the precision issues will be negligible for ward-level counts. Pedestrian crashes are identified via a field on the

STATS19 reporting form which indicates whether a casualty is a pedestrian. If a crash featured one or more of casualty identified as a pedestrian, then it was included in the present crash data.

The study period 2016–2018 was selected because the crash data were freely available and to have a clear separation from the effects of the COVID-19 pandemic, the effect of which is not within the scope of this paper. The decision was taken to aggregate the crash data over the three years of the study period as in Quddus (2008). This was found to give a better model fit than using crash data from a single year as this is subject to annual fluctuations which occur due to the fact that crashes are random events and would have led to a greater number of wards having zero counts.

2.2. Area-level variables

London has 32 boroughs which are divided into 629 wards according to boundaries which were accurate for the period 2014–2018, each of which has broadly uniform socio-demographic and built environmental characteristics. City of London is sometimes included in the count of boroughs to give a total of 33, although it is in reality a local government district and not a borough. The surface area of City of London is approximately that of a ward and its characteristics are broadly uniform in the same way as a ward would tend to be. Also, many variables were not available for the wards located in the City of London. Because of this, for the duration of this paper, the City of London has been treated as a ward giving 630 in total. For further context of the boroughs of London, Figs. 1 and 2 show the spatial distribution of some key variables for the individual boroughs and wards respectively with darker shades representing higher values. Notably in Fig. 1 the, traditionally more affluent, south west boroughs show less children in child benefit families than north and east London. Also, Fig. 1 indicates higher expenditure on alcohol around the central London area. In Fig. 2, the tendency for higher crime rates in inner London and higher car ownership in outer London can be seen.

To comprehensively investigate the characteristics of wards and boroughs which have an association with pedestrian crash frequency, the range of variables studied needed to be drawn from different categories, namely: exposure, socio-demographics, land use and built and natural environment. Beyond ensuring that these broad areas were represented in the data, inspiration and guidance was taken from the literature and the researchers' knowledge to select which variables could provide an informative analysis and were likely to yield significant results. Table 1 shows summary statistics of the variables which were investigated in the statistical analysis. It was possible to obtain much of the data required for analysis from two data sets kept by the Greater London Authority (GLA), these being London ward profiles (GLA, 2015a) and London borough profiles (GLA, 2015b). These profiles are comprised of a range of sources from various agencies and the years from which the data are taken reflect this. Data included in the present paper are from the 2016–2018 period to match the crashes where possible and as close as feasible if not. As Table 1 shows, variables are also taken at either ward or borough level as granulation to ward level is not always possible or practical for the data collectors. The exposure measures included were determined by availability to a certain extent and mostly proxy measures were used, particularly for active travel modes. The traffic flow in each borough is available in the form of vehicle kilometres travelled from the DfT's National Road Traffic Survey which is conducted at local authority level i.e., borough level, annually. However, the same cannot be said for walking and cycling and so commuting data from the 2011 census was used as well as data on how much people walk on a weekly basis from Sports England's active lives 2018 survey. London underground entry and exit data was also available which has been shown to be associated with active travel trips as using the underground is often accompanied by a travel stage of walking or cycling (TfL, 2021).

The land use characteristics investigated include the proportion of greenspace, domestic gardens, domestic and non-domestic buildings and water, all of which are available as part of the generalised land use database published by the Department for Communities and Local Government. The density of certain built environment features such as schools, cycle network and pubs as well as the coverage of roads is included. Also, we considered the number of Santander docking stations, a bike share scheme which has become an increasingly popular mode of transport in London (Lovelace et al., 2020; Chibwe et al., 2021). Socio-demographic factors describe the characteristics of the population which live in a ward or borough and information on these factors comes from a range of sources. Population estimates including the proportion of children are from the Strategic Housing Land Availability Assessment, provided by the GLA. Alcohol expenditure comes from a series of consumer expenditure estimates by Experian. The 2011 census, which represents the most up to date census data at the time of research, is the source for several variables including the average number of cars which a household has access to, the proportion of people who identify as BAME and the proportion of people with level four qualifications or above. Some economic characteristics like unemployed population, average weekly household earnings and the number of children in child benefit families are published by the Department for Work and Pensions (DWP). Also included is the crime rate, which along with income,

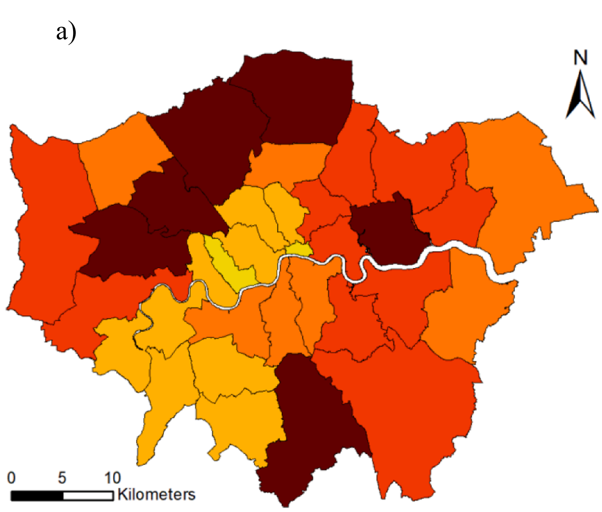

Fig. 1. Heat maps of the boroughs of Greater London showing a) the number of children in child benefit families and b) the expenditure on alcohol.

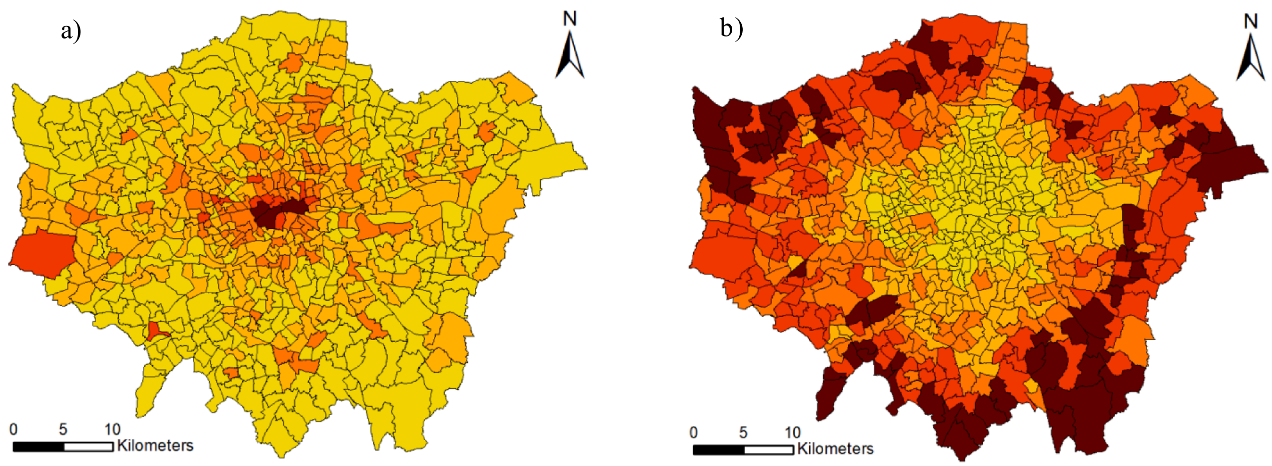

Fig. 2. Heat maps of the wards of Greater London showing a) the crime rate per unit population of the ward and b) the average number of cars per household.

Table 1
Summary statistics for area level variables included in the study.

Variable types	Variables	Spatial unit	Mean	SD	Minimum	Maximum
Crashes						
	Pedestrian crashes	Ward	28.468	26.942	2.000	365.000
Exposure						
	Population cycling to work (%)	Ward	4.011	3.395	0.242	19.092
	Population walking to work (%)	Borough	5.842	2.636	3.700	36.400
	Population walking at least once a week (%)[1]	Borough	73.156	5.408	62.800	86.510
	Population walking at least three times a week (%)[1]	Borough	49.591	5.945	40.071	64.922
	Population walking at least five times a week (%)[1]	Borough	38.223	5.224	30.090	51.830
Road network and traffic characteristics						
	Tube entries and exits (00 millions of travellers)	Borough	0.831	1.242	0.000	6.553
	Traffic flow (Billion vehicle kilometres travelled)	Borough	0.936	0.422	0.150	2.147
	Cycle network density	Borough	2.373	0.769	0.585	4.125
	Road coverage (km)	Borough	471.884	174.143	54.880	906.180
	Santander docking stations (00 s)	Borough	0.223	0.402	0.000	1.660
Land use, built and natural environment						
	Area of ward (km^2)	Ward	2.720	2.760	0.350	29.04
	Number of schools	Ward	5.146	2.791	0.000	24.000
	Land use with greenspace (%)	Ward	26.20	16.60	15.30	90.00
	Land use with domestic gardens (%)	Ward	26.208	12.092	0.120	59.161
	Land use with domestic buildings (%)	Ward	11.904	5.154	0.850	30.540
	Land use with non-domestic buildings (%)	Ward	6.582	5.701	0.390	42.180
	Land use with water (%)	Ward	2.149	6.304	0.000	74.240
	Density of pubs (pubs per km^2)	Borough	3.650	4.718	0.529	49.712
Socio-demographic						
	Number of cars per household	Ward	0.840	0.327	0.233	1.705
	Population (000 s)	Ward	14.132	3.083	4.622	32.046
	Population per square kilometre (0000 s)	Ward	0.875	0.520	0.019	2.766
	Child population (%)	Ward	19.805	3.761	6.473	32.695
	Black, Asian and minority ethnic (BAME) population (%)	Ward	0.389	0.189	0.041	0.937
	Population with level 4 qualifications and above (%)	Ward	37.678	12.840	12.500	68.700
	Crime rate (number of crimes committed per number of population)	Ward	0.088	0.073	0.026	0.894
	Children in child benefit families (000 s)	Borough	6.003	1.841	0.069	9.125
	Total Expenditure on alcohol (£00 millions)	Borough	0.525	0.244	0.029	1.125
	Average weekly earnings (£s)	Borough	561.680	70.385	462.367	902.000
	Unemployed population (%)	Borough	5.322	0.999	3.867	19.633
	Job density	Borough	1.115	4.758	0.410	118.950

[1] Walking measures include trips which are made for any purpose.

unemployment and education, represent some of the markers of deprivation which are often associated with a higher crash frequency in a given area (Graham and Stephens, 2008; Li et al., 2017).

2.3. Model formulation

The base model which is used for most crash frequency analysis is a

Poisson regression as this is the first choice for count data. Furthermore, as crash data often has a variance larger than its mean causing overdispersion (El-Basyouny and Sayed, 2009), a negative binomial approach is more appropriate as an overdispersion parameter is introduced (Washington et al., 2020). The crash frequency y_i can be expressed then as follows (as in Wang et al. (2017)) where λ_i is the expected value of y_i and follows a Poisson distribution. $\boldsymbol{\beta_0}$ represents the intercept, $\boldsymbol{\beta}$ a vector of estimable parameters, $\mathbf{X_i}$ a vector of explanatory variables and $e^{\varepsilon i}$ is a gamma distributed error term which can more appropriately address overdispersion.

$$\begin{aligned} &y_i\ Poisson(\lambda_i) \\ &\ln(\lambda_i) = \beta_0 + \boldsymbol{X_i\beta} + \varepsilon_i \end{aligned} \tag{1}$$

This provides the start point to a crash frequency analysis as in many similar studies (Amoh-Gyimah et al., 2016; Cai et al., 2016), however unobserved heterogeneity occurs in real world crash data and failing to account for this can lead to erroneous inferences. Unobserved heterogeneity (Mannering et al., 2016) is present mainly because crash data are limited, and many risk factors that may affect safety are not available in the data. This can cause a variation in the effect of certain risk factors on safety; for example, crash frequency across different geographical units. In order to help address this unobserved heterogeneity, random parameter effects can be introduced in to the analysis. One means of introducing random parameter effects is by using a multi-level model. The present data set has a hierarchical structure of wards nested within boroughs and so a multi-level random parameters model is utilised to capture dependencies induced by this structure (Dupont et al., 2013) as well as unobserved heterogeneity. Including these random effects in the model can allow the effect of one or multiple variables to change across boroughs (Ukkusuri et al., 2012; Heydari et al., 2018; Sacchi and El-Basyouny, 2018). The general form of the multi-level random parameters model is shown in Eq. (2), where λ_{ij} is the expected crash frequency in ward i, within borough j. $\boldsymbol{\beta_{0j}}$ represents the varying borough effects (random intercepts), $\mathbf{Z_j}$ are the explanatory variables the effects of which vary across boroughs and $\boldsymbol{\alpha_j}$ are their estimable parameters.

$$\ln(\lambda_{ij}) = \beta_{0j} + \boldsymbol{X_j\beta} + \boldsymbol{Z_j\alpha_j} + \varepsilon_i \tag{2}$$

For a detailed discussion on hierarchical data structures in the context of road safety, see Dupont et al. (2013). The advantage of this structure is its ability to capture the effects of unobservables that vary from one borough to another while accommodating dependency between observations that are nested within the same boroughs.

3. Results and discussion

3.1. Model performance

The model fit was assessed using both AIC (Akaike information criteria) and BIC (Bayesian information criterion) as a comparison between different combinations of variables as well as to identify if an improvement was observed in the model fit when introducing random intercepts and random parameters. The multilevel random parameters model selected represents the best fit to the data according to the abovementioned model fitting criteria.

3.2. Factors associated with pedestrian safety

The values of the estimable parameters for the variables used in the models can be seen in Table 2. The variables shown were those which were found to constitute a significant model, each being significant at the 5 % level. The coefficients show that the number of schools, BAME population, crime rate, children in child benefit families, alcohol expenditure and job density are positively associated with pedestrian crash frequency. Meanwhile, proportion of walking commuters, land used as greenspace and number of cars per household have a negative association. Also, the resultant coefficients are broadly similar across the models although the effect of the number of cars per household is notably stronger in the random parameters model as is the effect of the proportion of walking commuters, the percentage of the population which identifies as BAME and expenditure on alcohol. With regard to model fit, both the Akaike and Bayesian Information Criterion (AIC and BIC) improve with the addition of the multi-level structure, indicating an improved goodness-of-fit.

For practical implications in terms of understanding the impact of the explanatory variables on pedestrian safety and ease in the interpretation of the results, marginal effects of the explanatory variables are

Table 2
Regression parameter estimates.

	Coefficient	SD	P-value	95% Confidence Intervals	
Negative binomial model					
Ln(population walking to work)	−0.304	0.084	0.000	−0.468	−0.140
Number of schools	0.026	0.006	0.000	0.015	0.037
Greenspace	−0.353	0.115	0.002	−0.578	−0.128
Cars per household	−0.252	0.079	0.001	−0.407	−0.097
BAME population	0.333	0.101	0.001	0.134	0.531
Ln(crime rate)	0.881	0.042	0.000	0.799	0.963
Children in child benefit families	0.038	0.013	0.003	0.013	0.064
Alcohol expenditure	0.441	0.079	0.000	0.286	0.596
Job density	0.010	0.003	0.002	0.004	0.016
Constant	5.543	0.220	0.000	5.111	5.975
Model fit (AIC)	4598.600	–	–	–	–
Model fit (BIC)	4647.502	–	–	–	–

Multilevel random interceptsnegative binomialmodel

	Coefficient	SD	P-value	95 % Confidence Intervals	
Ln(population walking to work)	−0.305	0.112	0.007	−0.525	−0.085
Number of schools	0.027	0.006	0.000	0.016	0.038
Greenspace	−0.298	0.116	0.010	−0.525	−0.071
Cars per household	−0.266	0.089	0.003	−0.440	−0.092
BAME population	0.375	0.112	0.001	0.156	0.594
Ln(crime rate)	0.883	0.042	0.000	0.800	0.965
Children in child benefit families	0.036	0.017	0.037	0.002	0.070
Alcohol expenditure	0.439	0.110	0.000	0.223	0.656
Job density	0.010	0.003	0.004	0.003	0.017
Borough effect	5.536	0.279	0.000	4.990	6.082
Variance (borough effect)	0.008	0.004	–	0.003	0.021
Model fit (AIC)	4591.839	–	–	–	–
Model fit (BIC)	4645.187	–	–	–	–

Multilevel random parametersnegative binomialmodel

	Coefficient	SD	P-value	95 % Confidence Intervals	
Ln(population walking to work)	−0.345	0.120	0.004	−0.581	−0.110
Number of schools	0.028	0.006	0.000	0.017	0.039
Greenspace	−0.297	0.115	0.010	−0.523	−0.071
Cars per household	−0.314	0.093	0.001	−0.496	−0.133
Variance (cars per household)	0.017	0.009	–	0.006	0.045
BAME population	0.330	0.111	0.003	0.113	0.548
Ln(crime rate)	0.879	0.042	0.000	0.797	0.961
Children in child benefit families	0.039	0.164	0.018	0.007	0.071
Alcohol expenditure	0.400	0.115	0.001	0.174	0.627
Job density	0.010	0.003	0.003	0.004	0.017
Borough effect	5.654	0.291	0.000	5.083	6.225
Variance Borough effect	0.007	0.006	–	0.002	0.034
Model fit (AIC)	4531.764	–	–	–	–
Model fit (BIC)	4589.558	–	–	–	–

reported in Table 3. The log transformed variables are easily adapted from their respective parameter coefficients and the marginal effects are estimated as in Washington et al. (2020).

Examining these measures allows for the identification of the factors which have the greatest effect on pedestrian crash frequency. The marginal effects in Table 3 show the effect which a one unit increase of the independent variable would have on the expected pedestrian crash frequency. The results show that a unit increase in the proportion of greenspace, on average, would lead to a decrease of 8.47 in the ward-level expected pedestrian crashes. One unit increase in proportion of the population which define themselves as BAME would yield an increase of 9.43 crashes over the study period. An additional car per household could be expected to decrease pedestrian crashes, on average, by 8.62 crashes over the study period. Furthermore, it was found that the effect of cars per household on pedestrian safety varies across different boroughs. An additional school would increase expected pedestrian crashes by 0.79. A unit increase in children in child benefit families, alcohol expenditure and job density increases the number of pedestrian crashes over the study period by 1.11, 11.42 and 0.30, respectively.

A one unit increase in the log-transformed percentage of walking commuters would reduce the expected crash frequency by 9.86 and a one unit increase in the log transformed crime rate would increase crash frequency by 25.06 over the study period. As walking and crime rate are log transformed, a rather straightforward way of assessing their impacts on pedestrian crash frequency is by using semi-elasticities based on their respective estimated coefficients (see Table 2). These elasticities show that a 10 % increase in the crime rate of a ward would be expected to be associated with an 8.79 % increase in ward-level expected pedestrian crash frequencies over the study period in Greater London. As well as this, an increase of 10 % in the proportion of people walking to work leads to a reduction of 3.45 % in the ward-level expected pedestrian crash frequency. That is, interestingly, the higher the levels of walking, the lower expected pedestrian crash frequencies.

3.3. Spatial variations in expected pedestrian crash frequencies

The analysis in this paper has identified the wards and boroughs of Greater London which have a high expected pedestrian crash frequency. This process can give useful guidance to policy makers as resources can be more appropriately allocated to locations where they could have the greatest impact. Expected crash frequencies give more reliable inferences of where crash frequency will be high compared to observed crashes, which are prone to the regression to the mean bias (Hauer, 1997). This type of analysis is particularly important for pedestrian crash analysis because local characteristics are more likely to be associated with pedestrian crashes in the area than crashes of other vehicle types as people are likely to be walking closer to home (Haas, et al., 2015; Su et al., 2021).

Figs. 3 and 4 show heat maps of the expected number of pedestrian crashes over the study period in the boroughs and wards of Greater London, respectively. It can be observed from the borough map that the inner London boroughs are expected to experience a higher number of pedestrian-involved crashes than outer London boroughs, with Westminster having the highest number of expected crashes. Westminster is a large, central London borough with many popular tourist attractions and other places of interest. Referring back to Fig. 1, there are notable similarities in the spatial distribution of pedestrian crashes and the number of children in child benefit families which seems reasonable from the results of the model. The heat map of the wards of London shows a similar tendency for a higher number of expected crashes in inner London, but in more detail (higher spatial resolution) with several wards standing out as high-crash locations. Notably there are nine wards which the model indicates would have experienced more than 100 pedestrian crashes over the study period, these being: Marylebone High Street, St James's, West End, Westbourne (Westminster), Bishop's (Lambeth), Hoxton East and Shoreditch (Hackney), Fairfield (Croydon), City of London (shown in heat maps as both a ward and borough, therefore unsurprisingly high expected number of crashes compared to other wards). When comparing Fig. 4 to Fig. 2a there are considerable parallels between high pedestrian crash wards and wards where the crime rate is high. Also, Fig. 2b showing the number of cars per household appears to be fairly close to the opposite of the ward crash frequency map; both of these observations concur with the results of our model.

Table 3
Average marginal effects of the contributory variables.

Multilevel random parameters model	
Variables	Marginal effects
Ln(population walking to work)	−9.86
Number of schools	0.79
Greenspace	−8.47
Cars per household	−8.62
BAME population	9.43
Ln(crime rate)	25.06
Children in child benefit families	1.11
Alcohol expenditure	11.42
Job density	0.30

3.4. Limitations and future research

Some limitations are acknowledged within this study. An extension of the methodology to include heterogeneity in the mean and variance formulations may provide further insights as shown in previous research. Also, with regard to the methodology, issues concerning the multiple areal unit problem with differing levels of aggregation in the data would need to be considered, making an interesting follow-up study to the work presented here. Similarly, in future it would be interesting to conduct an analysis at MSOA (Middle Layer Super Output Area) level and investigate the differences in findings if any. With respect to spatial dependence in the data, while multilevel models are known to account spatial dependence in the data to some extent (Huang and Abdel-Aty, 2010; Dupont et al., 2013), other methods (e.g., conditional autoregressive models) that can accommodate spatial dependency should be investigated. It is interesting to compare the pedestrian exposure used in this study with those proposed in the crash literature, such as trip assignment or walking time (Sze et al., 2019; Li et al., 2020). However, it should be noted that the latter measures often rely on household survey data, which may be outdated. In this study, we considered data, which matches the study period, relating to various measures of area-level percent population walking (see Table 1) as the pedestrian exposure measure. The identified pedestrian exposure measure is suitable as it was among the most impactful variables associated with pedestrian safety; this can be inferred from average marginal effects estimates (see Table 3). Finally, some explanatory variables which may have been desirable for our analysis such as junction density and other network characteristics were not readily available, but would make interesting additions to future research.

4. Conclusion

This paper has examined the impact which a variety of built environment, socio-demographic and exposure variables can have on pedestrian crash frequency in the Greater London area for the period 2016–2018. To investigate these variables a multilevel random parameters negative binomial regression model was used which has been shown in the literature to be an appropriate tool for macroscopic crash frequency models as it can account to some degree for unobserved heterogeneity as well as spatial dependencies in the data.

As well as identifying high-crash locations among London wards and boroughs, this study reveals which factors are associated with higher pedestrian crash frequencies. This can be useful information for policy

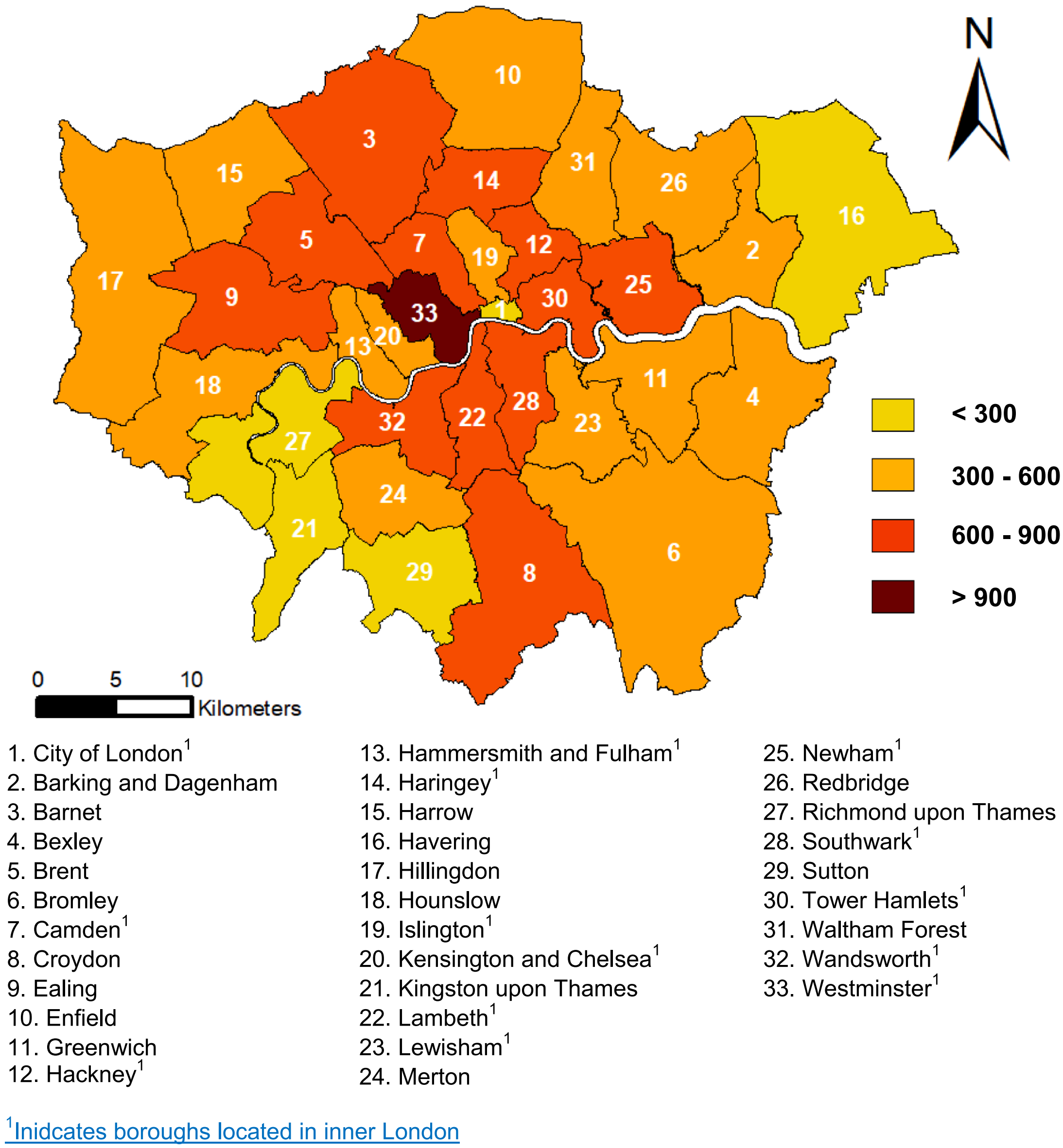

Fig. 3. Heat map showing expected pedestrian crash frequencies (crashes involving a pedestrian casualty) for the boroughs of Greater London over the study period.

makers as interventions can be designed specifically with these factors in mind. It should be noted that the factors listed as significant here are not exhaustive i.e., they are not the only factors which are important in determining pedestrian crash frequency, but they do represent a range of important factors which should be considered when designing interventions. It is also the case that there will be interdependence between these factors and changing one variable may have an effect on the others. This interdependence would make an interesting topic for further study but is deemed to not have a significant effect here.

Interestingly, we found that the proportion of people walking to work was associated with a decrease in pedestrian crash frequency. This initially appears counter intuitive as increased exposure is often associated with increased crash frequency, but this study indicated that areas where more people walk to work are likely to experience fewer crashes. One explanation could be that driver's change their behaviour in the presence of more numerous pedestrians, leading to an increase in pedestrian safety (Jacobsen, 2003). This is an encouraging result for policy makers wishing to promote walking as a travel mode for commuters - if adequate countermeasures are in place. A higher crime rate in a ward was associated with more pedestrian crashes, a relationship which has been found in previous literature (Cottrill and Thakuriah, 2010) but does not seem to have been explored in great detail. Nevertheless, this association provides an additional consideration for policy makers in linking the notions of crime, one of the markers of deprivation, and pedestrian safety.

Of particular note with regard to practical implications is that the number of schools was associated with an increase in pedestrian crash frequency in London over the study period, suggesting that schools decrease pedestrian safety. This indicates a need to focus on countermeasures and adequate infrastructure around schools and education campaigns for those walking to school. The effect of proximity to schools on pedestrian safety has previously been researched by Heydari et al. (2020) finding that proximity to schools decreased pedestrian safety at nearby intersections. The association of greenspace with fewer crashes was also revealed which offers another reason for the importance of devoting land to green spaces in cities alongside advantages in social integration and health promotion (Boulton et al., 2020).

The negative association shown with the number of cars per

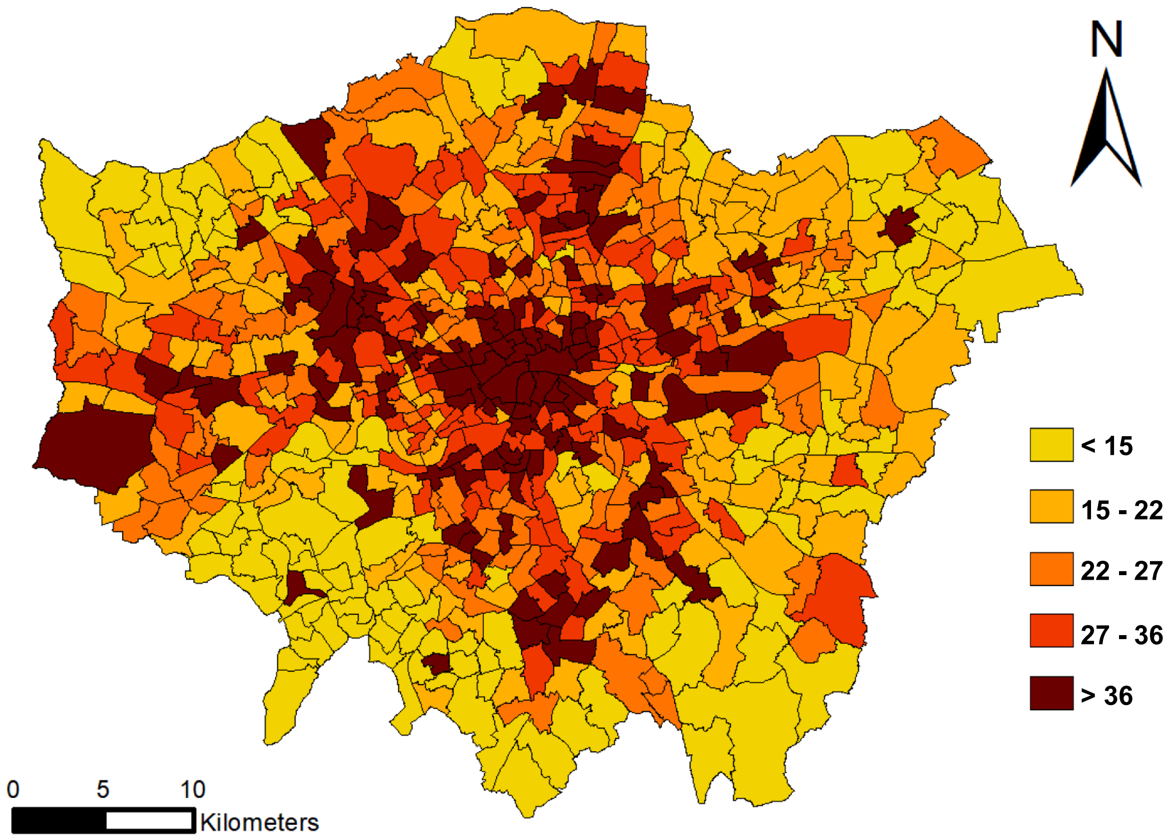

Fig. 4. Heat map showing expected pedestrian crash frequencies (crashes involving a pedestrian casualty) for the wards of Greater London over the study period.

household and positive with the number of children in child benefit households agree with research showing that areas which are more economically deprived experience higher pedestrian crash rates (Bernhardt and Kockelman, 2021). This relationship is one which should be considered by policy makers to identify the causes and address this inequality. The proportion of the population which identifies as BAME was also associated with a greater number of crashes and similarly should be a cause of concern for policy makers. Other macroscopic research has found similar disparities in crash frequency for different areal level proportions of certain ethnicities in the US (Lee et al., 2015) and Hong Kong (Su et al., 2021). An increase in pedestrian crash frequency being associated with both job density and alcohol expenditure demonstrates how activity centres can affect pedestrian safety and in turn the importance of considering this when designing safety interventions. The significance of alcohol expenditure in the model could also allude to issues with pedestrians and/or drivers under the influence of alcohol leading to distraction and reckless behaviour. The significant finding of alcohol expenditure is of particular importance as, while alcohol usage is noted to be a significant factor in crash-level analyses (Garrisson et al., 2021), the inclusion of this variable in macroscopic safety research is rare if non-existent to our knowledge, and represents an important finding for practitioners to consider.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

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

References

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2016. Macroscopic modeling of pedestrian and bicycle crashes: A cross-comparison of estimation methods. Accid. Anal. Prev. 93, 147–159.
Bernhardt, M., Kockelman, K., 2021. An analysis of pedestrian crash trends and contributing factors in Texas. J. Trans. Health 22, 101090.
Boulton, C., Dedekorkut-Howes, A., Holden, M., Byrne, J., 2020. Under pressure: Factors shaping urban greenspace provision in a mid-sized city. Cities 106, 102816.
Bozovic, T., Stewart, T., Hinckson, E., Smith, M., 2021. Clearing the path to transcend barriers to walking: Analysis of associations between perceptions and walking behaviour. Transport. Res. Part F: Traffic Psychol. Behav. 77, 197–208.
Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M., 2016. Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. Accid. Anal. Prev. 93, 14–22.
Chibwe, J., Heydari, S., Faghih Imani, A., Scurtu, A., 2021. An exploratory analysis of the trend in the demand for the London bike-sharing system: From London Olympics to Covid-19 pandemic. Sustainable Cities and Society 69.
Corazza, M.V., D'Alessandro, D., Di Mascio, P., Moretti, L., 2020. Methodology and evidence from a case study in Rome to increase pedestrian safety along home-to-school routes. J. Traffic Transport. Eng. (English Edition) 7 (5), 715–727.
Cottrill, C.D., Thakuriah, P.V., 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. Accid. Anal. Prev. 42 (6), 1718–1728.
DfT (Department for Transport), 2020. Reported road casualties Great Britain, annual report: 2019. Page 10. Available: https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2019. Accessed June 2021.
Ding, H., Sze, N.N., Li, H., Gup, Y., 2020. Roles of infrastructure and land use in bicycle crash exposure and frequency: A case study using Greater London bike sharing data. Accid. Anal. Prev. 144, 105652.
Dupont, E., Papadimitiou, E., Martensen, H., Yannis, G., 2013. Multilevel analysis in road safety research. Accid. Anal. Prev. 60, 402–411.
El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accid. Anal. Prev. 41 (4), 820–828.
Fonseca, F., Papageorgiou, G., Tondelli, S., Ribeiro, P., Conticelli, E., et al., 2022. Perceived Walkability and Respective Urban Determinants: Insights from Bologna and Porto. Sustainability 14, 9089.

Fossum, M., Ryeng, E.O., 2022. Pedestrians' and bicyclists' route choice during winter conditions. Urban, Plann. Trans. Res. 10 (1), 38–57.
Garrisson, H., Scholey, A., Ogden, E., Benson, S., 2021. The effects of alcohol intoxication on cognitive functions critical for driving: A systematic review. Accid. Anal. Prev. 154, 106052.
Graham, D., Glaister, S., Anderson, R., 2005. The effects of area deprivation on the incidence of child and adult pedestrian casualties in England. Accid. Anal. Prev. 37, 125–135.
Graham, D.J., Stephens, D.A., 2008. Decomposing the impact of deprivation on child pedestrian casualties in England. Accid. Anal. Prev. 40, 1351–1364.
Greater London Authority (GLA), 2015a. Ward Profiles and Atlas. Available: https://data.london.gov.uk/dataset/ward-profiles-and-atlas. Accessed November 2021.
Greater London Authority (GLA), 2015b. London Borough Profiles and Atlas. Available: https://data.london.gov.uk/dataset/london-borough-profiles. Accessed November 2021.
Green, J., Muir, H., Maher, M., 2011. Child pedestrian casualties and deprivation. Accid. Anal. Prev. 43 (3), 714–723.
Haas, B., Doumouras, A.G., Gomez, D., de Mestral, C., Boyes, D.M., et al., 2015. Close to home: An analysis of the relationship between location of residence and location of injury. J. Trauma Acute Care Surg. 78 (4), 860–865.
Hauer, E., 1997. Observational Before/After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety, Vol. 1st ed. Pergamon Press.
Heydari, S., Fu, L., Lord, D., Mallick, B., 2016. Multilevel Dirichlet process mixture analysis of railway grade crossing crash data. Analytic Methods in Accident Research 9, 27–43. https://doi.org/10.1016/j.amar.2016.02.001.
Heydari, S., Fu, L., Thakali, L., Joseph, L., 2018. Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: Applications to grade crossing safety analysis. Anal. Methods Acc. Res. 19, 33–48.
Heydari, S., Miranda-Moreno, L., Hickford, A.J., 2020. On the causal effect of proximity to school on pedestrian safety at signalized intersections: A heterogeneous endogenous econometric model. Anal. Methods Acc. Res. 26, 100115.
Hu, L., Wu, X., Huang, J., Peng, Y., Liu, W., 2020. Investigation of clusters and injuries in pedestrian crashes using GIS in Changsha, China. Saf. Sci. 127, 104710.
Huang, H., Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. Accid. Anal. Prev. 42 (6), 1556–1565.
Imprialou, M.M., Quddus, M., Pitfield, D.E., 2015. Multilevel Logistic Regression Modeling for Crash Mapping in Metropolitan Areas. Transport. Res. Board: J. Transport. Res. Board 2514 (1), 39–47.
Jacobsen, P.L., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Injury Prevention 9, 205–209.
Jayakody, A.A., Viner, R.M., Haines, M.M., Bhui, K.S., Head, J.A., et al., 2006. Illicit and traditional drug use among ethnic minority adolescents in East London. Public Health 120 (4), 329–338.
Kraidi, R., Evdorides, H., 2020. Pedestrian safety models for urban environments with high roadside activities. Saf. Sci. 130, 104847.
Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accid. Anal. Prev. 78, 146–154.
Lee, J., Abdel-Aty, M., Cai, Q., 2017. Intersection crash prediction modeling with macro-level data from various geographic units. Accid. Anal. Prev. 102, 213–226.
Li, H., Graham, D.J., Liu, P., 2017. Safety effects of the London cycle superhighways on cycle collisions. Accid. Anal. Prev. 99, 90–101.
Li, H., Wu, D., Graham, D.J., Sze, N.N., 2020. Comparison of exposure in pedestrian crash analyses: A study based on zonal origin-destination survey data. Saf. Sci. 131, 104926.
Lovelace, R., Beecham, R., Heinen, E., Vidal Tortosa, E., Yang, Y., et al., 2020. Is the London Cycle Hire Scheme becoming more inclusive? An evaluation of the shifting spatial distribution of uptake based on 70 million trips. Transport. Res. Part A: Policy Pract 140, 1–15.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Acc. Res. 11, 1–16.
Pljakića, M., Jovanović, D., Matović, B., Mićić, S., 2019. Macro-level accident modeling in Novi Sad: A spatial regression approach. Accid. Anal. Prev. 132, 105259.
Pulugurtha, S.S., Sambhara, V.R., 2011. Pedestrian crash estimation models for signalized intersections. Accid. Anal. Prev. 43, 439–446.
Quddus, M., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. Accident Analysis and Prevention 40 (4), 1486–1497. https://doi.org/10.1016/j.aap.2008.03.009.
Raveau, S., Guo, Z., Muñoz, J.C., Wilson, N.H.M., 2014. A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. Transport. Res. Part A: Policy Pract. 66, 185–195.
Sacchi, E., El-Basyouny, K., 2018. Multivariate linear intervention models with random parameters to estimate the effectiveness of safety treatments: Case study of intersection device program. Accid. Anal. Prev. 120, 114–121.
Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid. Anal. Prev. 45, 382–391.
Steinbach, R., Green, J., Edwards, P., Grundy, C., 2010. 'Race' or place? Explaining ethnic variations in childhood pedestrian injury rates in London. Health & Place 16 (1), 34–42.
Su, J., Sze, N.N., Bai, L., 2021. A joint probability model for pedestrian crashes at macroscopic level: Roles of environment, traffic, and population characteristics. Accid. Anal. Prev. 150, 105898.
Sze, N.N., Su, J., Bai, L., 2019. Exposure to pedestrian crash based on household survey data: Effect of trip purpose. Accid. Anal. Prev. 128, 17–24.
Transport for London (TfL), 2018. Walking action plan: Making London the world's most walkable city. Available: https://content.tfl.gov.uk/mts-walking-action-plan.pdf. Accessed June 2022.
Transport for London (TfL), 2021. Travel in London Report 14. Available: https://tfl.gov.uk/corporate/publications-and-reports/travel-in-london-reports. Accessed January 2022.
Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Saf. Sci. 50 (4), 1141–1151.
Wang, J., Huang, H., Zeng, Q., 2017. The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian. Accid. Anal. Prev. 98, 223–231.
Wang, X., Yang, J., Lee, C., Ji, Z., You, S., 2016. Macro-level safety analysis of pedestrian crashes in Shanghai, China. Accid. Anal. Prev. 96, 12–21.
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2020. Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall/CRC.
Zhu, M., Li, H., Sze, N.N., Ren, G., 2022. Exploring the impacts of street layout on the frequency of pedestrian crashes: A micro-level study. J. Saf. Res. 81, 91–100.