Cyclist crash rates and risk factors in a prospective cohort in seven European cities

Peer-reviewed author version

Branion-Calles, Michael; Gotschi, Thomas; Nelson, Trisalyn; Anaya-Boig, Esther; Avila-Palencia, Ione; Castro, Alberto; Cole-Hunter, Tom; de Nazelle, Audrey; DONS, Evi; Gaupp-Berghausen, Mailin; Gerike, Regine; Panis, Luc Int; Kahlmeier, Sonja; Nieuwenhuijsen, Mark; Rojas-Rueda, David & Winters, Meghan (2020) Cyclist crash rates and risk factors in a prospective cohort in seven European cities. In: ACCIDENT ANALYSIS AND PREVENTION, 141 (Art N° 105540).

DOI: 10.1016/j.aap.2020.105540
Handle: http://hdl.handle.net/1942/31886

Cyclist crash rates and risk factors in a prospective cohort in seven European cities

Michael Branion-Calles [a,b], Thomas Götschi [c], Trisalyn Nelson [d], Esther Anaya-Boig [e], Ione Avila-Palencia [f,g,h,i], Alberto Castro [j], Tom Cole-Hunter [k,l,m], Audrey de Nazelle [e], Evi Dons [n,o], Mailin Gaupp-Berghausen [p], Regine Gerike [q], Luc Int Panis [e,o,r], Sonja Kahlmeier [s], Mark Nieuwenhuijsen [f,g,h], David Rojas-Rueda [f,t], Meghan Winters [a,b]

[a] Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada
[b] Centre for Hip Health and Mobility, Vancouver, Canada
[c] School of Planning, Public Policy and Management, College of Design, University of Oregon, Eugene, USA
[d] School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, USA
[e] Centre for Environmental Policy, Imperial College London, London, United Kingdom
[f] ISGlobal, Barcelona, Spain
[g] Universitat Pompeu Fabra (UPF), Barcelona, Spain
[h] CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[i] Urban Health Collaborative, Dornsife School of Public Health, Drexel University, Philadelphia, USA
[j] Epidemiology, Biostatistics and Prevention Institute, University of Zürich, Zürich, Switzerland
[k] Centre for Air Pollution, Energy, and Health Research (CAR), University of New South Wales, Sydney, Australia
[l] International Laboratory for Air Quality and Health, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia
[m] Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia
[n] Flemish Institute for Technological Research (VITO), Mol, Belgium
[o] Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium
[p] Department of Spatial, Landscape, and Infrastructure Sciences, University of Natural Resources and Life Sciences, Vienna, Austria
[q] Institute of Transport Planning and Road Traffic, Dresden University of Technology, Dresden, Germany
[r] Transportation Research Institute (IMOB), Hasselt University, Diepenbeek, Belgium
[s] Department of Health, Swiss Distance University of Applied Science FFHS, Regensdorf/Zürich, Switzerland.
[t] Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, USA

Michael Branion-Calles (Corresponding Author)
Simon Fraser University
Blusson Hall, Room 11300
8888 University Drive
Burnaby, B.C.
V5A 1S6

michael_branion-calles@sfu.ca

Abstract

Increased cycling uptake can improve population health, but barriers include real and perceived risks. Crash risk factors are important to understand in order to improve safety and increase cycling uptake. Many studies of cycling crash risk are based on combining diverse sources of crash and exposure data, such as police databases (crashes) and travel surveys (exposure), based on shared geography and time. When conflating crash and exposure data from different sources, the risk factors that can be quantified are only those variables common to both datasets, which tend to be limited to geography (e.g. countries, provinces, municipalities) and a few general road user characteristics (e.g. gender and age strata). The Physical Activity through Sustainable Transport Approaches (PASTA) project was a prospective cohort study that collected both crash and exposure data from seven European cities (Antwerp, Barcelona, London, Örebro, Rome, Vienna and Zürich). The goal of this research was to use data from the PASTA project to quantify exposure-adjusted crash rates and model adjusted crash risk factors, including detailed sociodemographic characteristics, attitudes about transportation, neighbourhood built environment features and location by city. We used negative binomial regression to model the influence of risk factors independent of exposure. Of the 4,180 cyclists, 10.2% reported 535 crashes. We found that overall crash rates were 6.7 times higher in London, the city with the highest crash rate, relative to Örebro, the city with the lowest rate. Differences in overall crash rates between cities are driven largely by crashes that did not require medical treatment and that involved motor-vehicles. In a parsimonious crash risk model, we found higher crash risks for less frequent cyclists, men, those who perceive cycling to not be well regarded in their neighbourhood, and those who live in areas of very high building density. Longitudinal collection of crash and exposure data can provide important insights into individual differences in crash risk. Substantial differences in crash risks between cities, neighbourhoods and population groups suggest there is great potential for improvement in cycling safety.

Keywords: cycling safety, crash rates, risk factors, Europe, cohort

1.0 Introduction

Cycling for transport has many potential societal benefits. Increased cycling can improve population health outcomes through increased physical activity (de Hartog et al., 2010; Götschi et al., 2016; Mueller et al., 2015; Rojas-Rueda et al., 2016). Cycling also has potential harms, both real and perceived, that prevent concerned individuals from cycling. Negative safety perceptions are a main barrier to cycling (Heinen et al., 2010; Willis et al., 2015). Cyclists have higher risks of injury and/or fatality than other road users in highly motorized countries (Beck et al., 2007; Mindell et al., 2012; Reynolds et al., 2017; Scholes et al., 2018; Wegman et al., 2012).

It is critical to understand risk factors for cycling crashes to identify potential strategies for interventions. Studies of crash incidence require both crash and exposure data (e.g., cycling distance or duration) for a specified area and time (Götschi et al., 2016; Vanparijs et al., 2015). Exposure-based studies of cycling risk are typically conducted by compiling crash and exposure data from different sources, generally police databases (crashes) and travel surveys (exposure) (Castro et al., 2018; Hautzinger et al., 2007). Comparative studies of crash risk require attributes that are common to both the crash and exposure datasets. When combining crash and exposure data from different sources, common attributes tend to be limited to geography (e.g. countries, provinces, municipalities) and a few general road user characteristics (age and gender strata) (Beck et al., 2007; Blaizot et al., 2013; Mindell et al., 2012; Reynolds et al., 2017; Santamarina-Rubio et al., 2014; Scholes et al., 2018; Teschke et al., 2013). As a result, most exposure-based risk studies that combine disparate exposure and crash data are not able to provide detailed explorations of crash risk factors, such as individual user characteristics including cycling frequency, perception of their social environment, and neighbourhood features. Furthermore, different sources of data also make comparisons across different cities problematic.

Most exposure-based studies of cycling risk typically use crashes reported to police and/or hospital databases which under report less-serious injuries and crashes without injury (Amoros et al., 2006; de Geus et al., 2012; Elvik and Mysen, 1999; Juhra et al., 2012; Langley et al., 2003; Vanparijs et al., 2016; Veisten et al., 2007; Watson et al., 2015; Winters and Branion-Calles, 2017) and can make comparisons across different regions problematic due to potential differences in reporting practices (Yannis et al., 2014). Less severe crashes and crashes without injury are important to capture as they comprise the vast majority of crashes that occur and are a substantial economic cost to society (Aertsens et al., 2010; Veisten et al., 2007), considering treatment costs, productivity loss or leisure time loss (Aertsens et al., 2010). Furthermore, minor crashes and crashes without injury can negatively affect how individuals perceive cycling safety (Sanders, 2015), which may reduce cycling uptake and therefore minimize the net potential health and other benefits from cycling.

Prospective cohort studies offer an opportunity to address these limitations by collecting data on a range of crash types, including single bicycle crashes or crashes without injury (de Geus et al., 2012; Poulos et al., 2012). Furthermore, participant-specific travel behaviour can also be collected concurrently (Vanparijs et al., 2015), while also permitting the identification of individual sociodemographic, behavioural, social environment and built environment factors associated with crash risk. As a result, this design can allow for collection of less-severe crash types, more accurate calculation of crash rates and identification of individual level crash risk factors.

The Physical Activity through Sustainable Transport Approaches (PASTA) project was a prospective cohort study that used a longitudinal web survey of over 10,000 individuals residing in seven cities across Europe that collected crash and exposure data simultaneously (Gerike et al., 2016). The goal of this study was to use data from the PASTA project to quantify exposure-adjusted crash rates and model crash risk factors, including sociodemographic characteristics, social environment (including attitudes and social norms), and neighbourhood-built environment features.

2.0 Materials and Methods

2.1 Study Area

Our study area includes the cities in which participants were recruited for the PASTA project (Antwerp/Belgium, Barcelona/Spain, London/UK, Örebro/Sweden, Rome/Italy, Vienna/Austria and Zürich/Switzerland) (Gerike et al., 2016). These cities represent a range of environments in terms of size, population characteristics, mode shares, built environment, and culture (Table 1). Örebro and Antwerp have the highest levels of cycling, with 25% and 23% of trips being made by bicycle, respectively (Mueller et al., 2018). Örebro is also the least dense of the cities but is supported by a well maintained hierarchical network of cycling infrastructure, consisting of high speed regional cycling corridors that feed into local networks (PASTA Consortium, 2018a). Antwerp is a much more dense city than Örebro and is supported by a vast network of cycle paths and an extensive bike share program (PASTA Consortium, 2018b). Vienna has the next highest mode share at 6% (Mueller et al., 2018). This city is characterised by particularly high dynamics in cycling promotion. Having started with the active promotion of cycling not as long ago as some of the other cities in this study such as Antwerp or Örebro, it has today one of the largest cycling networks amongst the PASTA cities (PASTA Consortium, 2018c). Zürich has a modest 4% of trips made by bike (Mueller et al., 2018). Historically, other modes of transportation have been prioritized over cycling in Zürich resulting in an excellent public transportation system along with a high mode share of walking, but a fragmented cycling network (PASTA Consortium, 2018d). London has seen an increase in both investment in the cycling network and growth in cycling trips (Aldred and Dales, 2017) but still has a cycling mode share of only 3% (Mueller et al., 2018). Similar to London, Barcelona has expanded their cycling network significantly in recent years and is considered to be an emerging city for cycling (PASTA Consortium, 2018e) but currently only has a mode share of 2% (Mueller et al., 2018). Finally, Rome has the lowest cycling mode share at 1% (Mueller et al., 2018), a very limited cycling network and is considered to be a challenging place to get around by bicycle (PASTA Consortium, 2018f).

Table 1: Characteristics of PASTA Cities

City	**Antwerp**	**Barcelona**	**London**	**Örebro**	**Rome**	**Vienna**	**Zürich**
Country	**Belgium**	**Spain**	**United Kingdom**	**Sweden**	**Italy**	**Austria**	**Switzerland**
Population[a]	502,604	1,620,943	8,538,689	138,952	2,683,842	1,741,246	398,575
Area (Km^2)[a]	204	102	1,572	1,373	1,285	415	92
Population Density (pop/km^2)[a]	2,464	15,892	5,432	101	2,089	4,196	4,332
Cycling Mode Share (%)[b]	23	2	3	25	1	6	4
Cyclists/day[c]	113,509	26,532	235,288	26,538	18,846	75,685	16,416
Mean distance (km)[c]	3.84	3.5	3	3.3	7.7	3.3	2.77
Mean time (mins)[c]	14.4	16.2	22.8	16.2	24	18.6	14.4
Cycling network km (OSM)[d]	469.17	159.54	969.17	361.35	120.64	715.63	118.36
Street network km (OSM)[d]	1,651.74	1,554.56	16,439.74	3,045.27	8,281.36	3,946.11	1,193.59
Cycling network / street network[d]	0.28	0.10	0.06	0.12	0.01	0.18	0.10
Fatalities/year[e]	4	3	13	1	4	3	1
Cycling km/ year[e]	313,625,445	89,663,002	463,174,636	59,361,390	98,362,110	219,430,669	45,048,048
Fatalities/ billion km[e]	13	33	28	17	41	14	22

[a] Data compiled in Gerike et al. (2016) and refer to the year 2012.

[b] Data compiled in Mueller et al (2018). Antwerp, Barcelona, London, Örebro, Rome, Vienna and Zürich based on data from 2011, 2012, 2012, 2011, 2014, 2012, 2010, respectively.

[c] Data compiled in Mueller et al (2018). Antwerp, Barcelona, London, Örebro, Rome, Vienna and Zürich based on data from 2013, 2006/2015, 2013, 2011, 2014, 2013, 2010, respectively.

[d] Data compiled in Mueller et al (2018) from OpenStreetMap as of October 2017.

[e] Data compiled in Mueller et al (2018). Antwerp, Barcelona, London, Örebro, Rome, Vienna and Zürich based on traffic fatality data from 2011-2014, 2011-2015, 2014, 2012, 2015, 2010-2015, 2006-2010, respectively.

2.2 Study Design

Cycling crash and exposure data were collected as a part of the larger PASTA project (Gerike et al., 2016). The project used a longitudinal web-based survey (Dons et al., 2015). Data were collected between November 2014 (April 2015 in Örebro) and December 2016, primarily through an opportunistic sampling approach, though some participants in Örebro were recruited through random sampling. Participants were recruited with the same methods across all cities, which included press releases/editorials, consistent design of promotional materials, translation of promotional materials to local languages, close collaboration with local stakeholders networks to distribute information, promotion of the study through social media and participation incentivization through a prize lottery (except for Örebro where lotteries were not permitted) (Gaupp-Berghausen et al., 2019). A participant could enter (and leave) the study at any point within the data collection period. Participants were required to be at least 18 years of age, except for Zürich, where the minimum age was 16 years. The survey oversampled cyclists to ensure sufficient statistical power for analysis in cities with a low cycling mode share (Raser et al., 2018).

The PASTA project consisted of a comprehensive baseline questionnaire followed by follow-up surveys (Figure 1). The baseline questionnaire collected data on sociodemographic characteristics, travel behaviour, physical activity, information regarding locations of their home, work and school, as well as data on attitudes toward transportation. Follow-up survey invitations were sent every 13 days after completion of a questionnaire to collect prospective repeated measurements of travel, physical activity behavior, and safety incidents. Each follow-up survey included a modified version of the Global Physical Activity Questionnaire (GPAQ) aimed at estimating the duration and frequency of cycling in the previous week (World Health Organization, 2019). Every third follow-up included a 1-day travel diary. A custom designed web-survey platform automatically sent reminder emails for participants to complete questionnaires.

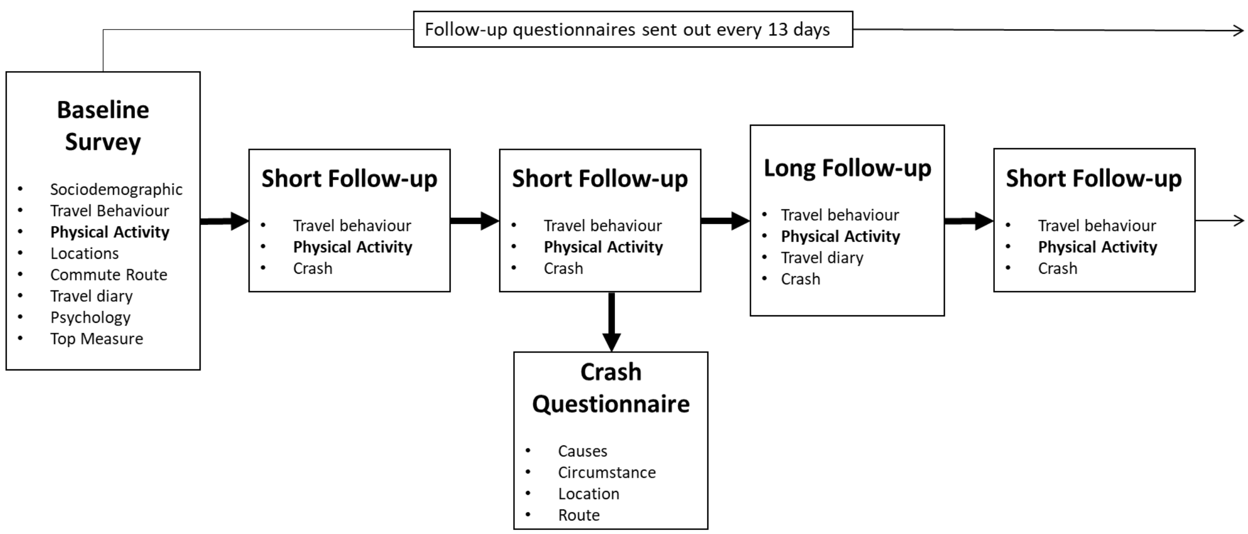

Figure 1: Longitudinal Survey Design for PASTA participants

2.3 Cycling Exposure and Crash Data

We estimated cycling duration from the modified version of the GPAQ administered in every follow-up survey. The questionnaire consisted of the following two questions: 1) “In the previous 7-days on how many days do you cycle for at least 10 minutes continuously to get to and from places?” and 2) “Typically, how much time do you spend cycling on such a day?”.

To obtain an estimate of weekly duration of cycling at each follow-up, we multiplied the number of days cycled by the typical time spent cycling. To estimate the total cycling exposure for the study, we multiplied the average of a participants weekly cycling over all follow-ups by the number of weeks between the date they entered the study and the date of the last follow-up survey they completed.

For capturing safety incidents, each follow-up survey asked, "Since the last time you filled out a questionnaire…, have you experienced any safety relevant incidents (i.e. a collision, fall or near miss as a pedestrian, cyclists, in public transport or driving)?" If participants had experienced a collision or fall, they were asked to complete a crash questionnaire for each case. This crash questionnaire collected details of circumstances including the crash type (fall, crash with motor vehicle, crash with cyclist or crash with pedestrian), injury (injury or non-injury) and medical treatment (none, treated without doctor, treated by doctor, brought to hospital, hospitalized). We only include collisions and falls while cycling in our analysis (e.g., we removed falls or crashes while using other modes of transport) referred to collectively as crashes in this paper.

2.4 Covariates

The PASTA study followed a comprehensive framework to understanding active travel behaviour, and aimed to not only measure sociodemographic characteristics, but also the characteristics of their social and built environments (Götschi et al., 2017). For this analysis, we selected the sociodemographic, social and built environment characteristics from the baseline survey that were either previously identified as risk factors in other cyclist cohorts (Degraeuwe et al., 2015; Poulos et al., 2015; Tin Tin et al., 2013; Vanparijs et al., 2015), or had a plausible association with crash risk, such as perceptions of traffic safety. Sociodemographic characteristics included age, gender, education, body mass index (BMI), and whether the cyclist had a driver's license. Perceptions of built and social environments were included as well, where participants rated their level of agreement with whether cycling for travel was comfortable, whether cycling for travel was safe with regards to traffic, whether cycling was well regarded in their neighborhood, and whether cycling was common in their neighbourhood. For each participant we also generated objective measures of the built environment around a participant's home (300m) including cycling infrastructure density, building density and a measure of "greenness" the Normalized Difference Vegetation Index (NDVI). These were derived by mapping the participants' residential locations to geospatial data from local partners, and/or open data infrastructure including from the European Environment Agency and OpenStreetMap.

2.5 Data cleaning and dealing with missing values

From the over 10,000 participants who completed the baseline questionnaire, we included those who completed at least one follow-up survey in which they were asked about cycling crashes (n=6,817). We removed participants who reported zero minutes of cycling (n=2,448), those who reported over 8 hours of daily cycling in 1 or more follow-ups (n=190), those that were in the study for less than 13 days (n=12), and those who provided incomplete data on crash type and injury (n=62). There were 4,180 participants who fulfilled the inclusion criteria.

Across the relevant baseline sociodemographic, social and built environment variables, the percentage of missing data amongst eligible participants ranged from 0 to 16.9%. The specific variables with missing data included age (n=1, <0.1%), BMI (n = 19, 0.5%), education level (n = 8, 0.2%), having young children (n=153, 3.7%), building density within 300 m of residence (n = 65, 1.6%), bike lane density within 300 m of residence (n = 707, 16.9%), street density within 300m of residence (n = 60, 1.4%), and NDVI within 300 m of residence (n = 65, 1.6%). In total 886 out of 4,180 eligible participants (21.2%) had incomplete sociodemographic, social or built environment data.

To address the missing values in sociodemographic, social and built environment variables we took a multiple imputation approach. Specifically, we used the multivariate imputation by chained equations (MICE) technique using fully conditional specification and the default settings of the mice 3.6 package in R (van Buuren and Groothuis-Oudshoorn, 2011). Multiple imputation creates multiple plausible versions of a complete dataset by filling in the missing values with reasonable estimates (Azur et al., 2011). We used the MICE algorithm to create 20 imputed datasets based on the rule of thumb that the number of imputations should approximate the proportion of incomplete cases (van Buuren, 2018).We converted building density, bike lane density, and NDVI to a categorical variable based on quintiles for each imputed dataset after imputation. We then calculated crash rates using the non-imputed data and statistically modelled crash risks using the imputed datasets.

2.6 Statistical Analysis

2.6.1 Crash rates

Using the non-imputed data, we calculated overall crash rate (number of crashes per 100,000 hours of cycling) by combining recorded crashes with exposure data. We also calculated crash rates by city and a range of sociodemographic, attitudinal, and built environment characteristics without data imputation. We used bootstrapping with 5,000 replications to generate 95% confidence intervals around crash rates.

To further understand differences in crash rates by city, we also examined crash rates for specific types of crashes based on which road users were involved and the injury severity. Crashes were defined as either involving a motor-vehicle, another cyclist, a pedestrian or a fall. There were 9 crashes that involved multiple other road users. These crashes were assigned to a category based on the most dangerous road user involved, where we ranked road users from most to least dangerous as follows: motor-vehicles, another cyclist, pedestrian, and finally no other road user (i.e. a fall). We also used medical treatment as a proxy for injury severity and assigned a crash as requiring medical treatment if the participant sought any kind of medical treatment, or else not requiring medical treatment.

2.6.2 Crash risk factors

To explore crash risk factors, we analysed the relationship between crash risks, exposure and other individual level factors. We applied this in each of the multiply imputed datasets and combined the results into a pooled model as per Rubin's rules (Azur et al., 2011). We used Generalized Linear Models with negative binomial error structures, and logarithmic links to quantify the relationship between the number of crashes a participant reported as a function of exposure, individual level factors, and city (Hilbe, 2014). We defined a base crash risk model as the following (Elvik, 2009):

$$\hat{E}(Y) = e^{\propto_0} \times EXP^{\propto_1} \times T^{\propto_2} \times e^{(b_1 city_1 + \ldots + b_6 city_6)} \quad (1)$$

where $\hat{E}(Y)$ is the predicted crashes for a participant, EXP is the average cycling exposure per month, T is the total months in the study and $city$ is an indicator variable for the city a participant resides in. We used city as an indicator variable to adjust for between city differences in individual crash risk (Cerin, 2011). Since participants spent differing amounts of time in the study there is potential for attrition bias. Here, attrition bias refers to the notion that there may be differences in crash risk between participants who participate for different lengths of time (Nunan et al., 2018). Therefore, we separated total exposure into two sub-components: average monthly exposure (EXP) and total number of observed months (T). The coefficients $\propto_0$, $\propto_1$, $\propto_2$ and b_i are estimated using maximum likelihood methods. If $\propto_1$ or $\propto_2$ are < 1 it means that the number of expected crashes increases less than proportionally to increases in average exposure or time in the study, respectively. We would expect $\propto_2$ to be ~1 if attrition was non-differential with regards to crash risk.

Each of the specified sociodemographic, social, and built environment characteristics was then initially examined separately by adding each to the base model:

$$\hat{E}(Y) = e^{\propto_0} \times EXP^{\propto_1} \times T^{\propto_2} \times e^{(b_1 city_1 + \ldots + b_6 city_6)} \times e^{(b_7 x_1 + \ldots + b_{(7+k)} x_k)} \quad (2)$$

Where x represented the one additional indicator variable of interest with k levels. We then estimated the incident rate ratio (IRR) for each level of x by exponentiating its coefficient, b. The IRR here represents the change in crash risk from the reference category in a specified sociodemographic, social, or built environment characteristic holding exposure and city-level differences constant. We will refer to these IRR's as "crude".

Finally, we developed a parsimonious crash risk model in a forward stepwise procedure. We added additional variables to the base model one at a time, based on the multivariate Wald statistic, from highest to lowest (van Buuren, 2018). A variable was kept in the model if the Wald statistic had a p-value under 0.2. The final parsimonious model is given by:

$$\hat{E}(Y) = e^{\propto_0} \times EXP^{\propto_1} \times T^{\propto_2} \times e^{(b_1 city_1 + \dots + b_6 city_6)} \times e^{(b_7 x_1 + \dots + b_n x_n)} \qquad (3)$$

Where there are n number of sociodemographic, social or built environment variables that have a p-value under 0.2. We use a high p-value to avoid excluding potentially important variables. We will refer to the IRRs based on this parsimonious model as "adjusted".

3.0 Results

Out of the 10,691 participants in the PASTA study, 4,180 participants provided cycling exposure data in at least one follow-up and did not provide outlier values or unreliable crash data (Table 2). We will refer to these participants as cyclists. The cyclists completed a median of seven follow-up surveys over a median of 7.3 months. At baseline, most reported being daily or almost daily cyclists (60.3%) and reported cycling for a median daily average of 16.3 minutes over follow-ups. Relative to other cities, London had the fewest cyclists (n=355), while Antwerp had the most (n=891). Cyclists were nearly evenly split between men and women and tended to be young and highly educated. Most cyclists agreed that cycling for transport was comfortable (72.9%) but only a minority agreed that it was safe from traffic (28.0%). Most participants agreed that cycling in their neighbourhood was well regarded (49.5%) and common (41.9%).

About one in ten cyclists experienced one or more crashes (10.2%) during their time in the study.

Table 2: Baseline characteristics of the cyclists the PASTA study

Variable	
No. Participants	4180
Months Observed (median [IQR])	7.3 [2.2, 16.6]
Follow-ups Completed (median [IQR])	7.0 [3.0, 17.0]
Total Exposure in Hours (median [IQR])	36.0 [10.5, 114.7]
Average Exposure in Minutes Per Day (median [IQR])	16.3 [6.4, 31.1]
Crashes per person (%)	
0	3752 (89.8)
>=1	428 (10.2)
City (%)	
Antwerp	891 (21.3)
Barcelona	523 (12.5)
London	355 (8.5)
Örebro	590 (14.1)
Roma	594 (14.2)
Vienna	637 (15.2)
Zürich	590 (14.1)
Cycling Frequency at baseline (%)	
Never	138 (3.3)
Less than once per month	247 (5.9)
on 1-3 days per month	370 (8.9)
on 1-3 days per week	893 (21.4)
Daily or almost daily	2522 (60.3)
Missing	10 (0.2)
Age (%)	
16-25 years	483 (11.6)
26-35 years	1313 (31.4)
36-45 years	1049 (25.1)
46-55 years	840 (20.1)
56-65 years	401 (9.6)
65+ years	93 (2.2)
Missing	1 (<0.1)
Gender (%)	
Women	2066 (49.4)
Men	2114 (50.6)

BMI (%)	
<25	2994 (71.6)
25-30	951 (22.8)
30+	216 (5.2)
Missing	19 (0.5)
Education (%)	
No degree/primary education	49 (1.2)
Secondary/further education	930 (22.2)
Higher/university education	3193 (76.4)
Missing	8 (0.2)
Income (%)	
< € 10,000	314 (7.5)
€ 10,000 - € 24,999	628 (15.0)
€ 25,000 - € 49,999	1232 (29.5)
€ 50,000 - € 74,999	799 (19.1)
€ 75,000 - € 99,999	309 (7.4)
€ 100,000 - € 150,000	171 (4.1)
€ >150,000	64 (1.5)
Missing	663 (15.9)
Drivers License (%)	
Yes	3812 (91.2)
No	368 (8.8)
Have Children (%)	
Yes	1460 (34.9)
No	2567 (61.4)
Missing	153 (3.7)
Cycling for transport is comfortable* (%)	
Agree	3047 (72.9)
Neutral	729 (17.4)
Disagree	404 (9.7)
Cycling for transport is safe from traffic* (%)	
Agree	1172 (28.0)
Neutral	1100 (26.3)
Disagree	1908 (45.6)
In my neighbourhood cycling is well regarded* (%)	
Agree	2070 (49.5)
Neutral	1306 (31.2)
Disagree	804 (19.2)
In my neighbourhood cycling is common* (%)	
Agree	1750 (41.9)
Neutral	1194 (28.6)
Disagree	1236 (29.6)

Building density of residence (m^2/km^2), 300 m buffer (%)	
Quintile 1: [0 – 111,080]	823 (19.7)
Quintile 2: (111,080 – 195,283]	823 (19.7)
Quintile 3: (195,283 – 285,409]	823 (19.7)
Quintile 4: (285,409 – 418,375]	823 (19.7)
Quintile 5: (418,375 – 659,249]	823 (19.7)
Missing	65 (1.6)
Bike lane density of residence (m/km^2), 300 m buffer (%)	
Quintile 1: [0]	988 (23.6)
Quintile 2: [0.031 – 1,530]	622 (14.9)
Quintile 3: (1,530 – 3,240]	621 (14.9)
Quintile 4: (3,240 – 5,700]	621 (14.9)
Quintile 5: (5,700– 20,400]	621 (14.9)
Missing	707 (16.9)
Street density of residence (m/km^2), 300 m buffer (%)	
Quintile 1: [570 – 11,600]	824 (19.7)
Quintile 2: (11,600 – 15,700]	824 (19.7)
Quintile 3: (15,700 – 19,300]	824 (19.7)
Quintile 4: (19,300 – 23,300]	824 (19.7)
Quintile 5: (23,300 – 49,000]	824 (19.7)
Missing	60 (1.4)
NDVI of residence, 300 m buffer (%)	
Quintile 1: [0.122 – 0.271]	828 (19.8)
Quintile 2: (0.271 – 0.360]	826 (19.8)
Quintile 3: (0.360 – 0.474]	815 (19.5)
Quintile 4: (0.474 – 0.595]	828 (19.8)
Quintile 5: (0.595 – 0.874]	818 (19.6)
Missing	65 (1.6)

*Collapsed from 5 category Likert scale, IQR: Interquartile range

3.1 Crash characteristics

Of the 4,180 cyclists in our study, 428 reported a total of 535 crashes (Table 3). Of these, two in five (40.4%) were falls (single bicycle crashes). The remaining crashes involved another road user, either a motor vehicle (35.3%), another cyclist (17.4%) or a pedestrian (6.9%). Just over half of all crashes resulted in an injury (considered to be a bruise or cramp at minimum) (55.3%). Just over a quarter of crashes required any medical treatment (26.5%), and there were just four hospitalizations (0.7%) . Most crashes went unreported in official sources: only 3.9% were reported as recorded by police and 9.3% were reported to an insurance company.

Table 3: Crash characteristics including involvement, injury, and medical treatment

	N Crashes (%)
Total s	535 (100)
Cycling Crash Types	
Fall	216 (40.4)
Crash with motor vehicle	189 (35.3)
Crash with other cyclist	93 (17.4)
Crash with pedestrian	37 (6.9)
Injury[a]	
Yes	296 (55.3)
No	239 (44.7)
Medical Treatment	
No	393 (73.5)
Yes, I treated it myself or by another person (no doctor).	80 (15.0)
Yes, I went to a doctor or hospital myself.	47 (8.8)
Yes, from an ambulance at the location of the crash.	0 (0.0)
Yes, I was brought to the hospital for medical treatment but could go home the same day.	11 (2.1)
Yes, I was hospitalized ≥ 1 night.	4 (0.7)
Official police report	
Yes, the police showed up and they officially reported the crash	16 (3.0)
Yes, I reported the crash later to the police (in the station, by phone or online).	5 (0.9)
No, the police showed up but they didn't officially report the crash.	7 (1.3)
No, the police didn't show up and the crash was not officially reported	493 (92.1)
Don't know	14 (2.6)
Reported to insurance company	
Yes	50 (9.3)
No	466 (87.1)
Don't know	19 (3.8)

[a]Defined as physical injury resulting from the crash including bruises or cramps

3.2 Crash Rates

Across the seven cities the crash rate was 137.9 crashes per 100,000 hours of cycling (95% CI, 125.2 - 152.1) or 1 crash every 725 hours. London had the highest crash rate with 220.8 crashes per 100,000 hours, while Örebro had the lowest crash rate of 32.8 crashes per 100,000 hours (Table 4, Figure 2). Zürich had the second highest crash rate of 188.6 per 100,000 hours followed by Vienna, Rome, Antwerp and Barcelona with rates of 154.3,144.9, 136.1 and 134.1, respectively (Table 4, Figure 2). The high crash rate in London was largely driven by its greater number of crashes which involved a motor-vehicle relative to the other PASTA cities, while falls appeared to be a greater issue in Rome compared to other crash types (Figure 3B). When stratifying by whether medical treatment was required or not, there was relatively little difference in the crash rates between cities with the exception of Örebro which had a substantially lower rate requiring treatment (Figure 3C). The overall differences in crash rates between cities appear to be largely driven by crashes that did not require any medical treatment (Figure 3C).

We also examined total crash rates by sociodemographic, social and built environment characteristics of cyclists. Crash rates decreased with increasing age category, increased with higher BMI category and were higher for men compared to women (Table 4). Crash rates tended to be highest for participants who disagreed that cycling for travel was comfortable, well regarded in their neighbourhood or common (Table 4). Participants who lived in neighbourhoods with a higher density of bike lanes, higher NDVI or lower building density tended to have lower crash rates (Table 4).

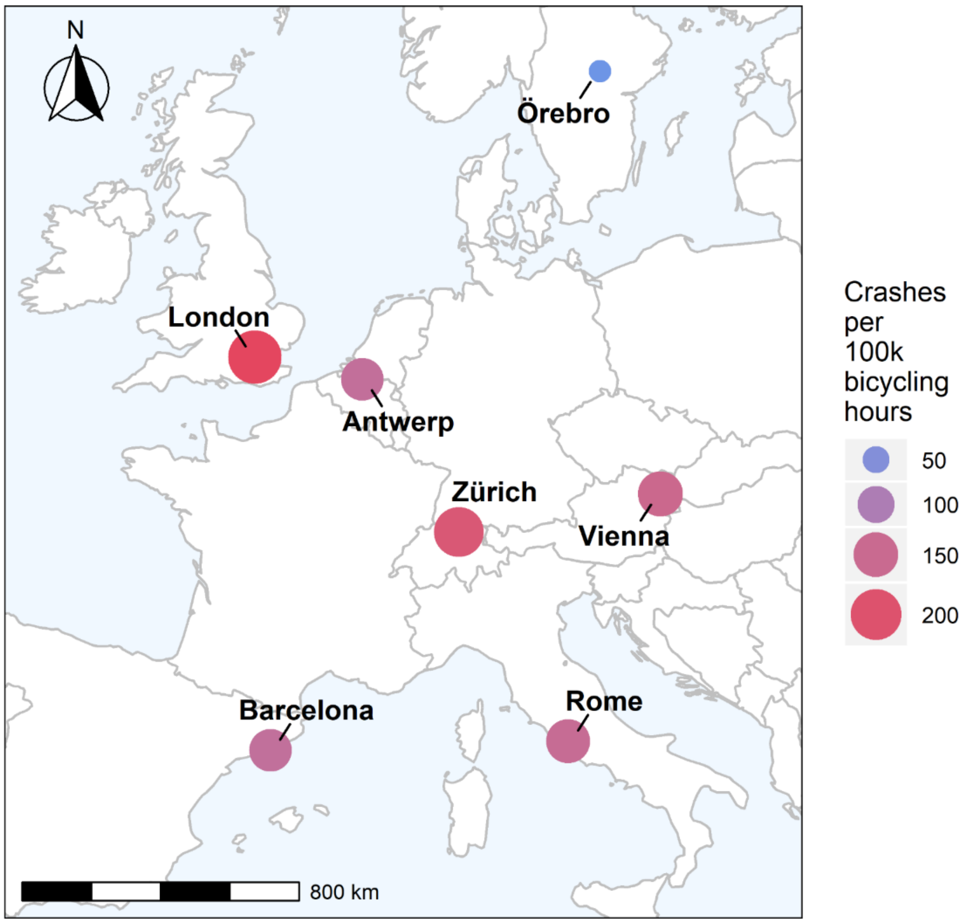

Figure 2: Map of crash rates by city

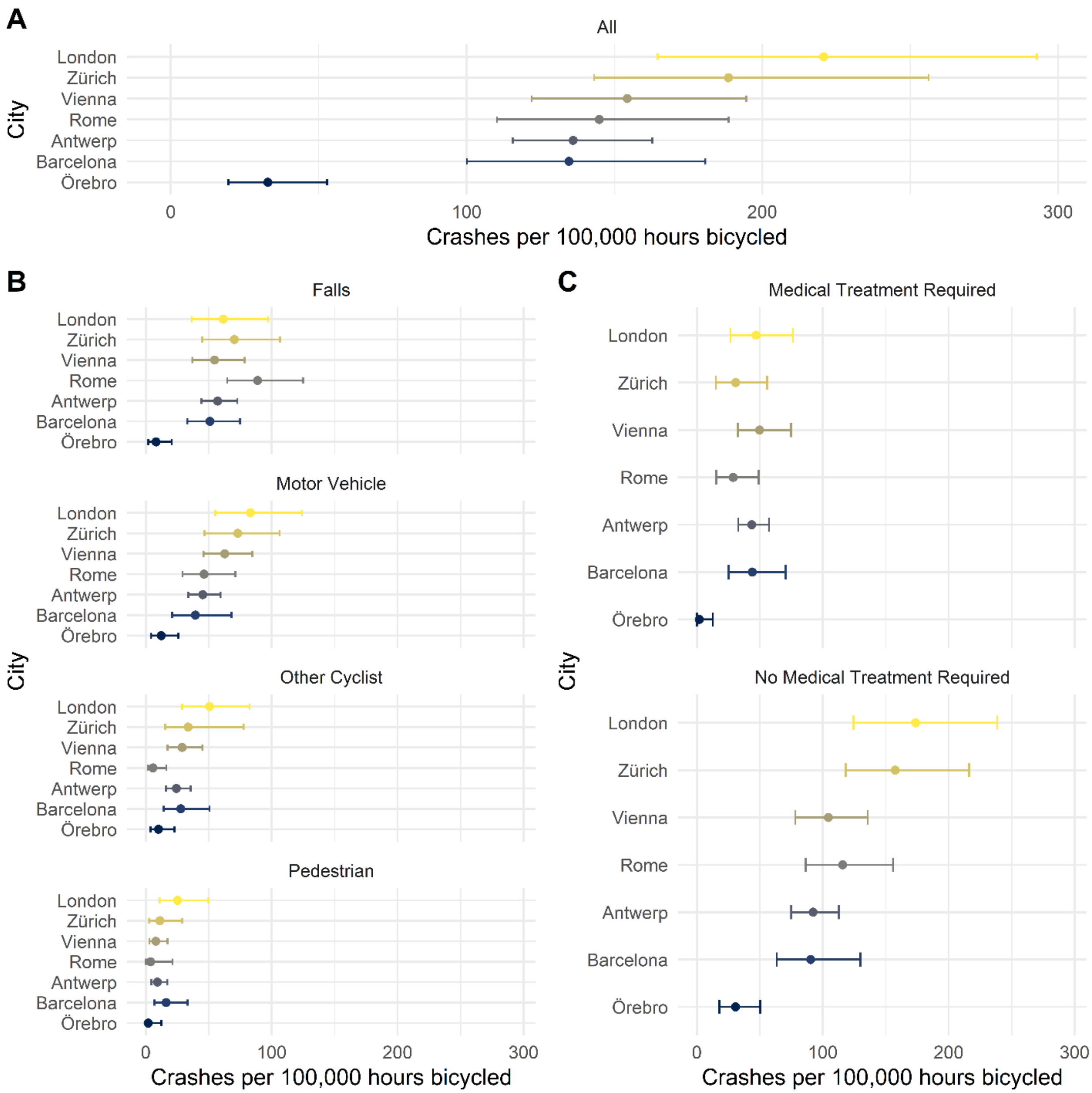

Figure 3: Crash rates by city for A) all crashes B) stratified by other road user involved in the crash and C) by injury severity

Table 4: Crash risk factors by city, sociodemographic, attitudinal, and neighbourhood characteristics.

Variable	*Level*	*% of total n (4,180)*	*Total Exposure Hours*	*Total number of crashes*	*Crash Rate per 100,000 hours (95% CI)[a]*	*Crude Incident Rate Ratio (95% CI)[b]*

Total		100.0	387,968	535	137.9 (125.2, 152.1)	
City	Antwerp	21.3	119,041	162	136.1 (115.6, 162.9)	Reference[c]
	Barcelona	12.5	43,084	58	134.6 (100.2, 180.7)	0.87 (0.63, 1.22)
	London	8.5	27,630	61	220.8 (164.6, 292.8)	**1.54 (1.09, 2.17)**
	Örebro	14.1	48,721	16	32.8 (19.5, 52.8)	**0.21 (0.13, 0.36)**
	Roma	14.2	51,767	75	144.9 (110.3, 188.6)	1.06 (0.78, 1.45)
	Vienna	15.2	62,198	96	154.3 (122.0, 194.6)	1.03 (0.77, 1.37)
	Zürich	14.1	35,529	67	188.6 (143.1, 256.2)	1.11 (0.80, 1.53)
Age (years)	16-25	11.6	32,140	56	174.2 (124.9, 238.0)	Reference
	26-35	31.4	115,659	167	144.4 (121.2, 172.3)	0.83 (0.59, 1.15)
	36-45	25.1	101,433	146	143.9 (118.4, 175.1)	0.87 (0.62, 1.23)
	46-55	20.1	89,817	117	130.3 (105.3, 157.2)	0.81 (0.57, 1.16)
	56-65	9.6	44,317	45	101.5 (73.8, 138.2)	0.73 (0.47, 1.13)
	65+	2.2	4,497	4	89.0 (20.8, 249.6)	0.53 (0.18, 1.56)
	Missing	0	105	0		
BMI	<25	71.6	277,398	369	133.0 (118.4, 149.9)	Reference
	25-30	22.8	90,476	130	143.7 (114.9, 177.1)	1.19 (0.95, 1.49)
	30+	5.2	18,466	34	184.1 (126.3, 266.9)	1.39 (0.93, 2.09)
	Missing	0.5	1,628	2	122.8 (0.0, 267.6)	
Gender	Women	49.4	166,862	187	112.1 (95.0, 132.5)	Reference
	Men	50.6	221,107	348	157.4 (140.1, 178.5)	**1.42 (1.16, 1.73)**
Education	No degree/Primary	1.2	5,547	6	108.2 (35.7, 216.5)	Reference
	Secondary/further	22.2	79,233	116	146.4 (118.0, 184.5)	1.06 (0.41, 2.76)
	Higher/university	76.4	302,491	410	135.5 (121.7, 151.2)	0.96 (0.38, 2.43)
	Missing	0.2	697	3	430.2 (0.0, 2110.6)	
Drivers License	Yes	91.2	356,032	483	135.7 (122.5, 150.3)	Reference
	No	8.8	31,936	52	162.8 (117.0, 222.7)	1.16 (0.84, 1.60)
Have Children under 18	Yes	34.9	143,338	176	122.8 (104.3, 145.3)	Reference
	No	61.4	67,299	335	143.2 (126.0, 162.1)	1.09 (0.89, 1.34)
	Missing	3.7	10,664	24	225 (133.8, 360.2)	
Cycling 'for travel' is comfortable	Agree	72.9	316,507	420	132.7 (119.0, 148.5)	Reference
	Neutral	17.4	53,633	79	147.3 (114.4, 191.9)	0.97 (0.74, 1.27)
	Disagree	9.7	17,828	36	201.9 (136.1, 293.6)	1.17 (0.80, 1.72)
Cycling 'for travel' is safe (with regards to traffic)	Agree	28	130,021	144	110.8 (91.7, 133.0)	Reference
	Neutral	26.3	109,699	160	145.9 (120.8, 175.5)	1.16 (0.90, 1.49)
	Disagree	45.6	148,248	231	155.8 (134.3, 178.6)	1.10 (0.87, 1.39)
In my neighbourhood is cycling is well regarded	Agree	49.5	202,584	248	122.4 (106.4, 141.0)	Reference
	Neutral	31.2	108,480	149	137.4 (113.8, 165.8)	1.16 (0.92, 1.46)
	Disagree	19.2	76,904	138	179.4 (146.9, 219.0)	**1.33 (1.03, 1.73)**
In my neighbourhood	Agree	41.9	159,337	201	126.1 (107.6, 145.9)	Reference
	Neutral	28.6	108,046	137	126.8 (105.7, 151.3)	1.03 (0.81, 1.32)

is cycling is common	Disagree	29.6	120,585	197	163.4 (135.9, 194.2)	1.26 (0.99, 1.62)
Building density (m²/km²) within 300 m of residential location	[0 – 111,080]	19.7	82,398	86	104.4 (83.6, 128.0)	Reference
	(111,080 – 195,283]	19.7	83,056	103	124.0 (99.6, 154.2)	0.88 (0.63, 1.22)
	(195,283 – 285,409]	19.7	70,103	92	131.2 (101.3, 169.9)	0.76 (0.54, 1.07)
	(285,409 – 418,375]	19.7	71,501	122	170.6 (138.8, 207.8)	1.09 (0.79, 1.52)
	(418,375 – 659,249]	19.7	75,708	124	163.8 (132.6, 201.3)	1.20 (0.84, 1.73)
	Missing	1.6	5,203	8	153.7 (60.9, 345.0)	
Bike lane density (m/km²) within 300 m of residential location	[0]	23.6	91,284	141	154.5 (127.2, 186.3)	Reference
	[0.031 – 1,530]	14.9	59,303	96	161.9 (124.1, 210.5)	1.03 (0.75, 1.42)
	(1,530 – 3,240]	14.9	55,515	80	144.1 (113.0, 180.3)	0.95 (0.68, 1.32)
	(3,240 – 5,700]	14.9	52,456	67	127.7 (96.6, 171.0)	0.97 (0.67, 1.41)
	(5,700– 20,400]	14.9	58,913	57	96.8 (70.8, 129.1)	0.84 (0.58, 1.21)
	Missing	16.9	70,495	94	133.3 (107.5, 163.4)	
Street density (m/km²) within 300 m of residential location	[570 – 11,600]	19.7	92273	98	106.2 (84.2, 132.4)	Reference
	(11,600 – 15,700]	19.7	79,924	99	123.9 (99.0, 153.0)	0.98 (0.71, 1.34)
	(15,700 – 19,300]	19.7	71,207	97	136.2 (108.5, 168.2)	0.98 (0.72, 1.37
	(19,300 – 23,300]	19.7	79,924	129	178.3 (144.5, 219.8	1.26 (0.92, 1.72)
	(23,300 – 49,000]	19.7	71,207	105	155.8 (124.9, 196.9)	1.19 (0.84, 1.70)
	Missing	1.4	72,348	7	145.4 (55.4, 349.7)	
NDVI within 300 m of home location	[0.122 – 0.271]	19.8	72,482	115	158.7 (126.1, 197.4)	Reference
	(0.271 – 0.360]	19.8	72,674	122	167.9 (137.0, 201.5)	0.90 (0.66, 1.22)
	(0.360 – 0.474]	19.5	70,742	108	152.7 (121.6, 194.1)	0.82 (0.59, 1.15)
	(0.474 – 0.595]	19.8	77,777	83	106.7 (83.8, 134.1)	**0.65 (0.45, 0.94)**
	(0.595 – 0.874]	19.6	89,194	100	112.1 (89.8, 136.5)	0.75 (0.51, 1.09)
	Missing	1.6	5,099	7	137.3 (50.0, 323.8)	

[a] Confidence intervals calculated using a bias corrected and accelerated bootstrap method (BCa) with 5,000 replications

[b] Adjusted for average cycling exposure per month, number of months participated in the study and city

[c] Adjusted for average cycling exposure per month, number of months participated in the study

Bold indicates significance at 95% confidence,

NDVI = Normalized Difference Vegetation Index.

3.3 Crash Risk Factors

In our parsimonious model, we identified average exposure per month, months of participation, city, gender, perceiving that cycling is well regarded in their neighbourhood, and building density as important factors affecting crash risk. The final pooled parsimonious model suggests a non-linear relationship between individual cycling exposure and number of crashes:

$$\hat{E}(Y) = 0.0005 \times EXP_{MonthlyAvg}^{0.58} \times T_{Months}^{0.80} \quad (4)$$

The exponents for *EXP* and *T* are less than 1, thus while the number of expected crashes for a participant increase with both increased cycling per month and number of months participated, the risk of a crash (expected crashes per unit of exposure) decreases. In other words crash risk is lower for participants who cycle more frequently as well as those who spend more time participating in the study. The effects are of differing strength, with attrition bias being weaker than the effect of exposure per month.

In the parsimonious model, additional risk factors for a crash included: being a man, living in a neighbourhood of very high building density, and perceiving that cycling was not well regarded in one's neighbourhood (Figure 4). London and Örebro stood out as the most and least risky cities, respectively. Relative to Antwerp, and holding exposure and other individual factors constant, the crash risk for a participant in London was 1.58 times higher, while in Örebro it was less than a quarter as risky (Figure 4). When we ran a sensitivity analysis with only observations with complete data for all variables, we found very similar results.

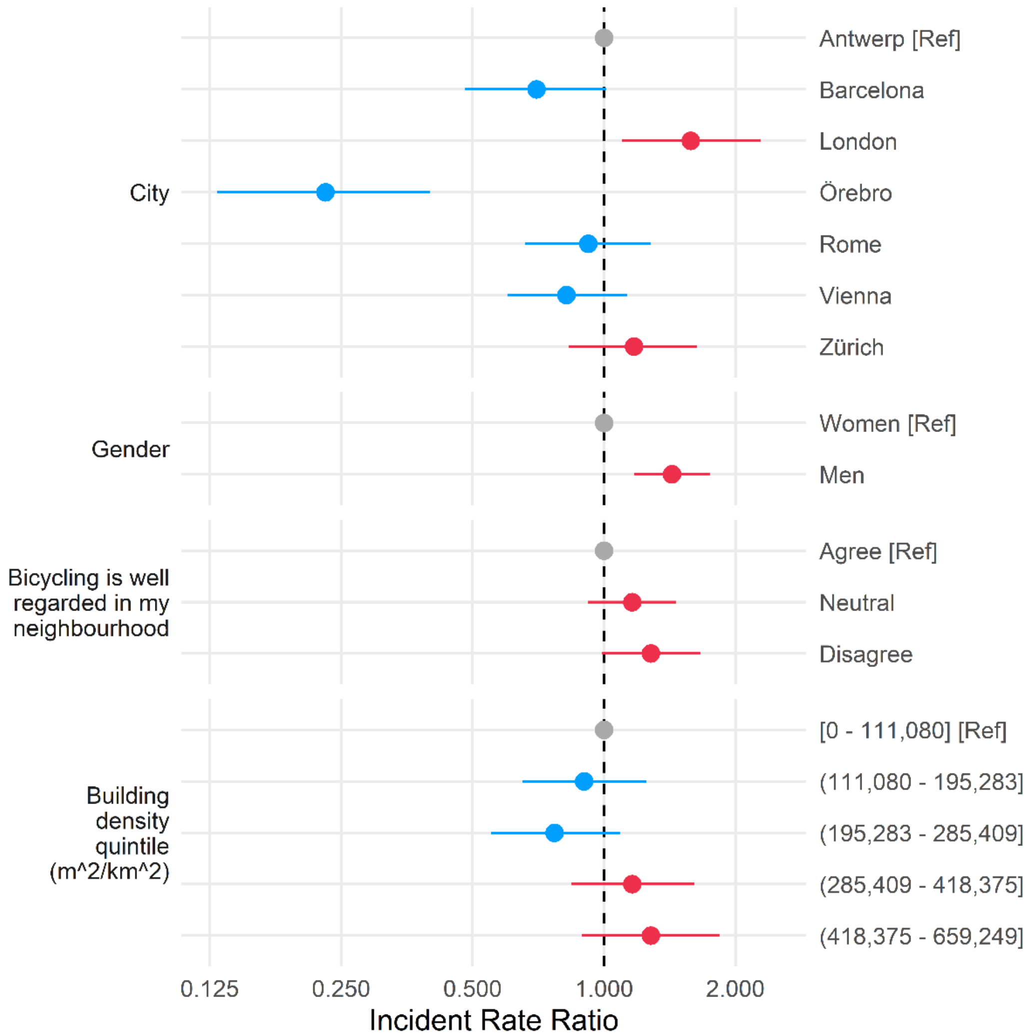

Figure 4: Adjusted Incident Rate Ratio (IRR) for the variables included in the stepwise model building. The IRR's represent the ratio of expected crashes for a given participant relative to the reference category, adjusting for average exposure per month, number of months participated in addition to the selected variables in the plot (City, Gender, Cycling well regarded in neighbourhood, and Building density quintile). Holding average exposure and time constant, participants with the highest number of expected crashes are those that live in London, are men, perceive that cycling in their neighbourhood is not well regarded, and live in a neighbourhood within the highest quintile of building density.

4.0 Discussion

This study analysed prospectively collected crash data in a cohort of cyclists across seven geographically diverse European cities, one of the largest studies of its kind. Of the seven PASTA cities, we found considerable variation in crash risk. Within cities, risk of a crash was highest for less frequent cyclists, men, those who perceive cycling to not be well regarded in their neighbourhood, and those who live in areas of very high building density. We show that crash risks differ by city, neighbourhood and individual level factors.

Our findings, like those in the literature, indicate that crash rates vary substantially across cities. Overall, the average crash rate was 137.9 crashes per 100,000 hours, from a low of 32.8 crashes per 100,000 hours in Örebro to nearly 7 times higher in London. There are only two other studies which collected crash and duration-based exposure data simultaneously; one in New South Wales, Australia (Poulos et al., 2015) and one in Belgium (de Geus et al., 2012). The Australian cohort had an incidence rate of 606.0 crashes per 100,000 hours (Poulos et al., 2015), while the Belgian cohort had an rate of 89.6 per 100,000 hours (de Geus et al., 2012). Difference in rates between these studies may be due to differences in the samples of cyclists, inclusion criteria for inclusion of crashes, methods in calculating and/or collecting exposure, as well as actual differences in the objective risk of a cycling crash between these areas. To illustrate, the Australian study included non-injury crashes and a substantial proportion of the cohort (40.1%) were “mainly recreational” cyclists (Poulos et al., 2015). The Belgian cohort did not include any recreational cycling and excluded non-injury crashes (de Geus et al., 2012). The discrepancy between the crash rate we estimated in Antwerp (136.1 per 100,000 hours) and the crash rate of 75.2 per 100,000 hours for Flanders (a larger region of Belgium in which Antwerp is located) may be partly explained by the crash inclusion criteria. In our study the crash rate for Antwerp drops to 74.8 crashes per 100,000 hours for crashes that resulted in injury and to 43.7 crashes per 100,000 for crashes that required any medical treatment.

When interpreting the differences we found in crash rates and/or adjusted crash risk between PASTA cities, the role of self selection should be considered (Castro et al., 2018). Here, self-selection refers to the idea that due to unsafe cycling conditions, many people may choose not to bicycle at all. Thus, the participants who choose to bicycle in these unsafe conditions may be overly brave and/or exceptionally skillful, the latter possibly having a moderating effect on crash risk (Castro et al., 2018). As such, cycling mode share provides important contextual information when interpreting differences between large geographic units. PASTA cities were selected in part to introduce variability in the samples in terms of culture, density, built environments, policies, and climates, and thereby cover a wide range of conditions related to cyclist safety (Dons et al., 2015). Our parsimonious model (which adjusted for other factors including exposure, time spent in the study, gender, social environment and neighbourhood building density) indicated that Örebro was the safest city for cycling, London the riskiest, and the remaining cities similar in terms of safety. Notably, Örebro was the least risky city and has the highest cycling mode share of our seven cities (25%). Antwerp was riskier than Barcelona, and was similar in risk to Rome, Vienna, and Zürich. Antwerp's population bicycles between 3.8 to 23 times more than Barcelona, Rome, Vienna or Zürich, which suggests that self-selection may play a role in risk differences. For example, Rome is known for highly challenging traffic conditions, reflected in the least extensive cycling network amongst the seven cities, and the lowest bicycle mode share (1%) (Mueller et al., 2018). Despite this, our model suggests a similar level of overall crash risk between participants in Antwerp and Rome. These results should be interpreted with some caution, as they reflected a definition of crashes that include those that resulted in no medical treatment. When we excluded crashes that did not result in an injury that required medical treatment, the differences between cities were much smaller, with the exception that Orebro was still far safer.

Individual level cyclist crash risk is a complex phenomenon comprised of interactions between individuals and their road environment in both space and time (Schepers et al., 2014). At the individual level, we found that crash risks varied based on frequency of cycling. Those that reported higher cycling were at a lower risk, compared to those that reported lower rates of cycling. The non-linear relationship between average cycling per month and number of expected crashes suggest a “safety in exposure” effect, such as safety from being a more experienced cyclist. This is consistent with the “safety in numbers” effect observed at aggregated spatial units (Elvik and Bjørnskau, 2015; Jacobsen, 2003), suggesting an individual level component to this phenomenon. The “safety in numbers” effect has been attributed to behavioural aspects, such as drivers being more used to cyclists in high cycling environments, as well as structural aspects, such as safer cycling infrastructure attracting higher numbers of cyclists (Götschi et al., 2016). Our findings suggest that one contributing factor to “safety in numbers” might be a lower number of inexperienced or infrequent cyclists and/or the improvement of safety-relevant cycling skills with increasing experience/frequency (Elvik and Bjørnskau, 2015; Fyhri et al., 2017).

Previous research has also found differences in crash risks between different sociodemographic groups, such as between men and women or older and younger adults (Vanparijs et al., 2015). In prospective studies, the relationship between gender and crash risk has been mixed, with women having higher crash rates in Belgian (Degraeuwe et al., 2015) and Australian cohorts (Poulos et al., 2015), but lower crash rates in a New Zealand cohort (although not statistically significant) (Tin Tin et al., 2013). Women have also been found to be at higher risk for serious or fatal injuries while cycling using the bike share scheme of London (Woodcock et al., 2014), but have similar risks across the UK (Aldred and Dales, 2017). In our study, our parsimonious model suggests that men had a crash risk 1.43 times higher than women. These same cohorts also had mixed results concerning the relationship between age and crash risk, with one study finding that the risk of a minor crash decreases with age (Degraeuwe et al., 2015), another that it is lower for the youngest and oldest age groups (Poulos et al., 2015), and another that the directionality depends on whether the collision occurred on-street (risk increases with age) or with a motor vehicle (risk decreases with age) (Tin Tin et al., 2013). In this study we observed that crash risk was lower amongst older cyclists, but the trend was not statistically significant. The overall sample of PASTA participants (including non-cyclists) were broadly representative of gender distribution, but tended to be relatively younger compared to city census data (Gaupp-Berghausen et al., 2019).

We can only speculate on the reasons behind why certain sociodemographic groups are lower risk than others, but we suggest lower risk is an association between belonging to a given sociodemographic group and a tendency to cycle at lower speeds, and/or engage in fewer risky behaviours (e.g. cycle in safer areas and/or cycle more cautiously). For example, our finding that women and older adults were at lower risk for a crash relative to men and younger adults, may reflect the fact women have been found to cycle at lower speeds than men (Aldred and

Crosweller, 2015) and have a stronger preference towards using safer infrastructure than men (Aldred et al., 2016). Similarly older adults may be more cautious when bicycling compared to younger adults (Bernhoft and Carstensen, 2008).

A new contribution of this research is an inquiry on the association between differences in social environment and crash risk. We found that individual perceptions of social norms around cycling were associated with crash risk, where those who agreed that cycling was a well-regarded mode of transport in their neighborhood were at lower risk for a crash than those that were neutral (1.16 times higher) or disagreed (1.28 times higher risk). The perception question was asked at baseline, so preceded any reported crashes. We suggest this variable may be in part capturing different built environment conditions, where those participants who think cycling is well regarded may live and travel in safer areas for cyclists, within their respective cities. A part of this may also be the safety in numbers effect, where a higher level agreement corresponds to an area with more cyclists due to the presumably more supportive social environment for cycling.

Prospective studies such as this one indicate that cyclist crashes (including non-injury crashes) are more common than would be suggested by more conventional analyses of police or insurance data and travel surveys (Amoros et al., 2006; de Geus et al., 2012; Elvik and Mysen, 1999; Juhra et al., 2012; Langley et al., 2003; Veisten et al., 2007; Watson et al., 2015; Winters and Branion-Calles, 2017). In this study, we found that on average across the cities, one crash occurs for every 725 hours bicycled. In contrast, a study in France that combined police recorded crashes and travel survey data found 1 crash per 93,023 hours of cycling (Blaizot et al., 2013). Only 3.9% of crashes reported to PASTA were recorded by police, at a rate of 1 per 18,475 hours of cycling. Of course, police, insurance and hospital data capture more severe (less frequent) events, resulting in lower rates of crashes (Amoros et al., 2006; Blaizot et al., 2013; Elvik and Mysen, 1999; Juhra et al., 2012; Veisten et al., 2007; Winters and Branion-Calles, 2017). While less severe crashes are under-reported in police records, it is likely more severe events are under-reported in the PASTA dataset as it is not possible to self-report a fatality and we cannot ascertain if a participant has dropped out due to severe injury. However, non-injury events may not result in direct healthcare costs, but have important implications for cycling in terms of perceived safety, and potentially future uptake (Aldred et al., 2016; Aldred and Crosweller, 2015; Sanders, 2015), which consequently may have costs from non-materialized health benefits from prevented cycling.

Our study has several strengths and limitations. The prospective design enabled the collection of detailed exposure data, as well as data on a range of different crash types including falls and non-injury events, for a large number of individuals. The data collection was consistent across cities, enabling more valid comparisons. Furthermore, the study design allowed for multivariable analysis to assess impacts of individual factors on crash risks, a refinement over what can be done with aggregate data. The relationships we found between the selected explanatory variables and crash risk should not be interpreted as causal. The extent to which our sample of cyclists are representative of the broader cycling population is not known, and results should be interpreted with some caution. PASTA participants are more educated and younger than the general population (Gaupp-Berghausen et al., 2019), although recruitment specifically oversampled cyclists, so there may be better representation of the cycling populations. The fact that our results show a lower risk of a crash with increasing time in the study may indicate some bias from loss to follow-up. There may be some reporting bias in crashes, although the repeated surveys (as little as 2 weeks apart) was designed to limit recall issues. This study primarily collected non-injury crashes and did not observe many serious injuries and is not able to record fatalities. We did not have traffic condition data to further explain neighbourhood level risks, and limitations to statistical power did not warrant further investigations of crash location attributes. The objective GIS measures of built environment (bike lane density, street density, building density, NDVI) only represent conditions within 300m of a participant's residence and may not reflect the route conditions in which participants typically ride, especially for longer trips. Spatially resolved exposure data would allow for further important analyses, such as risks associated with specific route characteristics, but at the beginning of PASTA large scale collection of spatially resolved route data from participants was not feasible due to limitations of available tracking apps at the time. Passive detection of cycling routes through mobile tracking apps (Geurs et al., 2015) could enable the widespread collection of spatially resolved exposure data and more detailed investigation of policy relevant risk factors in future studies.

5 Conclusions

The PASTA design provides comparable crash risks for cyclists, adjusted for differences in age and gender and other variables, across the diverse set of seven European cities. The large variations in crash risks indicate that cyclists' safety can still be improved considerably.

Longitudinal study designs can provide important insights into crash risk factors within cities, neighbourhoods, and population groups, in particular for minor crashes. Future research should focus on representative datasets that can integrate the most policy relevant crash risk factors such as route infrastructure and exposure to motorized modes, with individual characteristics and perceptions, benefitting from rapid progress in the collection of spatially resolved exposure data.

Declaration of Competing Interest

None.

Funding sources

This work was supported by the European PASTA project. PASTA is a 4-year project funded by the European Union's Seventh Framework Program under EC-GA No. 602624-2 (FP7-HEALTH-2013-INNOVATION-1). The funders had no role in study design, analysis, or writing of this manuscript. MBC is supported by a SSHRC Doctoral Fellowship. MW holds a Scholar Award from the Michael Smith Foundation for Health 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., Panis, L.I., 2010. Commuting by bike in Belgium, the costs of minor accidents. Accid. Anal. Prev. 42 6 , 2149–2157. doi:10.1016/j.aap.2010.07.008

Aldred, R., Crosweller, S., 2015. Investigating the rates and impacts of near misses and related incidents among UK cyclists. J. Transp. Heal. 2 3 , 379–393. doi:10.1016/j.jth.2015.05.006

Aldred, R., Dales, J., 2017. Diversifying and normalising cycling in London, UK: An exploratory study on the influence of infrastructure. J. Transp. Heal. 4, 348–362. doi:10.1016/j.jth.2016.11.002

Aldred, R., Elliott, B., Woodcock, J., Goodman, A., 2016. Cycling provision separated from motor traffic: a systematic review exploring whether stated preferences vary by gender and age. Transp. Rev. 1647 July , 1–27. doi:10.1080/01441647.2016.1200156

Amoros, E., Martin, J.-L., Laumon, B., 2006. Under-reporting of road crash casualties in France. Accid. Anal. Prev. 38 4 , 627–635. doi:10.1016/j.aap.2005.11.006

Azur, M.J., Stuart, E.A., Frangakis, C., Leaf, P.J., 2011. Multiple imputation by chained equations: what is it and how does it work? Int. J. Methods Psychiatr. Res. 20 1 , 40–49. doi:10.1002/mpr.329

Beck, L.F., Dellinger, A.M., O'Neil, M.E., 2007. Motor vehicle crash injury rates by mode of travel, United States: Using exposure-based methods to quantify differences. Am. J. Epidemiol. 166 2 , 212–218. doi:10.1093/aje/kwm064

Bernhoft, I.M., Carstensen, G., 2008. Preferences and behaviour of pedestrians and cyclists by age and gender. Transp. Res. Part F Traffic Psychol. Behav. 11 2 , 83–95. doi:10.1016/j.trf.2007.08.004

Blaizot, S., Papon, F., Haddak, M.M., Amoros, E., 2013. Injury incidence rates of cyclists compared to pedestrians, car occupants and powered two-wheeler riders, using a medical registry and mobility data, Rhône County, France. Accid. Anal. Prev. 58, 35–45. doi:10.1016/j.aap.2013.04.018

Castro, A., Kahlmeier, S., Gotschi, T., 2018. Exposure-Adjusted Road Fatality Rates for Cycling and Walking Fatality Rates for Cycling and Walking in European Countries Cycling and Walking in European Countries, in: International Transport Forum. Paris.

Cerin, E., 2011. Statistical Approaches to Testing the Relationships of the Built Environment with Resident-Level Physical Activity Behavior and Health Outcomes in Cross-Sectional Studies with Cluster Sampling. J. Plan. Lit. 26 2 , 151–167. doi:10.1177/0885412210386229

de Geus, B., Vandenbulcke, G., Int Panis, L., Thomas, I., Degraeuwe, B., Cumps, E., Aertsens, J., Torfs, R., Meeusen, R., 2012. A prospective cohort study on minor accidents involving commuter cyclists in Belgium. Accid. Anal. Prev. 45, 683–693. doi:10.1016/j.aap.2011.09.045

de Hartog, J.J., Boogaard, H., Nijland, H., Hoek, G., 2010. Do the health benefits of cycling outweigh the risks? Environ. Health Perspect. 118 8 , 1109–1116. doi:10.1289/ehp.0901747

Degraeuwe, B., de Geus, B., Thomas, I., Vandenbulcke, G., Meeusen, R., Int Panis, L., 2015. Cycling Behaviour and Accident Risk of Utilitarian Cyclists in Belgium. Cycl. Futur. From Res. into Pract.

Dons, E., Götschi, T., Nieuwenhuijsen, M., de Nazelle, A., Anaya, E., Avila-Palencia, I., Brand, C., Cole-Hunter, T., Gaupp-Berghausen, M., Kahlmeier, S., Laeremans, M., Mueller, N., Orjuela, J.P., Raser, E., Rojas-Rueda, D., Standaert, A., Stigell, E., Uhlmann, T., Gerike, R., Int Panis, L., 2015. Physical Activity through Sustainable Transport Approaches (PASTA): protocol for a multi-centre, longitudinal study. BMC Public Health 15 1 , 1126. doi:10.1186/s12889-015-2453-3

Elvik, R., 2009. The non-linearity of risk and the promotion of environmentally sustainable transport. Accid. Anal. Prev. 41 4 , 849–55. doi:10.1016/j.aap.2009.04.009

Elvik, R., Bjørnskau, T., 2015. Safety-in-numbers: A systematic review and meta-analysis of evidence. Saf. Sci. 0349 . doi:10.1016/j.ssci.2015.07.017

Elvik, R., Mysen, A.B., 1999. Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries. Transp. Res. Rec. J. Transp. Res. Board 1665, 133–140.

Fyhri, A., Sundfør, H.B., Bjørnskau, T., Laureshyn, A., 2017. Safety in numbers for cyclists—conclusions from a multidisciplinary study of seasonal change in interplay and conflicts. Accid. Anal. Prev. 105, 124–133. doi:10.1016/j.aap.2016.04.039

Gaupp-Berghausen, M., Raser, E., Anaya-Boig, E., Avila-Palencia, I., de Nazelle, A., Dons, E., Franzen, H., Gerike, R., Gotschi, T., Iacorossi, F., Hossinger, R., Nieuwenhuijsen, M., Rojas-Rueda, D., Sanchez, J., Smeds, E., Deforth, M., Standaert, A., Stigell, E., Cole-Hunter, T., Int Panis, L., 2019. Evaluation of Different Recruitment Methods: Longitudinal, Web-Based, Pan-European Physical Activity Through Sustainable Transport Approaches (PASTA) Project. J. Med. Internet Res. 21 5 , e11492. doi:10.2196/11492

Gerike, R., de Nazelle, A., Nieuwenhuijsen, M., Panis, L.I., Anaya, E., Avila-Palencia, I., Boschetti, F., Brand, C., Cole-Hunter, T., Dons, E., Eriksson, U., Gaupp-Berghausen, M., Kahlmeier, S., Laeremans, M., Mueller, N., Orjuela, J.P., Racioppi, F., Raser, E., Rojas-Rueda, D., Schweizer, C., Standaert, A., Uhlmann, T., Wegener, S., Götschi, T., 2016. Physical Activity through Sustainable Transport Approaches (PASTA): a study protocol for a multicentre project. BMJ Open 6, e009924. doi:10.1186/s12889-015-2453-3

Geurs, K.T., Thomas, T., Bijlsma, M., Douhou, S., 2015. Automatic trip and mode detection with move smarter: First results from the Dutch Mobile Mobility Panel. Transp. Res. Procedia 11, 247–262. doi:10.1016/j.trpro.2015.12.022

Götschi, T., de Nazelle, A., Brand, C., Gerike, R., 2017. Towards a Comprehensive Conceptual Framework of Active Travel Behavior: a Review and Synthesis of Published Frameworks. Curr. Environ. Heal. reports 4 3 , 286–295. doi:10.1007/s40572-017-0149-9

Götschi, T., Garrard, J., Giles-Corti, B., 2016. Cycling as a Part of Daily Life: A Review of Health Perspectives. Transp. Rev. 36 1 , 45–71. doi:10.1080/01441647.2015.1057877

Hautzinger, H., Pastor, C., Pfeiffer, M., Schmidt, J., 2007. Analysis Methods for Accident and Injury Risk Studies.

Heinen, E., van Wee, B., Maat, K., 2010. Commuting by Bicycle: An Overview of the Literature. Transp. Rev. 30 1 , 59–96. doi:10.1080/01441640903187001

Hilbe, J.M., 2014. Modeling Count Data. Cambridge University Press, New York.

Jacobsen, P.L., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj. Prev. 9 3 , 205–209. doi:10.1136/ip.9.3.205

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? Injury 43 12 , 2026–2034. doi:10.1016/j.injury.2011.10.016

Langley, J., Dow, N., Stephenson, S., Kypri, K., 2003. Missing Cyclists. Inj. Prev. 9 4 , 376–379. doi:10.1136/ip.9.4.376

Mindell, J.S., Leslie, D., Wardlaw, M., 2012. Exposure-Based, “Like-for-Like” Assessment of Road Safety by Travel Mode Using Routine Health Data. PLoS One 7 12 , 1–10. doi:10.1371/journal.pone.0050606

Mueller, N., Rojas-Rueda, D., Cole-Hunter, T., de Nazelle, A., Dons, E., Gerike, R., Götschi, T., Int Panis, L., Kahlmeier, S., Nieuwenhuijsen, M., 2015. Health impact assessment of active transportation: A systematic review. Prev. Med. (Baltim). 76, 103–114. doi:10.1016/j.ypmed.2015.04.010

Mueller, N., Rojas-Rueda, D., Salmon, M., Martinez, D., Ambros, A., Brand, C., de Nazelle, A., Dons, E., Gaupp-Berghausen, M., Gerike, R., Götschi, T., Iacorossi, F., Int Panis, L., Kahlmeier, S., Raser, E., Nieuwenhuijsen, M., 2018. Health impact assessment of cycling network expansions in European cities. Prev. Med. (Baltim). 109 January , 62–70. doi:10.1016/j.ypmed.2017.12.011

Nunan, D., Aronson, J., Bankhead, C., 2018. Catalogue of bias: attrition bias. BMJ Evidence-Based Med. 23 1 , 21–22. doi:10.1136/ebmed-2017-110883

PASTA Consortium, 2018a. Facts on Active Mobility: Örebro/Belgium [WWW Document]. URL http://www.pastaproject.eu/fileadmin/editor-upload/sitecontent/Publications/documents/AM_Factsheet_Oebrebro_WP2.pdf (accessed 6.1.19).

PASTA Consortium, 2018b. Facts on Active Mobility: Antwerp/Belgium [WWW Document]. URL http://www.pastaproject.eu/fileadmin/editor-upload/sitecontent/Publications/documents/AM_Factsheet_Antwerp_WP2.pdf (accessed 6.1.19).

PASTA Consortium, 2018c. Facts on Active Mobility: Vienna/Austria [WWW Document]. URL http://www.pastaproject.eu/fileadmin/editor-upload/sitecontent/Publications/documents/AM_Factsheet_Vienna_WP2.pdf (accessed 6.1.19).

PASTA Consortium, 2018d. Facts on Active Mobility: Zürich/Switzerland [WWW Document]. URL http://www.pastaproject.eu/fileadmin/editor-upload/sitecontent/Publications/documents/AM_Factsheet_Zurich_WP2.pdf (accessed 6.1.19).

PASTA Consortium, 2018e. Facts on Active Mobility: Barcelona / Spain [WWW Document]. URL http://www.pastaproject.eu/fileadmin/editor-upload/sitecontent/Publications/documents/AM_Factsheet_Barcelona_WP2.pdf (accessed 6.1.19).

PASTA Consortium, 2018f. Facts on Active Mobility: Rome / Italy [WWW Document]. URL http://www.pastaproject.eu/fileadmin/editor-upload/sitecontent/Publications/documents/AM_Factsheet_Rome_WP2.pdf (accessed 6.1.19).

Poulos, R.G., Hatfield, J., Rissel, C., Flack, L.K., Murphy, S., Grzebieta, R., McIntosh, A.S., 2015. An exposure based study of crash and injury rates in a cohort of transport and recreational cyclists in New South Wales, Australia. Accid. Anal. Prev. 78, 29–38. doi:10.1016/j.aap.2015.02.009

Poulos, R.G., Hatfield, J., Rissel, C., Grzebieta, R., McIntosh, a. S., 2012. Exposure-based cycling crash, near miss and injury rates: The Safer Cycling Prospective Cohort Study protocol. Inj. Prev. 18 1 , e1–e1. doi:10.1136/injuryprev-2011-040160

Raser, E., Gaupp-Berghausen, M., Dons, E., Anaya-Boig, E., Avila-Palencia, I., Brand, C., Castro, A., Clark, A., Eriksson, U., Götschi, T., Int Panis, L., Kahlmeier, S., Laeremans, M., Mueller, N., Nieuwenhuijsen, M., Orjuela, J.P., Rojas-Rueda, D., Standaert, A., Stigell, E., Gerike, R., 2018. European cyclists' travel behavior: Differences and similarities between seven European (PASTA) cities. J. Transp. Heal. January , 0–1. doi:10.1016/j.jth.2018.02.006

Reynolds, S., Tranter, M., Baden, P., Mais, D., Dhani, A., Wolch, E., Bhagat, A., 2017. Reported Road Casualties Great Britain 2016. London.

Rojas-Rueda, D., De Nazelle, A., Andersen, Z.J., Braun-Fahrländer, C., Bruha, J., Bruhova-Foltynova, H., Desqueyroux, H., Praznoczy, C., Ragettli, M.S., Tainio, M., Nieuwenhuijsen, M.J., 2016. Health impacts of active transportation in Europe. PLoS One 11 3 , 1–14. doi:10.1371/journal.pone.0149990

Sanders, R.L., 2015. Perceived traffic risk for cyclists: The impact of near miss and collision experiences. Accid. Anal. Prev. 75, 26–34. doi:10.1016/j.aap.2014.11.004

Santamarina-Rubio, E., Perez, K., Olabarria, M., Novoa, A.M., 2014. Gender differences in road traffic injury rate using time travelled as a measure of exposure. Accid. Anal. Prev. 65, 1–7. doi:10.1016/j.aap.2013.11.015

Schepers, P., Hagenzieker, M., Methorst, R., Van Wee, B., Wegman, F., 2014. A conceptual framework for road safety and mobility applied to cycling safety. Accid. Anal. Prev. 62, 331–340. doi:10.1016/j.aap.2013.03.032

Scholes, S., Wardlaw, M., Anciaes, P., Heydecker, B., Mindell, J.S., 2018. Fatality rates associated with driving and cycling for all road users in Great Britain 2005–2013. J. Transp. Heal. 8 September 2017 , 321–333. doi:10.1016/j.jth.2017.11.143

Teschke, K., Anne Harris, M., Reynolds, C.C.O., Shen, H., Cripton, P.A., Winters, M., 2013. Exposure-based traffic crash injury rates by mode of travel in British Columbia. Can. J. Public Heal. 104 1 , e75–e79.

Tin Tin, S., Woodward, A., Ameratunga, S., 2013. Incidence, risk, and protective factors of bicycle crashes: Findings from a prospective cohort study in New Zealand. Prev. Med. (Baltim). 57 3 , 152–161. doi:10.1016/j.ypmed.2013.05.001

van Buuren, S., 2018. Flexible imputation of missing data, Second edi. ed, Interdisciplinary statistics. Boca Raton, FL : CRC Press, Taylor & Francis Group.

van Buuren, S., Groothuis-Oudshoorn, K., 2011. mice : Multivariate Imputation by Chained Equations in R. J. Stat. Softw. 45 3 , 1–67. doi:10.18637/jss.v045.i03

Vanparijs, J., Int Panis, L., Meeusen, R., de Geus, B., 2016. Characteristics of bicycle crashes in an adolescent population in Flanders (Belgium). Accid. Anal. Prev. 97, 103–110. doi:10.1016/j.aap.2016.08.018

Vanparijs, J., Int Panis, L., Meeusen, R., de Geus, B., 2015. Exposure measurement in bicycle safety analysis: A review of the literature. Accid. Anal. Prev. 84, 9–19. doi:10.1016/j.aap.2015.08.007

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. Accid. Anal. Prev. 39 6 , 1162–1169. doi:10.1016/j.aap.2007.03.002

Watson, A., Watson, B., Vallmuur, K., 2015. Estimating under-reporting of road crash injuries to police using multiple linked data collections. Accid. Anal. Prev. 83, 18–25. doi:10.1016/j.aap.2015.06.011

Wegman, F., Zhang, F., Dijkstra, A., 2012. How to make more cycling good for road safety? Accid. Anal. Prev. 44 1 , 19–29. doi:10.1016/j.aap.2010.11.010

Willis, D.P., Manaugh, K., El-Geneidy, A., 2015. Cycling Under Influence: Summarizing the Influence of Perceptions, Attitudes, Habits, and Social Environments on Cycling for Transportation. Int. J. Sustain. Transp. 9 8 , 565–579. doi:10.1080/15568318.2013.827285

Winters, M., Branion-Calles, M., 2017. Cycling safety: Quantifying the under reporting of cycling incidents in Vancouver, British Columbia. J. Transp. Heal. 7, 48–53. doi:10.1016/j.jth.2017.02.010

Woodcock, J., Tainio, M., Cheshire, J., O’Brien, O., Goodman, A., 2014. Health effects of the London bicycle sharing system: Health impact modelling study. BMJ 348 February , 1–14. doi:10.1136/bmj.g425

World Health Organization, 2019. Global Physical Activity Surveillance [WWW Document]. URL http://www.who.int/ncds/surveillance/steps/GPAQ/en/ (accessed 6.15.19).

Yannis, G., Papadimitriou, E., Chaziris, A., Broughton, J., 2014. Modeling road accident injury under-reporting in Europe. Eur. Transp. Res. Rev. 6 4 , 425–438. doi:10.1007/s12544-014-0142-4