1   Effects of Changes in Road Network Characteristics on Road Casualties:

2   An Application of Full Bayes Models Using Panel Data

3

4

5   **Haojie Li***

6   **Daniel J. Graham**

7   **Arnab Majumdar**

8   Centre for Transport Studies

9   Dept of Civil and Environmental Engineering

10   Imperial College London

11   SW7 2AZ

12

13

14   *Corresponding author at: Centre for Transport Studies, Imperial College London,
15   London SW7 2AZ, UK.
16   Tel.: +44(0)20 75942706; fax: +44(0)20 75946102.
17   E-mail address: h.li09@imperial.ac.uk
18

19

20

21
22

1

1   **ABSTRACT**
2
3   In order to ensure a high level of road safety, road network planning needs to be based on the

4   best knowledge available of the effects of road design on road safety. In this study, we look

5   into how changes in road network characteristics affect road casualties. An approach based on

6   traffic assignment is proposed in order to estimate the traffic exposure at ward level. We apply

7   a widely used approach for before-after evaluation studies, the Bayesian method. We also use

8   a panel semi-parametric model to estimate the dose-response function for continuous

9   treatment variables. The result suggests that there are more casualties in areas with better

10   connectivity and accessibility, where more attention should be paid to safety countermeasures.

11   **Key words: Road Network; Panel Data; Continuous Treatment**

1. INTRODUCTION

The statistical relationship between road casualties and the characteristics of a road network has been investigated in the literature (e.g. Noland and Quddus, 2004; Wier et al., 2009; Huang et al., 2010; Dumbaugh and Li, 2010; Marshall and Garrick, 2011; Rifaat et al., 2011; Jones et al, 2008; Quddus, 2008). Specifically, road casualties are found to be significantly associated with road network characteristics, such as road length, road density and the number of nodes, i.e. junctions. The objective of this study is to investigate the causal relationship between road traffic crash casualties and road network characteristics, including traditional road network characteristics, the connectivity and accessibility of the road network and the curvature of the road network. A causal framework is established by employing full Bayes models. A panel data semiparametric model is used to estimate dose-response functions for continuous treatment variables.

This research contributes to the literature by addressing two key issues overlooked in previous studies. One constraint in previous research is that, to the best of our knowledge, no panel road network data has been analyzed in the UK. This implies that any variance in road network characteristics over time cannot be controlled for and inferences made on the impacts of the road network on road traffic crash casualties could therefore be biased. In contrast, in this study a panel data set for the road network in England and Wales between the years 2001 to 2010 is used to account for time-invariant heterogeneity in the road network.

Another key issue, which is critical in all analyses relevant to road traffic crash casualties, concerns the exposure variable. In analyses at the disaggregate level, the ideal variable used to control for risk exposure is the annual average daily traffic (AADT) for the unit of interest. In terms of aggregate analysis, however, the AADT is unavailable for an aggregate area and therefore typically proxy variables are used. Various such proxy exposure variables, include the usage of cars (Quddus, 2008), aggregated AADT (Jones et al., 2008; Marshall and Garrick, 2011) and a proxy variable derived from the gravity model using data of population and employment (Graham and Glaister, 2003). When applying these variables as exposure

3

1 variables, however, limitations arise. We will discuss this in detail later. This paper uses the

2 daily traffic trips generated within the study area as an exposure variable, and this is estimated

3 using the origin-destination (OD) data obtained from the Office for National Statistics. A two-

4 stage regression is used for this estimation, whereby the traffic trips are estimated with

5 ordinary least square regression in the first stage and the road casualties are analyzed using a

6 panel generalized linear model in the second stage. To account for the composite uncertainty

7 in the two-stage regression, the bootstrap technique is employed.

8 **2. PREVIOUS RESEARCH**

9 Numerous studies have been conducted to develop accident models at aggregate zonal levels.

10 One of the earliest studies by Levine et al. (1995) develops an aggregate model that relates

11 accidents in zones in Hawaii to various factors. However, the model in this study is linear and

12 inappropriate for accident data analysis. Generalized linear models, such as Poisson and NB

13 regression models, have become an prevalent approach to accommodating nonnegative

14 discrete data and have been widely used to relate accident frequencies to various

15 characteristics at zonal levels (Hadayeghi et al., 2003; Lovegrove and Sayed, 2007; Abdel-

16 Aty et al., 2011). For example, Hadayeghi et al. (2003) apply a series of macro-level crash

17 prediction models to relate the number of crashes to zonal characteristics in planning zones in

18 the city of Toronto. Recently, several studies have applied such zone-level models to analyse

19 the relationship between road casualties and road network characteristics.

20 A spatially disaggregated analysis of road casualties in England undertaken by Noland and

21 Quddus (2004) examined the effects of road characteristics and land use on road casualties.

22 Their results suggest that increased length of B roads could increase serious injuries,

23 irrespective of its length, although the coefficients for other types of road were not significant.

24 Marshall and Garrick (2011) investigated how street network characteristics affected road

25 safety in 24 Californian cities from 1997 to 2007. Street network characteristics, such as street

26 network density, street connectivity and street network patterns were controlled for in this

27 study and will be discussed in detail in the following sections. Marshall and Garrick's results

4

suggest that road casualties for all levels of crash severity are correlated with street network characteristics. A higher density of intersection counts (i.e. nodes) is associated with fewer crashes, while street connectivity (link to node ratio) is positively related to crashes. Dumbaugh and Li (2011) also find that miles of arterial roadways and numbers of four-leg intersections were major crash risk factors in Texas, using data from 2003 to 2007. Another study conducted by Rifaat et al. (2011) examined the effect on crash severity of different street patterns, including grid-iron, loops and lollipops, and mixed patterns. Although only pedestrian and bicycle crash data were analysed, the authors found significant effects of street pattern, road features and environmental conditions on the number of crashes. One limitation of all the above research is that they fail to examine the spatial correlation of road traffic crashes. According to LeSage (1998), two problems arise when sample data has a locational dimension: (1) spatial correlation exists between the observations, and (2) spatial heterogeneity occurs in the relationships that are modelled. Traditional zone-level models could violate the traditional Gauss-Markov assumptions, and hence lead to biased inferences.

If spatial correlation is present, appropriate spatial models need to be employed to account for the spatial dependence. Generally, there are two methods for the spatial analysis of road casualty data, traditional econometric models and Bayesian hierarchical models. Although the results from both models are very similar in many cases, it is indicated that inferences derived from traditional spatial models could be misleading when the spatial unit of analysis becomes smaller (Bhati, 2005). Bayesian hierarchical models have been suggested to be more appropriate and have been employed in many studies (Persaud et al., 2010; Aguero-Valverde and Jovanis, 2006; Haque et al., 2010). For example, Aguero-Valverde and Jovanis (2006) compared full Bayes (FB) hierarchical models with traditional negative binomial (NB) models using county-level crash data for Pennsylvania. The existence of spatial correlation in county-level crash data was revealed, although their results from FB hierarchical models were generally consistent with the NB estimates. This similarity in results from the NB and FB models has also been found in other studies (Quddus, 2008; Mitra and Washington, 2007).

This is because uncorrelated heterogeneity accounts for most of the variation and the traditional NB models can sufficiently control for this effect. Another probable reason is that when there is sufficient data, the NB method works well and the results from the two methods are similar.

The idea that the values of parameters could arise from distributions is a fundamental feature of Bayesian methods. Bayesian hierarchical models can accommodate distributions such as the hierarchical Poisson-Gamma distribution and the Poisson-Lognormal distribution (Miaou and Lord, 2003; El-Basyouny and Sayed, 2011; Siddiqui et al., 2012; Yanmaz-Tuzel and Ozbay, 2010). Different prior distributions have been discussed by Yanmaz-Tuzel and Ozbay (2010) with their results suggesting that a Poisson-Lognormal model structure with more informative priors and higher levels of hierarchy may reduce the biases in modeling parameters, hence leading to more robust estimates. In this study, hierarchical Poisson-Lognormal models are adopted to compare with the traditional NB models.

Spatial models have been also applied to analyse the relationship between road casualties and road network characteristics in recent years. A study by Jones et al. (2008) used district-level data to investigate the effects of various factors on traffic casualties in England and Wales. The authors found that traffic casualties were significantly related to road length, curvature, junction density and other geographical variables. However, no positive spatial autocorrelation existed at the district level. The study by Quddus (2008), though did find a significant positive spatial correlation among ward-level road traffic crashes in Greater London from 2002 to 2004. Quddus applied both traditional and spatial models with the results from the traditional NB models and the Bayesian hierarchical models being very similar in suggesting that road traffic crashes are associated with the road infrastructure, socioeconomic and traffic conditions. Substantial positive spatial correlation was also found in the analysis of crash data for Florida's 67 counties from 2003 to 2007 (Huang et al., 2010). One reason for the diverse results of spatial correlation tests could be due to the different levels of spatial aggregation in the papers mentioned above.

1   In addition to spatial correlation, two issues evident in previous studies examining the

2   relationship between the road network and road casualties remain to be adequately addressed.

3   One critical issue is the selection of appropriate traffic exposure variables. In analysis at the

4   unit-level, where the study object is usually road sections or intersections, the AADT or

5   vehicle miles travelled (VMT) is preferred as the traffic exposure variable (Huang et al.,

6   2010; Marshall and Garrick, 2011; Jones et al., 2008). These variables are not always

7   available when the analysis is conducted at the area-level, however. Although proxy variables

8   for traffic exposure have been developed (Graham and Glaister, 2003), there are certain

9   limitations to their use. In this study, a new method to construct the traffic exposure variable

10   is proposed.

11   The other issue concerns the usage of data for road network characteristics. Detailed data

12   regarding the road network, including road class, road length and nodes information can be

13   obtained from OS Meridian ^{TM} 2. Although this data set has been used in several studies in the

14   UK (Noland and Quddus, 2004; Haynes et al., 2007; Graham and Stephans, 2008; Jones et al.,

15   2008), the data availability is only for a single year and, consequently, the variance in road

16   network over time cannot be accounted for. To overcome this problem, OS Meridian ^{TM} 2 for

17   2001 to 2010 (except for 2005) is employed in this study. The data set is discussed in the next

18   section.

19   **3. DATA**

20   **3.1 Dependent Variable**

21   The data used in this analysis includes road casualties recorded in England and Wales from

22   2001 to 2010. The accident data (i.e. road traffic crash) was collected from the STATS 19 data

23   base and was further classified by severity type. Whilst it is recognized that inaccuracies exist

24   in the STATS 19 data, for the purposes of the research outlined in this paper, it is adequate.

25   The location of an accident was recorded using the British National Grid coordinate system.

26   Each individual accident was located on a map and these casualties were further aggregated at

7

1 the ward level, i.e. the primary unit of British administrative and electoral geography.

2 Geographical Information System (GIS) software, such as MapInfo and Arcmap were used to

3 process the data.

4 **3.2 Exposure Variable**

5 A number of exposure measures have been used in road safety, e.g. road length (Giles-Corti et

6 al., 2011), vehicle-kilometres (Huang et al., 2010), fuel consumption (Cardoso, 2005;

7 Fridstorm, 1999), driver population (Quddus, 2008), time spent in traffic (Santamarina-Rubio

8 et al., 2014), and quasi-induced exposure measures (Jiang et al., 2014). Most exposure

9 measures, however, are not applicable for the analysis in this paper due to data availability

10 and quality, as well as the context of the analysis. A review study on exposure data by

11 Papadimitriou et al. (2013) suggests that the most appropriate measure of exposure is VMT,

12 which are closer to the theoretical concept of exposure.

13 In practice, because of data unavailability, most studies estimate VMT using traffic counts,

14 e.g. AADT. In previous disaggregated analyses of traffic crashes, the AADT has frequently

15 been used to indicate the traffic exposure level (e.g. Abdel-Aty and Radwan, 2000). The

16 AADT was also employed in recently conducted area level analyses (Huang et al., 2010;

17 Marshall and Garrick, 2011; Jones et al., 2008). In studies conducted by Marshall and Garrick

18 (2011) and Huang et al. (2010), AADT count points were first located on each road and the

19 average AADT was then calculated for each road. The VMT was obtained by multiplying the

20 average AADT of each road by its centreline mile length and all VMT values were summed to

21 form the exposure variable for each study area. In the UK, a similar application of AADT data

22 was conducted by Jones et al. (2008). Traffic count data supplied by the UK Department for

23 Transport classifies the estimate of the average daily count of vehicles into six categories

24 ranging from pedal cycles to heavy goods vehicles at 5982 survey census points. With a grid

25 reference, each point was assigned to districts and the AADT was estimated by taking mean

26 count values across the points located on each road class.

27 While the usage of the AADT data provides substantial information on the traffic flow at an

1   area-level, there are two limitations in this approach. First, the AADT data is usually only

2   available at a limited number of data collection points. The under-representation of roads,

3   especially minor roads, could lead to an underestimation of traffic flow in each area.

4   Secondly, the fact that the AADT data only accounts for motorized vehicle travel means that it

5   cannot fully depict the level of overall activity, which consists of travels by foot, bicycle and

6   other transit modes. Given the promotion and increase in the amount of non-motorized modes

7   in recent years, this is an increasingly important deficiency of the AADT data.

8   Graham and Glaister (2003) developed a proxy variable for traffic exposure using a gravity

9   model. The idea of this approach is that the internal traffic generation of ward i is

10   proportionate to the population (P_i) and the number of employees (E_i) in ward i, while the

11   external traffic generation of ward i is affected by the population (PP_i) and employee (PE_i) of

12   its proximate ward j ($i \neq j$). The variable, proximate population and employees (PPE_i), which

13   reflects the external generation of ward i can be constructed as follows:

14    $$PPE_i = PP_i + PE_i = \sum_j^{i \neq j} \frac{P_j}{d_{ij}} + \sum_j^{i \neq j} \frac{E_j}{d_{ij}} = \sum_j^{i \neq j} \frac{P_j + E_j}{d_{ij}} \qquad (1)$$

15   where d_{ij} is the centroid distance from ward i to ward j. This variable can reflect the traffic

16   level to a certain extent. One question with this method, however, is how "proximate wards"

17   are defined.

18   A further issue with previous studies is that the proxy variables for traffic exposure are

19   estimated using deterministic models, e.g. the gravity model by Graham and Glaister (2003).
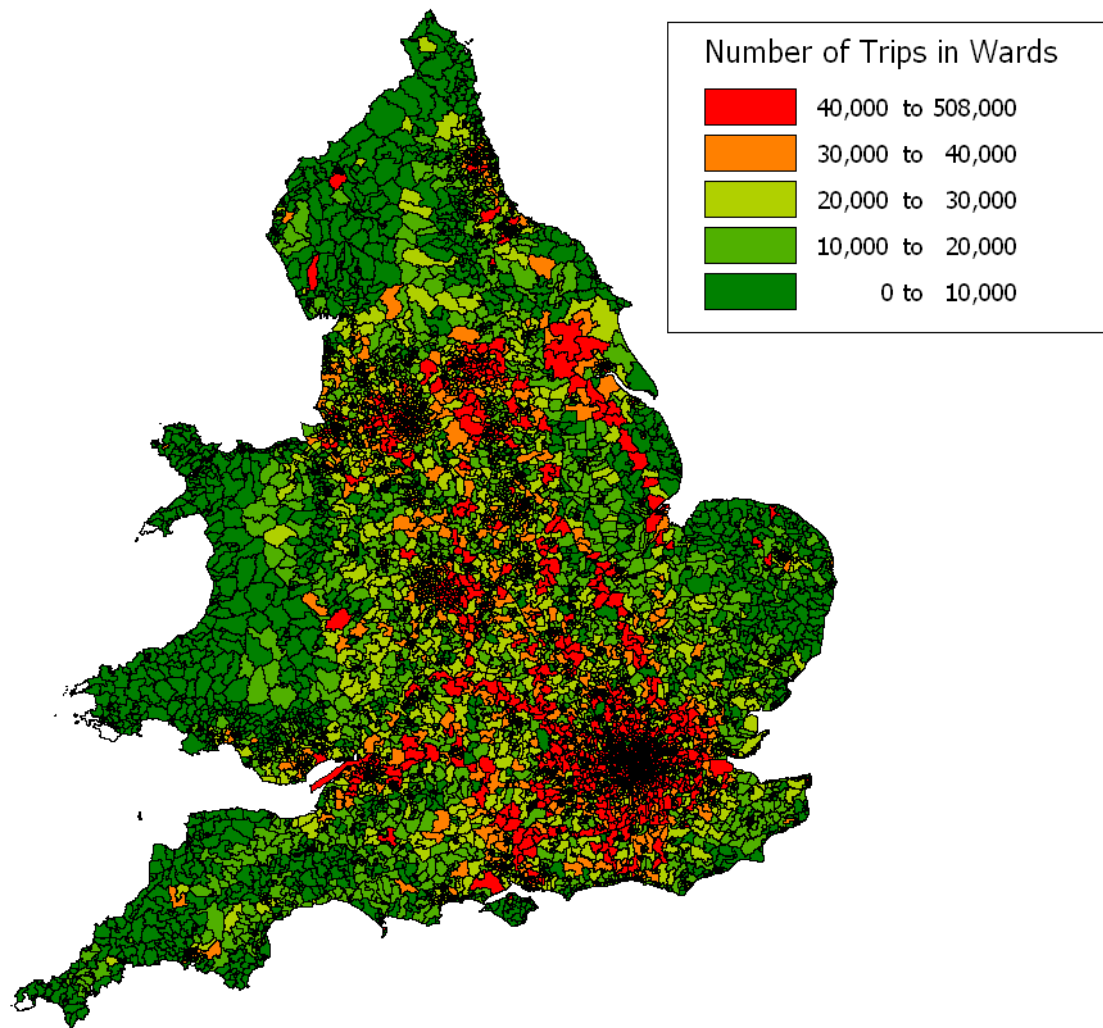
20   A deterministic model assumes that its outcome is certain if the input to the model is fixed. In

21   a dynamic transportation system, however, a completely deterministic model is unlikely to

22   account for uncertainties or random effects, which are closely related to spatial-temporal

23   variations in individual characteristics. In contrast, this study applies a two-stage regression

24   with the bootstrap method to estimate the traffic flow data. This method is superior to a

25   deterministic model since the former accounts for the uncertainty due to random effects. A

26   detailed discussion of this issue is provided later.

27   In this study, an approach based on traffic assignment is proposed in order to estimate the

traffic exposure at ward level. The idea underlying this method is that trips generated between origin-destination (OD) pairs are assigned to transportation networks and aggregated in each ward. The traffic assignment focuses on the selection of routes between OD and the traffic volume on each route in transportation networks. The centroids of wards are treated as origins and destinations, and transportation networks are constructed by links among the centroids. In the conventional transport forecasting model, traffic assignment is the fourth step following trip generation, trip distribution and mode choice. In this research, a wealth of information about OD statistics was obtained from the Office for National Statistics. In this data set, the daily trips of residents or workers from residence to workplace are provided as matrices, and discriminated according to the characteristics of the people and transit modes. To reflect the overall activities within each ward, the daily trips of all people and transit modes were included in the OD matrices.

The traffic assignment was implemented using TransCAD. The assignment method employed was the well-known user equilibrium method proposed by Wardrop (1952). Equilibrium model is widely used for the prediction of traffic patterns in transportation networks. It is assumed that travellers will choose the shortest path from origin to destination and network equilibrium occurs when no traveller can decrease his/her travel cost by shifting to a new path. The trips are assigned to links among the centroids and aggregated in each of the wards. The distribution of the number of trips in wards across England and Wales is presented in Figure 1. Subsequently the number of total trips is used as the proxy variable for traffic exposure. A word of caution needs to be made with this approach since the OD data was only available for 2004. The traffic exposure for other years is estimated by ordinary least square regression, discussed in following sections.

The Department for Transport published Road Transport Forecasts which presents the forecasting results for traffic demand in England up to 2040 (DfT, 2013). However this method is not used to estimate the traffic flow in this study because it forecasts long-term trends rather than individual years, e.g. in five year intervals.

Number of Trips in Wards

(red)	40,000 to 508,000
(orange)	30,000 to 40,000
(yellow-green)	20,000 to 30,000
(green)	10,000 to 20,000
(dark green)	0 to 10,000

1
2 **FIGURE 1 The distribution of trips in wards in England and Wales.**

3 **3.3 Road Characteristics**

4 A major contribution of this paper is that panel data of the road network was used to account

5 for effects due to the variation in road characteristics over time. Compared to time-series and

6 cross-section data, panel data provides several benefits., e.g. it is able to control for spatial-

7 and time-invariant individual heterogeneity, and also provides more informative data thus

8 better enabling the study of the dynamics of adjustment.

9 In this study, detailed information regarding the road network was obtained from Ordnance

10 Survey (OS) Meridian [TM] 2 for the period from 2001 to 2010, excluding for 2005. A set of

11 variables was employed to describe the characteristics of the road network at ward level.

12 (1) Traditional road network characteristics. The length, as well as the density, of the road

13 network was calculated according to road class as defined in the Design Manual for Roads

11

1 and Bridges (DMRB) (Department for Transport, http://www.dft.gov.uk/ha/standards/dmrb/.),

2 e.g. Motorway, A road, B road, Minor road. Road network nodes were defined as meeting

3 points of two or more roads, i.e. junctions. The total number and density of nodes and

4 roundabouts was also calculated.

5 (2) Connectivity and accessibility of the road network. It has been suggested that the degree

6 of connectivity and accessibility of a road network can influence the number of crashes

7 (Marshall and Garrick, 2011). Various measures of road network connectivity and

8 accessibility are utilized in the existing literatures (Dill, 2004; Tal and Handy, 2012; Wang et

9 al., 2013; Koohsari et al., 2014; Owen and Wong, 2013; Giles-Corti et al., 2011). In general,

10 measures used in previous studies include block length, block size, intersection density,

11 percent four way intersections, street density, connected intersection ratio, percentage of dead

12 ends, link node ratio, gamma index, alpha index, nearest neighbour, pedestrian walkable area,

13 and walking distance to activities. However, some measures are not applicable in this study.

14 For instance, measures, such as block length, block size and nearest neighbour, are focusing

15 on the street network at neighbourhood or block level. Such measures are not available when

16 studying nationwide road network. In addition, some variables, e.g. pedestrian walkable area

17 and walking distance to activities, are used to measure pedestrian activities and are dependent

18 of origins and destinations.

19 The measures used in this study were the link-to-node ratio and the density of dead ends

20 (Chin et al., 2008; Dill, 2004; Marshall and Garrick, 2011). The link-to-node ratio was

21 calculated by dividing the number of links by the number of nodes. A high link-to-node value

22 indicates a more connected road network than one with a low link-to-node value. A node with

23 only one link, also known as a dead end, is usually associated with a residential area. Dead

24 ends usually indicate barriers preventing direct routes and low street connectivity (Tal and

25 Handy, 2012).

26 (3) Curvature of the road network. Horizontal road curvature, as measured by vertices per km,

27 has been suggested as an important factor influencing road casualties (Haynes et al., 2007;

1 Jones et al., 2008; Quddus, 2008). The literature indicates that straighter roads have more

2 crashes than roads with more horizontal curves, i.e. bends. The variable used in this research

3 to measure curvature was also the number of vertices per km, with the number obtained using

4 ArcGIS and divided by the road length in each ward. Transition curves were not analysed as

5 research indicates that their impact on crashes is neutral. It is worth noting that the impact of

6 vertical curvature has not been studied in any great detail. Accidents increase with gradient

7 and down gradients have significantly higher accident rates than up gradients. However, the

8 crash implications of steep gradients may not be very severe since steeper gradients may not

9 be very severe. The geometry of vertical curves is not known to have a significant effect on

10 accident severity. The National Cooperative Highway Research Program (NCHRP) Synthesis

11 432 provides an overview of the research on vertical curve safety in North America, with the

12 emphasis on such aspects as the driver's view and headlamp angles, though no use is made of

13 crash analysis (NCHRP, 2012). Given the relatively minor impact of vertical curves compared

14 to horizontal curves, the former were not analysed in this study.

15 **3.4 Socio-demographic Characteristics**

16 Previous research has suggested an association between road traffic crashes and socio-

17 demographic characteristics, such as population, employment and deprivation (Wier et al.,

18 2009; Dissanayake et al., 2009). In particular, a positive relationship has been found in

19 relation to the size of the population and the level of employment, which implies that more

20 casualties may occur in areas with more residents and job opportunities. To consider this

21 effect, the data for population and employment at the ward level was obtained from the Office

22 for National Statistics (ONS).

23 Recent research also suggests that child injuries are influenced by factors related to area

24 deprivation (Graham and Stephens, 2008). Therefore, the Index of Multiple Deprivation (IMD)

25 was used as a control variable in this study. The Index of Multiple Deprivation integrates data

26 on the following seven deprivation domain indices into one overall deprivation score: income,

27 employment, housing and services, health, education, crime and the environment.

4. METHOD

A two-stage regression method was employed to explore the relationships between crash counts and road network characteristics. In the first stage, the traffic exposure for all years of the analysis was estimated using OLS regression. In the second stage, the fitted traffic exposure values were included, together with other variables, in a Generalized Linear Model (GLM). One important issue that needed to be addressed was that the standard errors from the second stage regression were biased because the traffic exposure is itself estimated. To correct the standard errors, therefore, the bootstrap approach was used in both stages.

4.1 Model for estimating travel activities

4.1.1 Within-ward travel activities

The first step involved identification of the relevant variables to be used in the model for estimating travel activities. As discussed previously, trips in each ward consisted of traffic generated within the ward and traffic passing through. There were also two main types of trips categorized by destinations: home-end and work-end trips as set out below.

For trips generated within the ward, the total number of trips leaving or returning to homes or work places is related to the population, employment, deprivation and total length of roads. Home-end trips can be described as a function of:

Home-end Trips= f (Population, Deprivation, Length of roads)   (2)

While work-end trips can be described as a function of:

Work-end Trips = f (Employment, Deprivation, Length of roads)   (3)

The total trips generated in the ward can thus be described as

Within-Ward Trips = Home-End Trips + Work-End Trips= f (Population, Employments, Deprivation, Length of roads)   (4)

4.1.2 Pass-through travel activities

In terms of pass-through trips, Graham and Glaister (2003) employed the idea of using the resident and employment population of proximate wards. A gravity model for trip distribution

1   was applied in their study. It was assumed that pass-through trips are related to trips generated

2   in proximate wards, but what these are is not clearly defined in the literature.

3   A study conducted by Dent and Bond (2008) investigated the commuting patterns of part-time

4   and full-time workers in the UK. One important finding reveals that the average commuting

5   distances in the UK were 7.5 km for part-time workers and 13 km for full-time workers

6   respectively. The average commuting distance given in the report DfT (2011), meanwhile,

7   was 8.6 miles for 2009.

8   In this study, wards within a certain distance called the "bandwidth", were taken into account

9   when selecting proximate wards from which to estimate the pass-through trips. The

10   bandwidth selected was 13 km, which is consistent with the average commuting distance for

11   full time workers.

12   The function of pass-through trips can be described as follows:

13   *Pass-Through Trips = f (Length of roads, Sum of Employments in Neighbour Wards, Sum of*

14   *Population in Neighbour Wards)*   *(5)*

15   To account for both in-ward and pass-through trips, the model used to estimate total trips was

16   developed as follows:

17    $T_i = \alpha + \beta_1 IMD_i + \beta_2 RL_i + \beta_3 E_i + \beta_4 P_i + \beta_5 \Sigma ((E_j + P_j)/D_{ij})$ *(6)*

18   where T_i is the total trips in ward i, IMD_i is the IMD score for ward i, RL_i is the length of roads

19   for ward i, E_i is the employments of ward i, P_i is the population of ward i. Ward j is a

20   neighbour ward of ward I within the bandwidth. Both the resident and employment population

21   of neighbour wards are indicated by E_j and P_j. D_{ij} is the distance between centroids of wards i

22   and j.

23   **4.2 Bootstrapping Generalized Linear Models**

24   In the two-stage regression, the traffic exposure was estimated using OLS and included as the

25   exposure variable in the GLM for accident analysis. It is commonly the case that the standard

26   errors from the second-stage regression are incorrect since the traffic exposure variable is

27   itself estimated. To correct the standard errors for the two-stage model, the bootstrap method

1   was used. The key analogy of bootstrap is "The Population is to the sample as the sample is to

2   the bootstrap sample" (see Fox, 2008, pp. 590). Bootstrap estimates relevant characteristics of

3   the population using the sample data. The sampling distribution of a statistic is then

4   constructed empirically by resampling from the sample. The resampling procedure parallels

5   the process by which sample observations were drawn from the population.

6   We calculated standard errors by adding a bootstrap on the traffic exposure regressions. That

7   is, we repeatedly estimated the traffic exposure regressions with bootstrapped samples,

8   calculated the residuals, and estimated the GLM for accident analysis with the new residuals.

9   The variance in the GLM estimates over the bootstrapped samples was added to the sampling

10   variance that is calculated for the GLM under fixed explanatory variables. This method can be

11   applied to produce accurate confidence intervals, standard errors and hypothesis tests (For

12   detailed discussion, please refer to Freedman, 1984). Let S_1 and S_2 denote the sample used in

13   the first and second regression respectively. The algorithm for bootstrapping the two-stage

14   regression in this study is introduced below.

15   (1)   Use S_1 to estimate the traffic exposure regression, *Equation (6)*,

16           $T_i = \alpha + \beta_1 IMD_i + \beta_2 RL_i + \beta_3 E_i + \beta_4 P_i + \beta_5 \Sigma ((E_j + P_j)/D_{ij})$, and store the estimates of

17   coefficients α', β_1', ..., β_5'.

18   (2)   Take a bootstrap sample of S_2 and call it S_{b2}.

19   (3)   Use S_{b2} to calculate and store the fitted value of traffic exposure $\widehat{T_i}$:

20           $\widehat{T_i} = \alpha' + \beta_1' IMD_i + \beta_2' RL_i + \beta_3' E_i + \beta_4' P_i + \beta_5' \Sigma ((E_j + P_j)/D_{ij})$ *(7)*

21   (4)   Regress crash counts on $\widehat{T_i}$ and other observed covariates in S_{b2} to estimate

22   the vector of coefficient $\boldsymbol{\pi}$ and standard errors.

23   (5)   Repeat steps (2)-(4) 1000times.

24   (6)   Compute the standard errors from the sampling distribution of the estimates.

16

1 There are two sources of random variation in terms of bootstrap inferences. Firstly, almost all

2 the variation among bootstrap distributions comes from the selection of the original sample

3 from the population. This variation can be reduced, however, using a large original sample.

4 Secondly, bootstrap resampling randomly from the original sample may introduce additional

5 variation. Usually, a bootstrap resampling process using 1000 or more resamples are required

6 to reduce additional variation.

7 **4.3 Bayesian Spatial Model**

8 To account for possible spatial autocorrelation of crash counts among adjacent wards,

9 Bayesian hierarchical models were employed as a comparison with the conventional NB

10 models. In this model, the area-specific random effects are decomposed into two components.

11 The first component models the effects that vary in a structured manner in space, such as

12 correlated heterogeneity, while the second models the effects that vary in an unstructured way

13 between areas, such as uncorrelated heterogeneity.

14 The Bayesian hierarchical model can be described as:

15 $Y_i \sim \text{Poisson}(\mu_i)$,

16 $\ln\mu_i = \alpha + \boldsymbol{\beta}\mathbf{X_i} + \varepsilon_i + u_i$ *(8)*

17 where u_i is the spatial correlated heterogeneity. The conditional autoregressive (CAR) prior

18 proposed by Besag et al. (1991) is presented below:

19 $[u_i|u_j, i{\neq}j, \tau^2_u] \sim N(\bar{u}_i, \tau^2_i)$

20 where:

21 $\bar{u}_i = \frac{1}{\sum_j \omega_{ij}} \sum_j u_j \omega_{ij}$ *(9)*

22 $\tau^2_i = \frac{\tau^2_u}{\sum_j \omega_{ij}}$ *(10)*

23 $\omega_{ij} = 1$ if i, j are adjacent, otherwise 0.

24 Both $\tau^2_\varepsilon, \tau^2_u$ are assigned as gamma distributions with priors $Ga \sim (0.5, 0.0005)$ as suggested

25 by Wakefield et al. (2000). The vector of coefficients $\boldsymbol{\beta}$ is assumed as a highly non-

17

1 informative normal distribution with N (0, 0.01) and the intercept α is assigned as a uniform

2 prior distribution to reflect the lack of precise knowledge of the value of the coefficients.

3 Two separate Markov Chain Monte Carlo (MCMC) analyses with different initial values were

4 used to assure convergence. The first 5000 samples were discarded as a burn-in and a further

5 20000 iterations were run for each chain. The overall goodness of fit was measured by the

6 Deviance Information Criterion (DIC), which is a generalization of the Akaike Information

7 Criterion (AIC).

8 **4.4 Dose-Response Function for Continuous Variables**

9 In the binary treatment case, the treatment variable is denoted as $D = \{0, 1\}$. The model can

10 be described as:

11 $Y_i = \alpha + \boldsymbol{\beta X_i} + \gamma d_i, d_i \in D$ *(11)*

12 The effect of the binary treatment variable D can be interpreted as the estimate of the

13 coefficient, γ.

14 $\hat{\gamma} = E\{E[Y_i(1)|X_i] - E[Y_i(0)|X_i]\} = E[Y(1) - Y(0) |X] = \gamma$

15 Such an estimate, however, is inappropriate in the continuous treatment case, where D is

16 allowed to be a continuous interval $[d_1, d_2]$. The relationship between the outcomes and the

17 treatment cannot be simply expressed as a linear relationship.

18 In this study, a panel data semiparametric model is applied to estimate dose-response

19 functions for continuous variables.

20 $Y_{it} = \alpha_i + \boldsymbol{\beta X_{it}} + f(d_{it}) + \varepsilon_{it}, d_{it} \in D$ *(12)*

21 where $f(d_{it})$ is assumed as a polynomial function, which can be expressed as

22 $f(d_{it}) = \mu_0 + \mu_1 d_{it} + \mu_2 d_{it}^2 + \mu_3 d_{it}^3 \ldots + \mu_m d_{it}^m$ *(13)*

23 The best fitting power, m, is selected by maximizing the likelihood of *Equation (13)*. This can

24 be done using the program **lpoly** in STATA.

25 The effect on road casualties of changes in the connectivity and accessibility of road networks

26 is of interest in this study. Specifically, *density of dead end* and *links per node* were treated as

18

1   continuous treatment variables. The dose-response functions of these two treatments were

2   estimated separately by holding other control variables constant and shown by spline curves.

3   **5.   RESULTS**

4   **5.1 Estimation of Traffic Exposure at Ward-Level**

5   In section 3, the traffic exposure in each ward was estimated using OD data. Since data was

6   only available for 2004, however, the traffic exposure for other years was estimated using

7   OLS regression. Predictor variables included in the regression were population, employment,

8   road length, IMD score and the employment and populations of neighbour wards.

9   In order to assess how well the traffic exposure for other years are likely to be predicted by

10   the model, two measures were used: the adjusted R^2 value of the model and the signs and

11   values of the regressors. Table 1 shows the results from the model with the adjusted R^2 value

12   of 0.81 indicating that the predictors sufficiently explained the variability in the data set. The

13   employment and population in neighbour wards are positively associated with the traffic

14   exposure. This is consistent with our assumption that pass-through trips need to be accounted

15   for when predicting the traffic exposure. In terms of within-ward trips, the employment

16   population was positively related to the traffic exposure, while the effect of the resident

17   population was less significant. This is probably because trips recorded in the OD data set are

18   mostly commuting trips. It is not surprising, therefore, that employment population is more

19   significant than the resident population in this model. Another finding is that less traffic

20   activities occur in deprived areas with higher IMD scores.

21

1 **TABLE 1 Model For the Traffic Exposure Regression**

	Coef.	Std.Err.	t	P>|t|	[95% Conf. Interval]	
Resident Population	-0.099	0.045	-2.2	0.028	-0.188	-0.011
Road length	0.104	0.006	18.88	0	0.093	0.115
IMD score	-61.685	14.163	-4.36	0	-89.448	-33.921
Employment Population	0.874	0.029	29.78	0	0.816	0.931
Sum of Pop and Emp in Neighbor Wards	0.098	0.001	124.25	0	0.096	0.099
Constant	-3003.12	368.619	-8.15	0	-3725.725	-2280.515
Observations=6952						
Adj R-squared=0.81						

2

3 **5.2 Estimation Results**

4 As discussed in section 4, the traffic exposure data estimated at the first stage was then

5 employed as the traffic exposure in the analysis of road casualties. Considering the fact that

6 the traffic exposure data was itself estimated, the bootstrapping approach was applied to

7 correct the standard errors. Table 2 shows the regression results from bootstrapped models for

8 total casualties, slightly injured casualties and killed and seriously injured casualties

9 respectively. Both standard errors and bootstrapped standard errors are shown, with the latter

10 for most of the variables being slightly larger than the former. This is because the variation of

11 traffic exposure predicted in the first regression was taken into account at the second stage. As

12 expected, the traffic exposure is significantly correlated with casualty numbers in all models.

13 The road network characteristics are divided into three categories: traditional road

14 characteristics of a road network, the degree of connectivity and accessibility, and the

15 horizontal curvature. As suggested in many other studies (e.g. Huang et al., 2010), road length

16 and density are positively associated with road casualties at all severity levels. In terms of

17 nodes, wards with higher node density were found to have fewer casualties for all categories

18 of casualties. This is consistent with previous research (Marshall and Garrick, 2011; Ladron

19 de Guevara et al., 2004).

20 An unexpected result is the positive relationship between the number of roundabouts and road

21 casualties and there are numerous possible explanations for this. This may be due to the fact

22 that roundabouts are relatively scarce compared to the number of nodes, with approximately

20

13000 of the former and 850000 of the latter. Many wards are observed with zero roundabouts. For example, 2907 in 6952 wards have no roundabouts, which may suggest that this variable may not pick up the real effect of roundabout on traffic casualties given high level of spatial aggregation (Quddus, 2008). The positive relationship is also probably associated with different types of road users. For example, cyclists and pedestrian are more often involved in injury crashes at roundabouts (Daniels et al., 2010). Future studies are recommended to fit models for particular user groups separately. It should be noted that the number and nature of crashes at roundabouts depends upon a number of factors. Previous studies indicate that the major physical factors that are statistically significant for crashes at roundabouts are: entry width, circulatory width, entry path radius, approach curvature, and the angle between entries to the roundabout (Bared and Kennedy, 2000; Daniels et al., 2010; Arndt, 1998). However, none of the physical factors important for roundabout crashes in the UK could be extracted and analysed in this paper. Future analyses will seek to enhance the database used in this paper with detailed information on the physical information for each roundabout.

It can be hypothesized that areas with a better-connected road network will have more casualties, because since pedestrians, cyclists and motor vehicles have better accessibility total traffic activities tend to be more frequent. Two variables were used as indicators of road network connectivity: the links per node and the density of nodes with one link (Chin et al., 2008). The results indicate that an increase in the links per node is associated with an increase in the casualty numbers for all severities. Lower densities of nodes with one link, also known as dead ends, usually indicate higher connectivity to streets. The results show that higher densities of dead ends are associated with fewer casualties and this will be discussed in detail in the next section.

There have been very few studies on the effect of horizontal curvature of the road network at aggregate level and, in those that have been conducted, different measures of curvature have been used. This paper used the number of vertices per km, and the results suggest that road

21

networks with a greater degree of horizontal curvature, i.e. more vertices per km, are associated with fewer crashes for all severity levels. This result is consistent with previous findings (e.g. Jones et al., 2008; Quddus, 2008) and in line with the fundamental physical forces and their impacts associated with vehicles travelling around a bend. The exact mechanisms underlying crashes on horizontal curved sections are complex. Several related mechanism might explain the protective effect of road curvature: reduction in speed, increased driver vigilance, and discouragement of risk-taking behaviour (Haynes et al., 2007). Lower speeds are to be expected on roads with frequent turns, and lower speeds are related to fewer crashes and less serious outcomes (Taylor et al., 2000). It is also likely that drivers on straight roads may underestimate the level of effort required to maintain optimal performance compared with those driving on curved roads (Mathews and Desmond, 2002). If driving conditions are perceived to be easy, this could reduce drivers' attention levels as well as encouraging underestimation of risks, leading to an increase in collisions (Mahalel and Szternfeld, 1986).

TABLE 2 Bootstrapped Models Results

	Total				Slightly Injured				Killed and Seriously Injured			
	Coef.	Std. Err.	Bootstrap Std. Err		Coef.	Std. Err.	Bootstrap Std. Err.		Coef.	Std. Err.	Bootstrap Std. Err.	
Motorway	7.40E-02	2.35E-03	4.66E-03	*	7.72E-02	2.42E-03	2.58E-03	*	5.06E-02	4.99E-03	6.06E-03	*
Aroad	4.02E-02	8.15E-04	8.36E-04	*	3.73E-02	8.42E-04	8.76E-04	*	5.04E-02	1.89E-03	1.98E-03	*
Broad	1.95E-02	8.45E-04	8.52E-04	*	1.72E-02	8.73E-04	9.05E-04	*	2.89E-02	2.00E-03	2.02E-03	*
Minor road	8.45E-04	4.01E-04	4.04E-04	*	---	---	---		5.86E-03	8.50E-04	8.81E-04	*
Number of Nodes	1.22E+00	3.32E-02	3.35E-02	*	1.18E+00	3.42E-02	3.53E-02	*	1.27E+00	6.84E-02	7.10E-02	*
Vertices Density	-2.55E-01	1.15E-02	1.50E-02	*	-2.34E-01	1.19E-02	1.27E-02	*	-3.12E-01	2.37E-02	2.43E-02	*
Motorway density	1.45E-01	2.18E-02	2.37E-02	*	1.53E-01	2.25E-02	2.46E-02	*	---	---	---	---
Aroad density	2.28E-01	6.11E-03	6.23E-03	*	2.42E-01	6.28E-03	6.33E-03	*	1.62E-01	1.39E-02	1.63E-02	*
Broad density	9.36E-02	7.79E-03	7.99E-03	*	1.05E-01	8.02E-03	8.27E-03	*	4.90E-02	1.86E-02	2.03E-02	*
Minor road density	3.86E-02	2.57E-03	2.66E-03	*	4.46E-02	2.64E-03	2.85E-03	*	1.80E-02	6.01E-03	6.44E-03	*
Nodes density	-6.93E-03	3.56E-04	4.65E-04	*	-7.60E-03	3.66E-04	3.79E-04	*	-5.14E-03	8.28E-04	8.30E-04	*
Number of Roundabouts	5.01E-02	1.47E-03	1.63E-03	*	5.35E-02	1.51E-03	1.27E-03	*	7.51E-03	3.00E-03	3.30E-03	*
Density of Dead Ends	-1.82E-02	7.96E-04	7.71E-04	*	-1.89E-02	8.25E-04	8.27E-04	*	-8.91E-03	1.60E-03	1.65E-03	*
Links per Node	1.15E+00	2.36E-02	2.39E-02	*	1.17E+00	2.48E-02	2.48E-02	*	9.49E-01	5.00E-02	5.22E-02	*
Traffic Exposure	9.06E-03	1.32E-04	2.75E-04	*	8.64E-03	1.35E-04	1.66E-04	*	9.08E-03	2.95E-04	3.69E-04	*
Constant	-1.36E+00	6.26E-02	6.27E-02	*	-1.58E+00	6.57E-02	6.57E-02	*	-6.69E-01	1.43E-01	1.43E-01	*

2   *Significant at 99% level

3   **Significant at 95% level

To account for the extra variation due to spatial dependence among the observations, full Bayesian models are applied and compared with the traditional NB models. As discussed in section 2, the candidate models include the following: Poisson-Lognormal model, Poisson-Lognormal with random effects and Poisson-Lognormal with random effects and spatial effects. DIC values are used to compare different Bayesian models. As expected the Poisson-Lognormal model accounting for spatial correlation exhibited the lowest DIC value, indicating that it performs the best among all the Bayesian models.

The results from Bayesian spatial models and traditional NB models are very similar. This is probably because the extra variation was largely due to area-specific heterogeneity, which was controlled for in both models. Results show that increased number of vertices per km, density of nodes, and density of dead ends are associated with reduced road casualties, while there were positive relationships between road casualties and other factors, such as road length, road density, traffic exposure and the ratio of link-to-node.

1   **TABLE 3 Full Bayesian Models Results**

Model	PL		PL with Random Effects		PL with Random Effects and Spatial Effects	
	Total Casualties					
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Motorway	4.60E-02	6.10E-04	5.28E-02	8.04E-04	5.28E-02	7.98E-04
Aroad	3.24E-02	2.74E-04	3.38E-02	3.51E-04	3.38E-02	3.51E-04
Broad	2.07E-02	3.19E-04	1.65E-02	3.94E-04	1.65E-02	3.90E-04
Minor road	5.05E-03	1.12E-04	4.18E-03	1.93E-04	4.16E-03	1.64E-04
Number of Nodes	1.30E+00	8.78E-03	9.69E-01	1.61E-02	9.67E-01	1.19E-02
Vertices Density	-3.55E-01	2.99E-03	-2.48E-01	5.47E-03	-2.46E-01	4.00E-03
Motorway density	1.32E-01	6.88E-03	7.28E-02	8.23E-03	7.39E-02	8.16E-03
Aroad density	2.18E-01	1.90E-03	2.02E-01	2.34E-03	2.03E-01	2.31E-03
Broad density	8.99E-02	2.73E-03	9.46E-02	3.28E-03	9.56E-02	3.29E-03
Minor road density	5.94E-02	7.41E-04	2.93E-02	1.26E-03	3.02E-02	1.15E-03
Nodes density	-7.80E-03	1.08E-04	-2.16E-03	1.59E-04	-2.20E-03	1.49E-04
Number of Roundabouts	3.61E-02	4.19E-04	3.65E-02	5.43E-04	3.68E-02	5.24E-04
Density of Dead Ends	-1.72E-02	1.96E-04	-1.59E-02	3.23E-04	-1.58E-02	2.92E-04
Links per Node	1.43E+00	4.88E-03	1.08E+00	1.31E-02	1.07E+00	1.04E-02
Traffic Exposure	4.44E-03	2.50E-05	2.39E-03	5.27E-05	2.34E-03	5.20E-05
Constant	-1.80E+00	1.31E-02	-9.07E-01	3.57E-02	-8.80E-01	2.83E-02
DIC Value	20916		17562		17283	

2
3
4
5
6

1 TABLE 3 Full Bayesian Models Results (Continued)

	Slightly Injured					
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Motorway	4.71E-02	6.64E-04	5.63E-02	8.66E-04	5.61E-02	8.62E-04
Aroad	3.03E-02	3.14E-04	3.23E-02	3.75E-04	3.22E-02	3.82E-04
Broad	1.91E-02	3.45E-04	1.54E-02	4.22E-04	1.53E-02	4.19E-04
Minor road	4.77E-03	1.50E-04	3.92E-03	1.83E-04	3.83E-03	1.82E-04
Number of Nodes	1.33E+00	1.30E-02	1.01E+00	1.40E-02	9.92E-01	1.23E-02
Vertices Density	-3.61E-01	4.55E-03	-2.56E-01	4.70E-03	-2.51E-01	4.40E-03
Motorway density	1.40E-01	7.14E-03	6.64E-02	8.71E-03	6.80E-02	9.00E-03
Aroad density	2.25E-01	1.99E-03	2.11E-01	2.53E-03	2.12E-01	2.49E-03
Broad density	9.79E-02	2.93E-03	1.00E-01	3.55E-03	1.01E-01	3.51E-03
Minor road density	6.14E-02	9.17E-04	3.32E-02	1.25E-03	3.41E-02	1.27E-03
Nodes density	-7.98E-03	1.20E-04	-2.53E-03	1.57E-04	-2.56E-03	1.58E-04
Number of Roundabouts	3.86E-02	4.49E-04	3.83E-02	5.76E-04	3.86E-02	5.72E-04
Density of Dead Ends	-1.82E-02	2.36E-04	-1.68E-02	3.69E-04	-1.68E-02	3.20E-04
Links per Node	1.48E+00	1.12E-02	1.13E+00	1.57E-02	1.12E+00	1.18E-02
Traffic Exposure	4.24E-03	2.58E-05	2.36E-03	5.61E-05	2.31E-03	5.65E-05
Constant	-2.09E+00	2.98E-02	-1.20E+00	4.27E-02	-1.18E+00	3.32E-02
DIC Value	17584		15495		15275	

2
3
4
5
6
7
8

1 **TABLE 3 Full Bayesian Models Results (Continued)**

| | Killed and Seriously Injured | | | | | |
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Motorway	4.16E-02	1.57E-03	3.75E-02	1.99E-03	3.71E-02	1.96E-03
Aroad	4.37E-02	6.95E-04	4.23E-02	8.27E-04	4.18E-02	7.98E-04
Broad	2.92E-02	7.72E-04	2.33E-02	9.23E-04	2.29E-02	8.74E-04
Minor road	7.22E-03	3.56E-04	6.23E-03	5.01E-04	5.91E-03	4.07E-04
Number of Nodes	1.19E+00	3.10E-02	9.25E-01	4.30E-02	8.90E-01	3.33E-02
Vertices Density	-3.44E-01	1.06E-02	-2.56E-01	1.50E-02	-2.43E-01	1.11E-02
Motorway density	4.78E-02	1.92E-02	8.67E-02	2.27E-02	9.00E-02	2.27E-02
Aroad density	1.77E-01	5.34E-03	1.58E-01	6.02E-03	1.59E-01	6.06E-03
Broad density	4.26E-02	7.59E-03	5.49E-02	8.87E-03	5.72E-02	8.77E-03
Minor road density	4.48E-02	2.15E-03	1.50E-02	2.97E-03	1.58E-02	2.87E-03
Nodes density	-6.80E-03	3.02E-04	-2.03E-03	3.93E-04	-1.89E-03	3.86E-04
Number of Roundabouts	1.90E-02	1.21E-03	2.16E-02	1.44E-03	2.30E-02	1.46E-03
Density of Dead Ends	-1.18E-02	5.88E-04	-1.23E-02	7.15E-04	-1.22E-02	7.81E-04
Links per Node	1.17E+00	1.58E-02	9.70E-01	2.22E-02	9.37E-01	2.24E-02
Traffic Exposure	5.76E-03	6.47E-05	4.46E-03	1.22E-04	4.13E-03	1.25E-04
Constant	-2.98E+00	4.27E-02	-2.49E+00	6.03E-02	-2.41E+00	6.16E-02
DIC Value	10997		8228		8070	

2

5.3 Dose-Response Functions for Continuous Variables

As discussed previously, it may be inappropriate to assume a linear relationship between the outcomes and the treatment in the continuous treatment case, where the treatment effect heterogeneity has become a primary interest of researchers. The panel data semi-parametric model was applied in this study to investigate how the treatment effect varied over the treatment doses.

Figure 2 displays the dose-response curves for links per node and density of dead end, with the shaded sections indicating the 95% confidence interval. It can be seen that there is an increased trend in the total casualties over the value of links per node, which is consistent with the results obtained from both the bootstrapped model and the FB model. It is obvious, however, that the relationship does not follow a linear trend. The treatment effect remains largely unchanged under the value of 2.5 of links per node, whilst above this value there is a significant increase. Regarding the marginal effect of the dead end density on total casualties, it is evident that there are more casualties in an area with lower dead end density. Specifically, the number of total casualties reaches a peak when the dead end density is near zero, which indicates a highly connected and accessible road network.

Both treatment variables indicate the degree of the connectivity and complexity of the road network. The figures show that a better connected and more complex road network (i.e. with a higher value of links per node and lower dead end density) may experience more crashes. Given this, then it is imperative to ensure the education of driver behaviour is such that crashes are minimised on such complex networks.
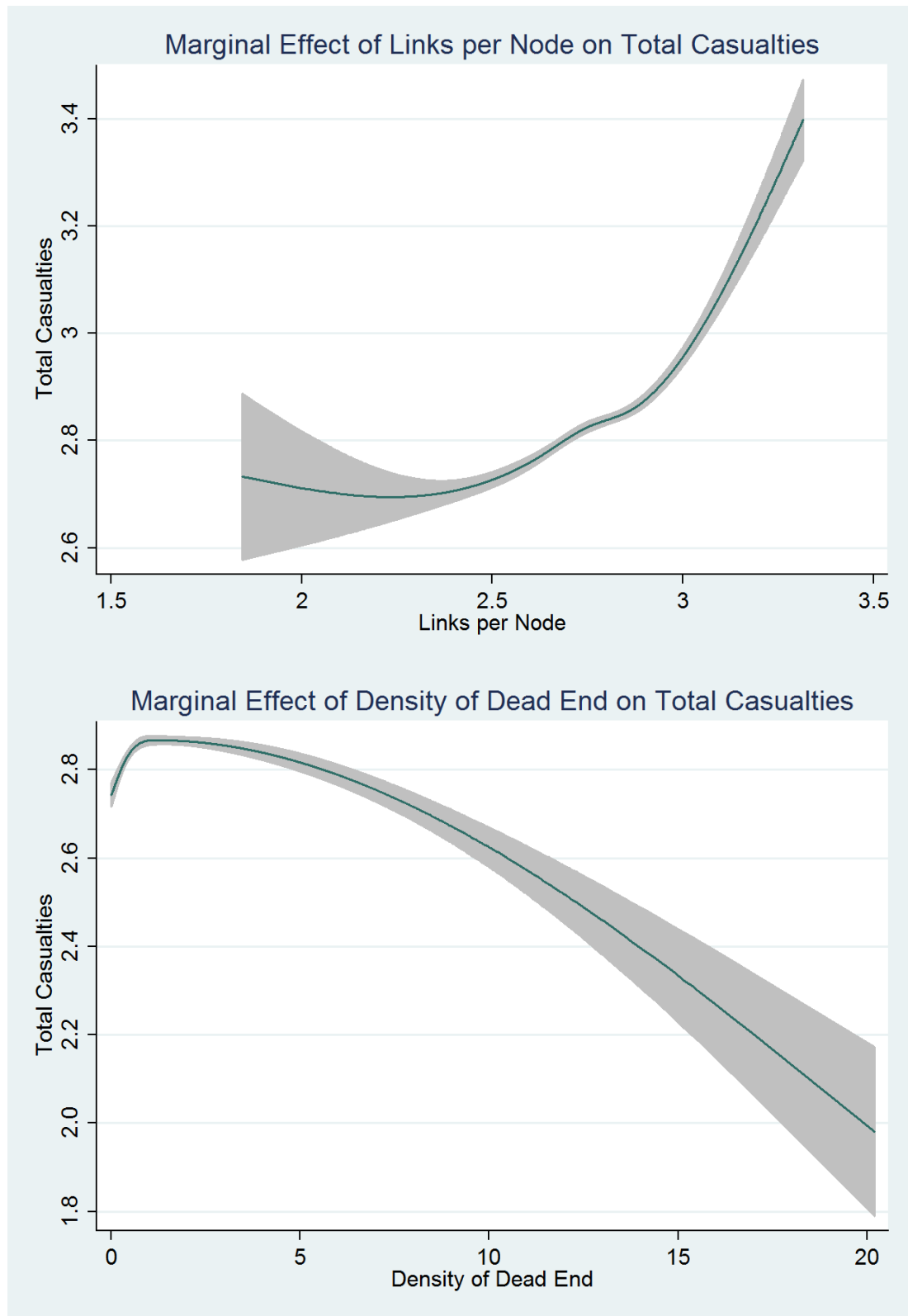
FIGURE 2 **Dose-response functions for total casualties.**

6. CONCLUSIONS

In this study NB models and full Bayesian models were employed to analyse the relationships

between road casualties and various road network characteristics in the UK. The results are

29

1   consistent with the general conclusions of previous research in this field. This study

2   contributes to the literature by addressing two outstanding issues in previous research.

3   Several studies at an aggregation level, concerned with analysing the effects of road network

4   factors on road casualties, have been conducted in UK. One common problem with these

5   studies is that the data for the road network is not longitudinal due to the poor availability of

6   data. In this research, OS Meridian^{TM} 2 from 2001-2010 was obtained, which enables this

7   study to control for the variation in the road network across time.

8   The traffic exposure is a critical issue in road accident analysis; however, there is a lack of

9   appropriate variables that can be used for traffic exposure at an aggregation level. In this

10 research, a method is proposed for constructing traffic exposure at the ward-level based on

11 OD data. A synthetic road network was first built among wards and OD trips were assigned to

12 this network using traffic assignment method. Trips were aggregated at the ward-level and

13 fitted by a regression model. The high adjusted R^2 value indicates that the traffic exposure

14 data was well fitted by the model, which validates the predictions of the traffic exposure. To

15 control for the additional variation due to the traffic exposure data, a two-stage bootstrapped

16 model was developed. Most variables in this model were still significant despite a slight

17 increase in standard errors.

18 Various factors related to road networks have been proposed previously. To the best of the

19 author's knowledge, however, these have not been examined simultaneously at the ward-level.

20 By using GIS software, such as ArcGIS, this study has been able to generate the potential

21 explanatory variables at the ward-level. In particular, these variables are divided into three

22 categories: (1) traditional road characteristics, such as road length and road class; (2)

23 accessibility and connectivity of road network; (3) curvature of road. Applying the panel data

24 for the road network, most of the findings are in line with previous research. Marshall and

25 Garrick (2011) include dead ends as explanatory variable in their model, however, no

26 significant relationship is found. In this research, it has been shown that the accessibility and

27 connectivity of road network is an important factor affecting road casualties using ratio of

30

node to link and density of dead end as measures. Furthermore, a panel semiparametric model was used to estimate the dose-response functions of these two continuous treatments to show the heterogeneity of treatment effects.

In summary, devising a safe road network is a difficult task, especially when the existing network has been in use for a long time. However, a better knowledge of the impacts of road network design on road safety can provide both useful policy implications and also help to improve the design standards. For instance, in this study, the results show that a road network with high connectivity and accessibility has more casualties. Although the road network cannot be changed in the short term, e.g. the design life of roads in the United Kingdom is typically 40 years, it is possible for policy makers to enhance road safety countermeasures in areas with such road networks. Design standards can also be enhanced given the findings in this paper, e.g. the findings on horizontal curvature can be considered in ensuring minimum radii for such curves according to the type of road.

REFERENCES

Abdel-Aty, M. A., Radwan, A. E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention, 32, pp. 633-642.

Abdel-Aty, M. A., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating trip and roadway characteristics to manage safety in traffic analysis zones. Transportation Research Record, no. 2213, pp. 20-28.

Aguero-Valverde, J., Jovanis, P. P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention, 38, pp. 618-625.

Arndt, O., 1998. Road Design Incorporating Three Fundamental Safety Parameters. Technology Transfer Forum 5 and 6, Transport Technology Division, Main Roads Department, Queensland, Australia.

Bared, J. G., Kennedy, K., 2000. Safety Impacts of Modern Roundabouts. The Traffic Safety Toolbox: A Primer on Traffic Safety, Chapter 28, Institute of Transportation Engineers.

1 Besag, J., York, J., Mollie, A., 1991. Bayesian image restoration with two applications in

2 spatial statistics. Annals of the Institute of Statistical Mathematics, 43, pp. 1-59.

3 Bhati, A. S., 2005. Modeling count outcomes with spatial structures: An information-theoretic

4 approach. Discussion Paper, Justice Policy Center, The Urban Institute, Washington DC.

5 Cardoso, J. L., 2005. The use of international data on fuel sales and vehicle fleet for the

6 estimation of yearly national traffic volumes. Accident Analysis & Prevention, 37 (1), 207-

7 215.

8 Chin, G. K. W., Van Niel, K. P., Giles-Corti, B., Knuiman, M., 2008. Accessibility and

9 connectivity in physical activity studies: The impact of missing pedestrian data. Preventive

10 Medicine, 46, pp. 41-45.

11 Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining variation in safety performance of

12 roundabouts. Accident Analysis and Prevention, 42, pp. 393-402.

13 Dent, A., Bond, S., 2008. An investigation into the location and commuting patterns of part-

14 time and full-time workers in the United Kingdom, using information from the 2001 Census.

15 Office for National Statistics.

16 Department for Transport, 2011. Personal Travel Factsheet.

17 Department for Transport, 2013. Road Transport Forecasts 2013.

18 Dill, J., 2004. Measuring Network Connectivity for Bicycling and Walking. Presented at 83rd

19 Transportation Research Board Annual Meeting, Washington, D. C.

20 Dissanayake, D., Aryaija, J., Wedagama, D. M. P., 2009. Modelling the effects of land use

21 and temporal factors on child pedestrian casualties. Accident Analysis and Prevention, 41, pp.

22 1016-1024.

23 Dumbaugh, E., Li, W., 2010. Designing for the Safety of Pedestrians, Cyclists, and Motorists

24 in Urban Environments. Journal of the American Planning Association, 77:1, pp. 69-88.

25 El-Basyouny, K., Sayed, T., 2011. A full Bayes multivariate intervention model with random

26 parameters among matched pairs for before–after safety evaluation. Accident Analysis and

27 Prevention, 43, pp. 87-94.

1   Fox, J. , 2008. Applied Regression Analysis and Generalized Linear Models. Sage, Thousand

2   Oaks, CA, second edition.

3   Freedman, D., 1984. On Bootstrapping Two-Stage Least-Squares Estimates in Stationary

4   Linear Models. The analysis of statistics, Vol.12, No.3, 827-842.

5   Fridstrom, L., 1999. Econometric Models of Road Use, Accidents, and Road Investment

6   Decisions. TOI Report 457/1999. Institute of Transport Economics, Norway.

7   Gelman, A., Rubin, D. B., 1992. Inference from Iterative Simulation Using Multiple

8   Sequences. Statistical Science, Vol.7, No.4, pp. 457-511.

9   Giles-Corti, B., Wood, G., Pikora, T., Learnihan, V., Bulsara, M., Van Niel, K., Timperio, A.,

10 McCormack, G., Villanueva, K., 2011. School site and the potential to walk to school: The

11 impact of street connectivity and traffic exposure in school neighborhoods. Health & Place,

12 17, pp. 545-550.

13 Graham, D. J., Glaister, S., 2003. Spatial Variation in Road Pedestrian Casualties: The Role

14 of Urban Scale, Density and Land-use Mix. Urban Studies, Vol. 40, No. 8, pp. 1591-1607.

15 Graham, D. J., Stephens, D. A., 2008. Decomposing the impact of deprivation on child

16 pedestrian casualties in England. Accident Analysis and Prevention, 40, pp. 1351-1364.

17 Hadayeghi, A., Shalaby, A. S., Persaud, B. N., 2003. Microlevel accident prediction models

18 for evaluating safety of urban transportation systems. Transportation Research Record, 1840,

19 pp. 87-95.

20 Haque, Md. M., Chin, H. C., Huang, H., 2010. Applying Bayesian hierarchical models to

21 examine motorcycle crashes at signalized intersections. Accident Analysis and Prevention, 42,

22 pp. 203-212.

23 Haynes, R., Jones, A. P., Kennedy, V., Harvey, I., Jewell, A., 2007. District variations in road

24 curvature in England and Wales and their association with road traffic crashes. Environment

25 and Planning A, 39, pp. 1222-1237.

1   Huang, H., Abdel-Aty, M. A., Darwiche, A. L., 2010. County-Level Crash Risk Analysis in

2   Florida Bayesian Spatial Modelling. Transportation Research Record: Journal of the

3   Transportation Research Board, 2148, pp. 27-37.

4   Jiang, X., Lyles, R. W., Guo, R., 2014. A comprehensive review on the quasi-induced

5   exposure technique. Accident Analysis and Prevention, 65, pp. 36-46.

6   Jones, A. P., Haynes, R., Kennedy, V., Harvey, I. M., Jewell, T., Lea, D., 2008. Geographical

7   variations in mortality and morbidity from road traffic casualties in England and Wales.

8   Health & Place, 14, pp. 519-535.

9   Koohsari, M. J., Sugiyama, T., Lamb, K. E., Villanueva, K., Owen, N., 2014. Street

10   connectivity and walking for transport: Role of neighborhood destinations. Preventive

11   Medicine, 66, pp. 118-122.

12   Ladron de Guevara, F., Washington, S. P., Oh, J., 2004. Forecasting Crashes at the Planning

13   Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona.

14   Transportation Research Record: Journal of the Transportation Research Board, 1897, pp.

15   191-199.

16   LeSage, J. P., 1998. Spatial econometric. Working paper, Department of Economics,

17   University of Toledo.

18   Levine, N., Kim, K. E., Nitz, L. H., 1995. Spatial Analysis of Honolulu Motor Vehicle

19   Crashes: I. Spatial Patterns. Accident Analysis and Prevention, Vol. 27, No. 5, pp. 663-674.

20   Li, H., Graham, D. J., Majumdar, A., 2012. The effects of congestion charging on road traffic

21   casualties: A causal analysis using difference-in-difference estimation. Accident Analysis and

22   Prevention, 49, pp. 366–377.

23   Lord, D., Manar, A., Vizioli, A., 2005. Modeling crash-flow-density and crash-flow-V/C ratio

24   relationships for rural and urban freeway segments. Accident Analysis and Prevention, 37, pp.

25   185-199.

Lord, D., van Schalkwyk, I., Chrysler, S., Staplin, L., 2007. A strategy to reduce older driver injuries at intersections using more accommodating roundabout design practices. Accident Analysis and Prevention, 39, pp. 427-432.

Lovegrove, G. R., Sayed, T., 2007. Macrolevel collision prediction models to enhance traditional reactive road safety improvement programs. Transportation Research Record, no. 2019, pp. 65-73.

Mahalel, D., Szternfeld, Z., 1986. Safety improvements and driver perception. Accident Analysis and Prevention, 18, pp. 37-42.

Marshall, W. E., Garrick, N. W., 2011. Does street network design affect traffic safety? Accident Analysis and Prevention, 43, pp. 769-781.

Matthews, G., Desmond, P. A., 2002. Task-induced fatigue states and simulated driving performance. Quarterly Journal of Experimental Psychology, 55A, pp. 659-686.

Miaou, S. P., Lord, D., 2003. Modelling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes. Transportation Research Record, 1840, pp. 31-40.

Mitra, S., Washington, S., 2007. On the nature of over-dispersion in motor vehicle crash prediction models. Accident Analysis and Prevention, 39, pp. 459-468.

National Cooperative Highway Research Program (NCHRP) Synthesis 432: Recent Roadway Geometric Design Research for improved safety operations: A synthesis of highway practice. Transportation Research Board 2012, Washington DC.

Noland, R. B., Quddus, M. A., 2004. A spatially disaggregate analysis of road casualties in England. Accident Analysis and Prevention, 36, pp. 973-984.

Papadimitriou, E., Yannis, G., Bijleveld, F., Cardoso, J. L., 2013. Exposure data and risk indicators for safety performance assessment in Europe. Accident Analysis and Prevention, 60, pp. 371-383.

1 Persaud, B., Lan, B., Lyon, C., Bhim, R., 2010. Comparison of empirical Bayes and full

2 Bayes approaches for before–after road safety evaluations. Accident Analysis and Prevention,

3 42, pp. 38-43.

4 Quddus, M. A., 2008. Modelling area-wide count outcomes with spatial correlation and

5 heterogeneity: An analysis of London crash data. Accident Analysis and Prevention, 40, pp.

6 1486-1497.

7 Owen, K. K., Wong, D. W., 2013. An approach to differentiate informal settlements using

8 spectral, texture, geomorphology and road accessibility metrics. Applied Geography, 38, pp.

9 107-118.

10 Rifaat, S. M., Tay, R., Barros, A., 2011. Effect of street pattern on the severity of crashes

11 involving vulnerable road users. Accident Analysis and Prevention, 43, pp. 276-283.

12 Santamarina-Rubio, E., Perez, K., Olabarria, M., Novoa, A. M., 2014. Gender differences in

13 road traffic injury rate using time travelled as a measure of exposure. Accident Analysis and

14 Prevention, 65, pp. 1-7.

15 Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and

16 bicycle crashes. Accident Analysis and Prevention, 45, pp. 382-391.

17 Tal, G., Handy, S., 2012. Measuring Nonmotorized Accessibility and Connectivity in a

18 Robust Pedestrian Network. Transportation Research Record: Journal of the Transportation

19 Research Board, No. 2299, pp. 48-56.

20 Taylor, M. C., Lyman, D. A., Baruya, A., 2000. The effects of drivers' speed on the frequency

21 of road accidents. TRL Report 421, Crowthorne, Berkshire: Transport Research Laboratory.

22 Wakefield, J. C., Best, N. G., Waller, L., 2000. Bayesian Approaches to Disease Mapping, on

23 Spatial Epidemiology: Methods and Applications. Oxford University Press.

24 Wang, F., Wen, M., Xu, Y., 2013. Population-adjusted street connectivity, urbanicity and risk

25 of obesity in the U.S. Applied Geography, 41, pp. 1-14.

26 Wang, X., Abdel-Aty, M., 2006. Temporal and Spatial Analyses of Rear-end Crashes at

27 Signalized Intersections. Accident Analysis and Prevention, 38, pp. 1137-1150.

1   Wardrop, J. G., 1952. Some theoretical aspects of road traffic research. Proceedings,

2   Institution of Civil Engineers, PART II, Vol.1, pp. 325-378.

3   Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., Bhatia, R., 2009. An area-level model of

4   vehicle-pedestrian injury collisions with implications for land use and transportation planning.

5   Accident Analysis and Prevention, 41, pp. 137-145.

6   Yanmaz-Tuzel, O., Ozbay, K., 2010. A comparative Full Bayesian before-and-after analysis

7   and application to urban road safety countermeasures in New Jersey. Accident Analysis and

8   Prevention, 42, pp. 2099-2107.