Denne artikel er publiceret i det elektroniske tidsskrift
Artikler fra Trafikdage på Aalborg Universitet
(Proceedings from the Annual Transport Conference
at Aalborg University)
ISSN 1603-9696
www.trafikdage.dk/artikelarkiv

Infrastructure and spatial effects on the frequency of cyclist-motorist collisions in the Copenhagen region

Carlo G. Prato, Sigal Kaplan, Thomas K. Rasmussen, Tove Hels
Department of Transport, Technical University of Denmark

Abstract

Promoting cycling aims at reducing congestion and pollution and encouraging healthy and sustainable lifestyles, but generally clashes with the perception of crash risk while riding a bicycle that is still the most significant disincentive to cycling. The current study analyzed the factors contributing to increase crash risk while riding a bicycle by focusing on the variation of 5349 cyclist-motorist collisions within 269 traffic zones in the Copenhagen Region. The model controlled for traffic exposure for both bicycles and motorized transport modes, evaluated the effects of infrastructure and socio-economic characteristics of the zones, and accounted for heterogeneity and spatial correlation across the zones. A Poisson-lognormal model with second-order CAR priors confirmed the existence of the safety in numbers phenomenon, contradicted previous literature about bicycle facilities not being helpful in reducing crash risk, highlighted the need for Copenhagen-style bicycle paths especially in suburban areas, and emphasized how heterogeneity and spatial correlation play a significant role in explaining the probability of cyclist-motorist crash occurrence.

1. Introduction

In a world where natural resources are rapidly depleting and health concerns are constantly growing, promoting bicycle use has been embraced by urban and transport planners as a way to encourage a healthy and environmentally sustainable lifestyle (e.g., Jacobsen, 2003; Chapman, 2007; Elvik, 2009; Vandelbulcke et al., 2009; Buehler et al., 2011; de Nazelle et al., 2011). This lifestyle offers commuters and leisure travelers the possibility of a healtier and more affordable door-to-door transport solution and bears the potential of reducing congestion and pollution that deplete natural resources worldwide (e.g., Pucher et al., 2011a; Pucher et al., 2011b).

Currently, the debate about whether cycling actually provides more benefits or disbenefits is vivid. While some evidence points in the direction of cycling offering advantages because of the physical activity that reduces the health risks inherent to a sedentary life (e.g., de Hartog et al., 2010; Rabl and de Nazelle, 2011; de Geus et al., 2014), other evidence points in the direction of cycling bringing along negative effects because of the exposure to traffic pollution and crash risk (e.g., Aertsens et al., 2010; Int Panis et al., 2010;

Rojas-Rueda et al., 2011). When considering that the crash risk on a bicycle is definitely higher than the one on a motorized vehicle (e.g., Joshi et al., 2001; Reynolds et al., 2009; Wood et al., 2009) and factoring that the crash risk is the most significant disincentive to cycling (e.g., Parkin et al., 2007; Winters et al., 2011; Lawson et al., 2013), investigating the factors contributing to the risk of being involved in a crash while riding a bicycle assumes a fundamental role for urban and transport planners aiming to increase the appeal of their cities and regions to potential cyclists.

The literature on cycling crashes presents several attempts of understanding the factors contributing to crash occurrence. In particular, a great deal of attention has been posed to infrastructure characteristics and specific conditions leading to a higher number of conflicts between cyclists and other road users. Intersections and roundabouts have been often identified as blackspots for cyclists, with particular problems in signalized intersections or roundabouts replacing other types of intersections (e.g., Wang and Nihan, 2004; Quddus, 2008; Møller and Hels, 2008; Reynolds et al., 2009). Interruptions to the traffic flow such as the presence of public transport stops or the adoption of traffic calming measures have been generally identified as problematic for cyclists, with specific hindrances linked to the interaction with other road users (e.g., Quddus, 2008; Chong et al., 2010; Pei et al., 2010). Cycling facilities have been related to higher cycling crash occurrence and these findings have been interpreted as on-road cycling being safer, with specific problems related to bidirectional facilities and discontinuous ones (e.g., Rodgers, 1997; Aultman-Hall and Hall, 1998: Pucher et al., 1999). Traffic congestion has been analyzed with respect to cycling crash occurrence via surrogate measures such as measuring rates in peak or off-peak hours, considering speed limits, and observing traffic composition, with identifiable problems related both to higher traffic volumes in the peak hours and higher speeds in the off-peak hours (e.g., Li et al., 2007; Wang et al., 2009; Yiannakoulias et al., 2012; Vandenbulcke et al., 2014).

Limitations are identifiable in the existing literature on cycling crashes. Firstly, existing research efforts generally examined the correlation of crash occurrence with one specific factor and hence missed on the multifaceted and interrelated nature of the contribution of different factors to the number of cycling crashes. Secondly, existing research endeavors largely overlooked the heterogeneity of the zone or road where the crashes happened as well as the spatial correlation effects that evidently exist as crashes are more frequent in certain locations. Thirdly, existing research studies ignored the exposure in terms of traffic and hence missed on correcting for the amount of cyclists in the zone or road where the crashes took place while likely introducing bias in the estimates of the effects related to crash occurrence. The current study overcomes these limitations in the existing literature by proposing a comprehensive analysis of the frequency of cyclist-motorist crashes that considers heterogeneity and spatial correlation across zones.

The current study estimated a crash frequency model of 5349 cyclist-motorist collisions occurred in the Copenhagen Region between 2009 and 2013. Crash data were available from the National Crash Database maintained by the Danish Road Directorate and compiled from police records, and their geographical coordinates were matched to the road network of the study region. The model formulation considered traffic zones within the region as the level of aggregation of the crash counts and accounted for both heterogeneity and spatial correlation across the zones. The model specification focused on the infrastructure and the characteristics of the traffic zones in the region, but most relevantly corrected for the traffic exposure of bicycles, cars, vans and heavy vehicles. In fact, the availability of a large dataset of actual travel behavior from the on-going national travel survey and the accessibility to traffic estimates from the regional and national transport models allowed overcoming the main limitation of crash occurrence studies not correcting for traffic exposure. The estimation of Poisson-gamma and Poisson-lognormal models with conditional autoregressive (CAR) priors within a full hierarchical Bayesian framework allowed answering the requirements for the formulation and the specification of the model.

Analyzing cyclist-motorist crashes in the Copenhagen Region has several advantages. Firstly, it allows providing an outer marker for emphasizing safety issues and suggesting traffic and policy measures that are relevant for the realization of cycling safety in Denmark as well as in other countries (see, e.g., Pucher and Buehler, 2008). Established cycling cultures could find interesting looking at the Danish experience and evaluating similarities and differences while estimating a similar model. Emerging cycling cultures could find inspiration in the Danish example and reflect upon policies bearing the potential to reduce the probability of cyclist-motorist crashes. Secondly, it allows suggesting the type of information that is necessary for analyzing cyclist-motorist collisions and properly designing policies with the aim of promoting bicycle use, while providing the availability of detailed information about exposure and hence avoiding to obtain biased estimates that might lead to counterintuitive results. Thirdly, it answers a specific need for a Danish society where the safety of cyclists is at the top of the priority list when considering cycling strategies being designed and implemented at the municipal, regional and national level.

The remainder of this paper is structured as follows. The next section describes the data collection effort and introduces the model formulation and specification. Then, model estimates are presented and discussed with respect to findings from the existing literature. Last, conclusions are drawn and further research direction are proposed.

2. Methods

2.1 Data

The current study focused on 5349 cyclist-motorist collisions that occurred in the Copenhagen Region from 2009 to 2013. Crashes were extracted from the National Crash Database that the Danish Road Directorate compiles from police reports. These reports enclose information about the characteristics of the crash, the injured persons and the involved vehicles (for details, see, e.g., Kaplan and Prato, 2013; Kaplan et al., 2014). Most importantly, these reports contain a georeferenced location of the crashes that enables matching them with the road network.

The selection of collisions between cyclists and motorists was motivated by the possible bias that would have been introduced by considering also single-bicycle crashes or collisions between cyclists and vulnerable road users that are subject to a high degree of underreporting (see, e.g., Veisten et al., 2007; Aertsens et al., 2010; Juhra et al., 2012). Collisions between cyclists and motorists are far more reported also because insurance claims require the police to be called. The choice of a five-year period was induced by the need for having a long enough period to assemble a sample of adequate size and a short enough period to limit changes in road and traffic conditions. The focus of the study on the Copenhagen Region was prompted by the need for a very detailed road network and a calibrated transport model for planning purposes in order to be able to consider infrastructure characteristics and traffic volumes within the model specification.

The geocoded crashes were matched to the bicycle network of the Copenhagen Region. This is a high-resolution network of 110,893 nodes and 272,586 links that was constructed from a variety of sources for planning purposes. The network contains paths only accessible by bicycles and pedestrians, bicycle lanes and paths alongside roads for motorized traffic, bicycle paths along motorways and expressways, and all the road network available to motorists in the region. The network covers an area of about 3000 sq.km. and a population of about 2 million inhabitants divided in 18 municipalities and 269 traffic zones according to the zone system of the Danish national transport model (LandsTrafikModel, LTM). The information about the traffic on the network was obtained by working on three sources: the national travel survey

(TransportvaneUndersøgelse, TU), the regional transport model (ØresundTrafikModel, OTM), and the LTM. The TU is an on-going collection of 1000 travel diaries per month of a representative sample of the Danish population between 10 and 84 years old. The OTM is the transport planning model traditionally used for traffic evaluations in the Copenhagen Region. The LTM is the transport planning model that will be used for traffic and transport policy assessments in Denmark. Both the OTM and the LTM use the TU travel diaries for the construction of the origin-destination matrices and the estimation of the demand model. The current study used the OTM matrices for cars, vans, heavy vehicles and bicycles in order to assign the inter-zonal traffic on the network, and enriched the information with the TU travel diaries in order to assign the bicycle intra-zonal traffic on the network. In fact, considering intra-zonal traffic was very relevant to properly assess the bicycle traffic as traditionally the distances covered by bicycle are lower than the ones traveled by motorized transport modes. The current study then used the traffic zones from the LTM because of the higher level of detail about the zones in which the Copenhagen Region is divided.

In light of the described data availability, the variables for the model specification were constructed as follows: (i) crash data were considered as count variables representing the number of crashes in a spatial unit, and this unit corresponded to the LTM traffic zones; (ii) infrastructure data for each LTM traffic zone were obtained by overlaying the network data with the traffic zone boundaries; (iii) traffic exposure was retrieved by taking the average daily traffic (vehicle × km for cars, vans, heavy vehicles and bicycles) at the link level and then aggregating the values at the LTM traffic zone level. While a limitation of the current study might consist in considering the average daily traffic over the five-year period, it is offset by the benefit of considering traffic exposure within crash frequency models. Notably, the traffic exposure was validated against a large database of traffic counts and the variations in the counts over the period were not conspicuous, especially for the cycling traffic that has not shown significant alterations in the period between 2000 and 2011 (City of Copenhagen, 2012).

Table 1 - Sample characteristics

Variable	mean	st.dev.	min	25% pct.	median	755 pct.	max
Number of crashes	19.9	22.7	0	6	13	26	194
Population (unit)	6867.7	4624.1	0	3590	6059	9472	30190
Area (sq.km.)	11.5	18.0	0.161	1.391	4.253	13.982	123.579
Average income (1000 DKK)	228.8	50.5	85.5	199.3	220.2	249	501.2
Full-time employed (unit)	3931	2784	0	2025	3468	5416	18342
Part-time employed (unit)	163	125	0	87	138	194	822
Students or pupils (unit)	1263	837	0	688	1128	1769	5435
Retired (unit)	1139	737	0	535	1066	1622	3104
Unemployed seeking job (unit)	117	160	0	25	59	144	1092
Unemployed not seeking job (unit)	255	227	0	96	200	348	1577
Road without bicycle lane (km)	58.760	58.378	1.025	16.996	44.567	79.875	403.113
Road with bicycle lane (km)	6.496	5.103	0.798	3.120	5.894	8.860	31.600
Road with bicycle path (km)	22.511	21.570	1.241	4.905	18.043	32.518	135.831
Footpath segregated (km)	2.217	2.574	0.162	0.453	1.427	2.828	14.523
Cycling traffic (bicycle × km)	8,633	6,656	137	4,349	7,117	11,281	62,431
Car traffic (vehicle × km)	101,489	101,750	4,438	34,820	62,181	135,637	523,433
Van traffic (vehicle × km)	10,316	11,010	477	3,217	6,113	14,092	62,341
Heavy vehicle traffic (vehicle × km)	5,049	5,891	54	1,355	2,909	6,587	32,197

Table 1 presents the most significant variables in the dataset for model estimation. This study focused on the Copenhagen Region, where the LTM traffic zones are quite heterogeneous in terms of area and population, with smaller zones in the Copenhagen city center and larger zones in the outskirts of the region. The distribution of the average income does not reflect completely the one of the jobs, as higher income is found in the north of the metropolitan area while most jobs are found in the Copenhagen city center (see, e.g., Kaplan et al., 2014). Notably, the development of job opportunities and the expansion of the metropolitan area has followed the transit-oriented fingerplan and accordingly has been around the five train lines from the center of the city (see, e.g., Kaplan et al., 2014). The Copenhagen Region has an extremely developed bicycle infrastructure, especially in the Copenhagen city center and the residential zones that have adopted the typical Copenhagen-style design for bicycle paths (i.e., road-curb-bicycle path-curb-sidewalk). This design has been found to increase the perception of safety with respect to alternative mixed designs and hence to be at the root of the high bicycle market share (Chataway et al., 2014). Cyclists share the road though with motorists, in particular in the outskirts of the metropolitan area and the suburban areas of the region.

2.2 Model

The current study intended to analyze the frequency of collisions between cyclists and motorists by estimating a count data model at the zone level while considering heterogeneity and spatial correlation across the zones. Accordingly, Poisson-based models with heterogeneity effects and conditional autoregressive priors were estimated as recommended in the literature (for a review, see Lord and Mannering, 2010; Mannering and Bhat, 2014).

Considering the crash counts within the traffic zones in the Copenhagen Region, the base form of the Poisson-based model was expressed as:

$$Y_i \sim Poisson(\mu_i) \tag{1}$$

$$\log(\mu_i) = \alpha + \boldsymbol{\beta X}_i + v_i + u_i \tag{2}$$

where Y_i is the observed number of crashes in traffic zone i, μ_i is the expected Poisson crash rate in traffic zone i, X_i is a vector of explanatory variables, α is the intercept to be estimated, $\boldsymbol{\beta}$ is a vector of parameters to be estimated, v_i is a random term that captures the heterogeneity across traffic zones, and u_i is a random term that captures the spatial correlation across traffic zones. The vector X_i of explanatory variables includes the traffic exposure (i.e., cars, vans, trucks, bicycles), the infrastructure characteristics and the zone characteristics. Different model formulations were considered according to different specifications of the error terms v_i and u_i: Poisson-lognormal and Poisson-gamma models were tested to evaluate which distribution of the random term v_i was the most suitable to account for heterogeneity across zones, while two neighboring structures were tested to assess which structure expressed in the random term u_i was the most appropriate to account for spatial correlation across zones.

Both the Poisson-lognormal and the Poisson-gamma formulations assigned a uniform prior distribution to the intercept α and a highly non-informative normal prior to all β's with zero mean and 100,000 variance (see, e.g., Spiegelhalter et al., 2002). In the Poisson-lognormal formulation, the prior distribution for the heterogeneity random term v_i was a normal prior with distribution $N(0,\tau_v^{-1})$, where τ_v is the precision (i.e., the inverse of the variance) with a vague gamma prior $Gamma(0.5,0.001)$. It should be noted that the parameterization of the Gamma distribution $Gamma(a,b)$ has mean a/b and variance a/b^2. In the Poisson-gamma formulation, the prior distribution for the exponential of the heterogeneity random term $exp(v_i)$ was a gamma prior with distribution $Gamma(\phi,\phi)$ where ϕ was assigned to a non-vague hyper prior with $Gamma(0.1,1.0)$.

Both formulations for the correlation structure across zones considered the same priors for the intercept α and the parameters β's, and both considered the two formulations of the heterogeneity random term v_i. The spatial correlation term u_i was represented with a conditional autoregressive model (Besag, 1974):

$$u_i \left| u_j \right., i \neq j \sim N\left(\frac{\sum_j u_j w_{ij}}{w_{i+}}, \frac{\tau_{ij}^2}{w_{i+}} \right) \tag{3}$$

where w_{ij} is the weight between zone i and zone j, w_{i+} is the sum over j of the weights w_{ij}, and τ_{ij}^2 is a scale parameter with a prior with distribution *Gamma*(*0.5,0.001*). The two formulations of the correlation structure were differentiated according to the definition of the weights w_{ij}. In the first formulation, first-order neighbours were defined for traffic zone j sharing a border with traffic zone i with $w_{ij} = 1$, and for traffic zones not sharing any border with $w_{ij} = 0$. In the second formulation, second-order neighbours were also defined for traffic zone j being connected to first-order neighbors of zone i with $w_{ij} = 0.5$, and with $w_{ij} = 0$ otherwise.

Models were estimated with the software package Openbugs (Lunn et al., 2009) by using the Markov Chain Monte Carlo (MCMC) method under the full hierarchical Bayesian framework. As several models were estimated (i.e., the two heterogeneity formulations without spatial correlation, and the four combinations of the formulations for the error terms v_i and u_i), and the deviance information criterion (DIC) was used to compare the goodness-of-fit and select the best model (Spiegelhalter et al., 2002).

3. Results

The specification of the model followed a traditional iterative process looking for the best expression of variables and their transformation in order to explain the variation of the number of collisions between cyclists and motorists in the Copenhagen Region. All the six model formulations were then estimated with the same specification of variables and parameters throughout the iterative process.

The average daily traffic values for bicycles, cars, vans, and heavy vehicles, were transformed in logarithmic scales in order to reduce the large variation of these explanatory variables across zones. The same applied to the number of kilometers of road with and without cycling facilities, whose effect was investigated also relatively to their location in either urban or suburban areas of the Copenhagen Region. The socio-economic characteristics of the zones were also considered in the model specification in order to include proxies relative to the availability of resources for infrastructure development and maintenance, as the income in a zone reflects the taxation resources of the local authorities, and the typology of demand, as the profile of the population in a zone reflects the potential heterogeneity in the demand composition. The various specifications were tested with all six model formulations described in the methodological section, and the comparison of the DIC revealed that Poisson-lognormal models performed better than Poisson-gamma models, and that second-order spatial correlation effects led to better model fits than first-order effects only. Table 2 presents the comparison of the DIC for the six model formulations with the best model specification, with differences that are significant according to the indications by Spiegelhalter et al. (2002).

Table 2 - Comparison of model formulations

Model	Heterogeneity	Spatial correlation	DIC
Aspatial Poisson-lognormal	Lognormal	-	1345.7
Aspatial Poisson-gamma	Gamma	-	1361.8
Spatial Poisson-lognormal	Lognormal	1st order CAR	1325.5
Spatial Poisson-gamma	Gamma	1st order CAR	1337.1
Spatial Poisson-lognormal	Lognormal	2nd order CAR	1305.9
Spatial Poisson-gamma	Gamma	2nd order CAR	1315.8

The posterior means and standard deviations of the intercept α, the β's for the explanatory variables, the standard deviation of the heterogeneity error term v_i and the standard deviation of the spatial correlation error term u_i were estimated using the MCMC method. For all the model formulations and specifications, two chains were simulated with different initial values and the initial 25,000 iterations were discarded as burn-ins to reach the convergence of the two chains. Then, both chains were simulated for other 75,000 iterations with the aim of calculating the posterior means and standard deviations of the estimated parameters for the best model, namely the Poisson-lognormal with second-order CAR priors, that are presented in table 3.

Table 3 - Poisson-lognormal CAR model estimates

Variable	Mean	st.dev.	MC err.	sig.
log (bicycle × km)	0.6690	0.0361	0.0021	**
log (car × km)	0.3490	0.0877	0.0052	**
log (van × km)	0.3150	0.1097	0.0065	**
log (heavy vehicle × km)	0.1746	0.0812	0.0048	**
Copenhagen/Frederiksberg	-0.0492	0.0447	0.0019	
Suburban areas	-0.6141	0.2039	0.0093	**
log (km road without bicycle lane)	0.2755	0.0540	0.0019	**
log (km road with bicycle lane)	0.0565	0.0247	0.0002	**
log (km road with bicycle path)	-0.2928	0.0642	0.0016	**
log (km road without bicycle lane) in suburban areas	0.0412	0.0260	0.0003	
log (km road with bicycle lane) in suburban areas	-0.1230	0.1309	0.0059	**
log (km road with bicycle path) in suburban areas	0.1555	0.0644	0.0006	*
number of intersections	0.0414	0.0188	0.0003	**
number of intersections in suburban areas	-0.0246	0.0116	0.0002	**
log (income)	-0.1499	0.0989	0.0059	*
full-time employed	1.1385	0.3903	0.0231	**
part-time employed	0.0549	0.0306	0.0018	*
students or pupils	-0.4595	0.0902	0.0053	**
Retired	0.1788	0.0604	0.0035	**
population on welfare	0.8984	0.3201	0.0186	**
Constant	-2.5614	1.1020	0.0657	**
st.dev. (v)	0.3160	0.0554	0.0018	**
st.dev. (u)	0.2622	0.0596	0.0020	**

* statistically significant difference from zero (90% credible set shows the same sign)
** statistically significant difference from zero (95% credible set shows the same sign)
mean: mean of the distribution of the estimated parameter; st.dev.: standard deviation of the distribution of the estimated parameter; MC err: Monte Carlo error

The average daily traffic is related to the number of collisions between cyclists and motorists in each zone. Interestingly, the estimated parameter shows a non-linear relationship between the number of crashes and the bicycle × km with a statistically significant difference from 1, as credible sets do not contain 1 neither at the 95% nor at the 90% confidence level. This is an important result because it shows that the crash rates fall as cycling exposure increases, ceteris paribus, most likely because higher numbers of cyclists increase awareness in drivers and hence reduce risk. Also, the estimated parameters show a non-linear relationship between the number of crashes and the vehicle × km for all the types of motorized traffic considered, namely cars, vans and heavy vehicles. This is also a relevant result because it suggests that the crash rates decrease with increasing volumes of traffic, or in other words with increasing congestion.

After controlling for the exposure, the location of the zone is significantly associated with the number of crashes, as more crashes are related to suburban areas. Most relevantly, a higher extension of bicycle facilities is significantly related to a decrease in the number of crashes, in particular when longer bicycle paths that are segregated from the vehicle traffic are constructed. The same effect is not observed for bicycle lanes alongside the roads, although not as remarkedly as for the absence of bicycle facilities, thus confirming that the Copenhagen-style bicycle path design is the safest solution to reduce the risk on the bicycles. Notably, when considering the interaction with the location in urban or suburban areas, it emerges that bicycle paths are less effective in suburban areas. When looking at the socio-economic characteristics of each zone, there is a negative relationship between the number of crashes and the average income, suggesting that a lower number of crashes occurs in zones that are in average richer and hence have higher taxation intake. The number of crashes is also associated with a higher number of potential commuters, identifiable with the number of full-time and part-time workers, a lower number of students, and a higher number of people to various extents being maintained by the welfare systems.

When looking at the heterogeneity and spatial correlation, the posterior mean of the standard deviation of the error term accounting for heterogeneity is statistically significant and suggests that heterogeneity effects are relevant in these crash data. Analogously, the posterior mean of the standard deviation of the error term representing the spatial correlation is statistically significant and indicates that spatial correlation effects play a significant role in these crash data. It should be noted that the best model specification has spatial effects concerning not only the obvious correlation with zones that are direct neighbors, but also the more loose correlation with zones sharing a common neighbor. The ratio between the *st.dev.* (u) of the error term representing the spatial correlation and the sum of the standard deviations *st.dev.* (u) and *st.dev.* (v) of both error terms allows assessing the posterior proportion of model error explained by spatial random effects. The value of this ratio indicates that the spatial correlation accounts for 54.7% of the stochastic variation in the model and hence assumes a slightly more prominent role with respect to the heterogeneity.

4. Discussion and conclusion

The current study analyzed the frequency of cyclist-motorist collisions in the Copenhagen Region by estimating zone-based Poisson-based models accounting for heterogeneity and spatial correlation across zones. In particular, model estimates illustrate the importance of infrastructure effects and the relevance of spatial correlation to explain the variation in the number of crashes within the region studied. The best model for these crash data is a Poisson-lognormal with second-order CAR priors that account for correlation not only between neighboring zones, but also between zones sharing a neighbor.

An extremely relevant finding is the non-linear relationship between crashes and average bicycle daily traffic in the zone. As the traffic is expressed as bicycle × km and the parameter is significantly lower than 1, this non-linearity confirms the safety in numbers hypothesis that the crash rates diminish when the bicycle traffic increases (e.g., Jacobsen, 2003; Elvik, 2009). Having a higher number of cyclists can be motivated from both the cyclists' perspective, namely they feel safer in an environment designed for them that they share with a lot of fellow cyclists, and the drivers' perspective, namely they reach higher awareness because they see cyclists everywhere. Non-linearity is observed also with the vehicle traffic, and again as the traffic is expressed as vehicle × km and the parameter is significantly lower than 1, this non-linearity suggests that the crash rates diminish when vehicular traffic augments. This finding is not surprising, especially when considering that congestion has been generally related to lower crash rates because of the lower speeds and the consequent higher margins of error in recognizing a potential conflict (e.g., Parkin et al., 2007; Møller and Hels, 2008; Reynolds et al., 2009).

Another extremely important finding is the positive correlation between the presence of bicycle infrastructure and the reduction in the number of crashes. Findings from the current study are in disagreement with previous literature showing that cycling in mixed traffic is safer than cycling on bicycle infrastructure (e.g., Rodgers, 1997; Aultman-Hall and Hall, 1998: Pucher et al., 1999). These findings are however in line with recent studies pointing out that bicycle infrastructure not only increases safety (de Rome et al., 2014; Kaplan et al., 2014), but also increases the perception of safety that inherently encourages more people to cycle (Chataway et al., 2014). It should be noted that the estimated model controls for traffic exposure and for the extensive infrastructure in one of the cycling capital regions of the world, while previous studies from the 90's were from regions with very little bicycle traffic that was not controlled for and very limited infrastructure. Accordingly, non-significant relationships between crash rates and bicycle infrastructure could result from lack of observations rather than actual absence of correlation. It should also be noted that the findings of the current study extend the discussion by underlying the importance of the infrastructure in suburban areas where the separation between cyclists and motorists is even more beneficial because of the risk inherent in the higher speeds of the vehicles (e.g., Yiannakoulias et al., 2012; Vandenbulcke et al., 2014). Findings about intersections being problematic are also in general in line with previous research (Wang and Nihan, 2004; Quddus, 2008; Reynolds et al., 2009), and also in this case they extend the current knowledge by emphasizing in particular that the effect is less marked in suburban areas than in urban ones.

The relevance of heterogeneity and spatial correlation is evident from the model estimates, as is the relative importance that shows how the variation of the error term is slightly towards the spatial effects. This confirms the importance of considering spatial correlation within frequency models (see, e.g., Lord and Mannering, 2010; Mannering and Bhat, 2014) and looking for specific solutions in specific locations where more crashes occur in order to achieve a substantial reduction in crash occurrence. Findings from this study underline how promoting bicycle use can create a virtuous circle in which the more people bike and the safer are the roads for cyclists, how the Copenhagen-style bicycle path design is the most effective for achieving a decrease in the number of crashes and hence an increase in encouraging cycling, and how intersection solutions are necessary to reduce the number of conflicts. Future research should investigate whether the possibility of a model estimated at the link level is feasible, and whether a joint model of frequency and severity could help further understanding the determinants of cyclist-motorist collisions.

Acknowledgements and remarks

The authors gratefully acknowledge the financial support to the project "Improving road safety" by the Danish Council for Strategic Research.

References

Aertsens, J., de Geus, B., Vandenbulcke, G., Degraeuwe, B., Broekx, S., De Nocker, L., Liekens, I., Mayeres, I., Meeusen, R., Thomas, I., Torfs, R., Willems, H., Int Panis, L., 2010. Commuting by bike in Belgium, the costs of minor accidents. *Accident Analysis and Prevention*, 42, 2149-2157.

Aultman-Hall, L., Hall, F., 1998. Ottawa-Carleton commuter cyclist on- and off-road incident rates. *Accident Analysis and Prevention*, 30, 29-43.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 36, 192-236.

Buehler, R., Pucher, J., Merom, D., Bauman, A., 2011. Active travel in Germany and the U.S. - contributions of daily walking and cycling to physical activity. *American Journal of Preventive Medicine*, 41, 241-250.

Chapman, L., 2007. Transport and climate change: a review. *Journal of Transport Geography*, 15, 354-367.

Chataway, E.S., Kaplan, S., Nielsen, T.A.S., Prato, C.G., 2014. Safety perceptions and behavior related to cycling in mixed traffic: a comparison between Brisbane and Copenhagen. *Transportation Research Part F*, 23, 32-43.

Chong, S., Poulos, R., Olivier, J., Watson, W.L., Grzebieta, R., 2010. Relative injury severity among vulnerable non-motorised road users: comparative analysis of injury arising from bicycle-motor vehicle and bicycle-pedestrian collisions. *Accident Analysis and Prevention*, 42, 290-296.

City of Copenhagen, 2012. Good, better, best - The city of Copenhagen's bicycle strategy 2011-2025. http://www.kk.dk.

de Geus, B., Degraeuwe, B., Vandenbulcke, G., Int Panis, L., Thomas, I., Aertsens, J., De Weerdt, Y., Torfs, R., Meeusen, R., 2013. Utilitarian cycling in Belgium: A cross-sectional study in a sample of regular cyclists. *Journal of Physical Activity and Health*, Epub ahead of print.

de Hartog, J., Boogaard, H., Nijland, H., Hoek, G., 2010. Do the health benefits of cycling outweigh the risks? *Environmental Health Perspectives*, 118, 1109-1116.

de Nazelle, A., Nieuwenhuijsen, M.J., Antó, J.M., Brauer, M., Briggs, D., Braun-Fahrlander, C., Cavill, N., Cooper, A.R., Desqueyroux, H., Fruin, S., Hoek, G., Int Panis, L., Janssen, N., Jerrett, M., Joffe, M., Andersen, Z.J., van Kempen, E., Kingham, S., Kubesch, N., Leyden, K.M., Marshall, J.D., Matamala, J., Mellios, G., Mendez, M., Nassif, H., Ogilvie, D., Peiró, R., Pérez, K., Rabl, A., Ragettli, M., Rodríguez, D., Rojas, D., Ruiz, P., Sallis, J.F., Terwoert, J., Toussaint, J.F., Tuomisto, J., Zuurbier, M., Lebret, E., 2011. Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment. *Environment International*, 37, 766-777.

de Rome, L., Boufous, S., Georgeson, T., Senserrick, T., Richardson, D., Ivers, R., 2014. Bicycle crashes in different riding environments in the Australian Capital Territory. *Traffic Injury Prevention*, 15, 81-88.

Elvik, R., 2009. The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis and Prevention*, 41, 849-855.

Int Panis, L., de Geus, B., Vandenbulcke, G., Willems, H., Degraeuwe, B., Bleux, N., Mishra, V., Thomas, I., Meeusen, R., 2010. Exposure to particulate matter in traffic: a comparison of cyclists and car passengers. *Atmospheric Environment*, 44, 2263-2270.

Jacobsen, P., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prevention*, 9, 205-209.

Joshi, M.S., Senior, V., Smith, G.P., 2001. A diary study of the risk perceptions of road users. *Health Risk and Society*, 3, 261-279.

Juhra, C., Wieskötter, B., Chu, K., Trost, L., Weiss, U., Messerschmidt, M., Malczyk, A., Heckwolf, M., Raschke, M., 2012. Bicycle accidents – Do we only see the tip of the iceberg? A prospective multi-centre study in a large German city combining medical and police data. *Injury*, 43, 2026-2034.

Kaplan, S., Vavatsoulas, K., Prato, C.G., 2014. Aggravating and mitigating factors associated with cyclist injury severity in Denmark. *Journal of Safety Research*, 50, 75-82.

Kaplan, S., Prato, C.G., 2013. Cyclist-motorist crash patterns in Denmark: A latent class clustering approach. *Traffic Injury Prevention*, 14, 725-733.

Lawson, A.R., Pakrashi, V., Ghosh, B., Szeto, W.Y., 2013. Perception of safety of cyclists in Dublin City. *Accident Analysis and Prevention*, 50, 499-511.

Li, L., Zhu, L., Sui, D., 2007. A GIS-based Bayesian approach for analyzing spatial-temporal patterns of intra-city motor vehicle crashes. *Journal of Transport Geography*, 15, 274-285.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A*, 44, 291-305.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: Evolution, critique and future directions (with discussion). *Statistics in Medicine*, 28, 3049-3082.

Mannering, F., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research*, 1, 1-22.

Møller, M., Hels, T., 2008. Cyclists' perceptions of risk in roundabouts. *Accident Analysis and Prevention*, 40, 1055-1062.

Parkin, J., Wardman, M., Page, M., 2007. Models of perceived cycling risk and route acceptability. *Accident Analysis and Prevention*, 39, 364-371.

Pei, X., Wong, S.C., Sze, N.N., 2010. A joint-probability approach to crash prediction models. *Accident Analysis and Prevention*, 43, 1160-1166.

Pucher, J., Buehler, R., 2008. Making cycling irresistible: lessons from The Netherlands, Denmark and Germany. *Transport Reviews*, 28, 495-528.

Pucher, J., Buehler, R., Seinen, M., 2011. Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A*, 45, 451-475.

Pucher, J., Garrard, J., Greaves, S., 2011. Cycling down under: a comparative analysis of bicycling trends and policies in Sydney and Melbourne. *Journal of Transport Geography*, 19, 332-345.

Pucher, J., Komanoff, C., Schimek, P., 1999. Bycicling renaissance in North America? Recent trends and alternative policies to promote bicycling. *Transportation Research Part A*, 33, 625-654.

Quddus, M., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity. An analysis of London crash data. *Accident Analysis and Prevention*, 40, 1486-1497.

Rabl, A., de Nazelle, A., 2011. Assessment of health costs and benefits for active transport. *Transport Policy*, 19, 121-131.

Reynolds, C., Harris, M., Teschke, K., Cripton, P., Winters, M., 2009. The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. *Environmental Health*, 8, 47.

Rodgers, G., 1997. Factors associated with the crash risk of adult bicyclists. *Journal of Safety Research*, 28, 233-241.

Rojas-Rueda, D., de Nazelle, A., Tainio, M., Nieuwenhuijsen, M.J., 2011. The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. *British Medical Journal*, 343, d4521.

Spiegelhalter, D., Best, N.G., Carlin, B.P., Van der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583-639.

Vandenbulcke, G., Thomas, I., de Geus, B., Degraeuwe, B., Torfs, R., Meeusen, R., Int Panis, L., 2009. Mapping bicycle use and the risk of accidents for commuters who cycle to work in Belgium. *Transport Policy*, 16, 77-87.

Vandenbulcke, G., Thomas, I., Int Panis, L., 2014. Predicting cycling accident risk in Brussels: A spatial case-control approach. *Accident Analysis and Prevention*, 62, 341-357.

Veisten, K., Sælensminde, K., Alvær, K., Bjørnskau, T., Elvik, R., Schistad, T., Ytterstad, B., 2007. Total costs of bicycle injuries in Norway: Correcting injury figures and indicating data needs. *Accident Analysis and Prevention*, 39, 1162-1169.

Wang, C., Quddus, M., Ison, S., 2009. Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. *Accident Analysis and Prevention*, 41, 798-808.

Wang, Y., Nihan, N., 2004. Estimating the risk of collisions between bicycles and motor vehicles at signalized intersections. *Accident Analysis and Prevention*, 36, 313-321.

Winters, M., Davidson, G., Kao, D., Teschke, K., 2011. Motivators and deterrents of bicycling: comparing influences on decisions to ride. *Transportation*, 38, 153-168.

Wood, J.M., Lacherez, P.F., Marszalek, R.P., King, M.J., 2009. Drivers' and cyclists' experiences of sharing the road: Incidents, attitudes and perceptions of visibility. *Accident Analysis and Prevention*, 41, 772-776.

Yiannakoulias, N., Bennet, A.S., Scott, D.M., 2012. Mapping commuter cycling risk in urban areas. *Accident Analysis and Prevention*, 45, 164-172.