



Identifying urban features for vulnerable road user safety in Europe

Marina Klanjčić^{1,2}, Laetitia Gauvin¹, Michele Tizzoni¹ and Michael Szell^{1,3*} 

*Correspondence: misz@itu.dk

¹ISI Foundation, Via Chisola 5,
10126, Torino, Italy

³Computer Science Department, IT
University of Copenhagen, Rued
Langgaards Vej 7, 2300, København,
Denmark

Full list of author information is
available at the end of the article

Abstract

One of the targets of the UN Sustainable Development Goals is to substantially reduce the number of global deaths and injuries from road traffic collisions. To this aim, European cities adopted various urban mobility policies, which has led to a heterogeneous number of injuries across Europe. Monitoring the discrepancies in injuries and understanding the most efficient policies are keys to achieve the objectives of Vision Zero, a multi-national road traffic safety project that aims at zero fatalities or serious injuries linked to road traffic. Here, we identify urban features that are determinants of vulnerable road user safety through the analysis of inter-mode collision data across European cities. We first build up a data set of urban road crashes and their participants from 24 cities in 5 European countries, using the widely recommended KSI indicator (killed or seriously injured individuals) as a safety performance metric. Modelling the casualty matrices including road infrastructure characteristics and modal share distribution of the different cities, we observe that cities with the highest rates of walking and cycling modal shares are the safest for the most vulnerable users. Instead, a higher presence of low-speed limited roads seems to only significantly reduce the number of injuries of car occupants. Our results suggest that policies aimed at increasing the modal share of walking and cycling are key to improve road safety for all road users.

Keywords: Urban data science; Mobility; Casualty matrix; Road safety

1 Introduction

Road traffic crashes result in yearly 1.3 million deaths and 50 million injuries, and are the world's leading cause of death for children and young adults 5–29 years of age [1]. The World Health Organization quantifies the economic costs of road traffic crashes to 3% of the global GDP, or 2.3 trillion USD. Because of this pressing societal issue, the UN has declared in 2015 the global sustainability goal to halve the number of global deaths and injuries from road traffic crashes by 2020 [2]. However, traffic deaths and injuries have kept rising worldwide instead of decreasing, and the UN goal has been missed [1].

On a global level, the WHO's explanation for this failure is the heterogeneity of progress: while casualties from road traffic have overall stagnated or decreased in high income countries, they have increased in low and middle income countries. For example, on the one hand, road fatalities have decreased in the EU (although EU-wide targets to significantly

© The Author(s) 2022. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

lower traffic crashes have been missed [3]). On the other hand, in most African and South-East Asian countries, road fatalities have stagnated or grown exceptionally high [1].

The WHO report also shows that vulnerable road users – pedestrians, cyclists, and motorcyclists – are disproportionately affected. Increased urbanization has therefore made clear that implementing effective urban planning policies at scale is necessary to overcome such failures [4]. In particular, the UN's current sustainability goal 11 to “Make cities inclusive, safe, resilient and sustainable” [2] is a key to decrease road casualties worldwide [5].

In this study, we seek to identify urban features that are determinants of vulnerable road user safety through the analysis of inter-mode collision data across European cities. We first build up a high-quality data set of urban road collisions and collision participants from 24 cities in 5 European countries, using the widely recommended KSI indicator (killed or seriously injured individuals) as a safety performance metric [6]. We then apply machine learning tools on this established data set to identify (1) the biggest danger to vulnerable traffic participants per city, and (2) the most relevant urban features – extracted from OpenStreetMap [7] – that are associated with higher safety for road users. This approach follows a human-centric urban data science [8] that aims to generate value for citizens by applying data science methods on large-scale urban data sets.

Our work follows in the footsteps of a wide literature of data-driven studies on road safety. Previous studies investigating the determinants of road safety have typically considered a subset of dimensions, including vehicle type, road infrastructure, traffic and control, environmental factors, through the regression analysis of individual crash data [9–12]. Most of them have a limited geographical coverage, usually focusing on one particular city or region, with some notable exceptions typically on policy questions [13–17]. Also, many of these studies took into account a single transport mode [18, 19] (e.g. cyclists, or pedestrian), yet increasingly on vulnerable road users [20–23], but usually only limited to the victim participant in the crash [24, 25]. In particular, among vulnerable road users, cyclists have received considerable attention by recent studies. Cycling is one of the most sustainable mobility solutions for short and medium distance trips, but faces considerable risks imposed by motorized vehicles. The risk for injury has been quantified recently in London using a multilevel regression model accounting for exposure, finding that lower speed limits and more cycling routes can be a crucial factor [26]. A more recent study of data from Spain followed a Bayesian network approach to identify the most relevant features for cyclist injury severity, finding higher risk posed by heavy goods vehicles and lower risk from certain route conditions [27]. Other approaches use GIS methods to link objective and subjective risks [28], bicycle trip data of a public bicycle rental system to proxy the bicycle crash exposure [29], crowd-sourced bicycle incident reports to characterize patterns of injury [30], spatio-temporal trends [31], and analysis of intersections or bicycle infrastructure [32–36].

To summarize, the majority of studies on urban road safety focus on crash victims, often from a single mode, and only in specific cities or regions. However, there is a clear lack of research that considers both sides of a crash from *all* traffic modes to identify inter-mode hazards, together with *multiple cities* to control for regional peculiarities.

Here we fill this gap by following the three main recommendations of the OECD for developing evidence-based approaches to road safety [37, 38]: (1) to collect and analyze crash data “from a larger set of cities”, (2) to investigate “the relationships between urban

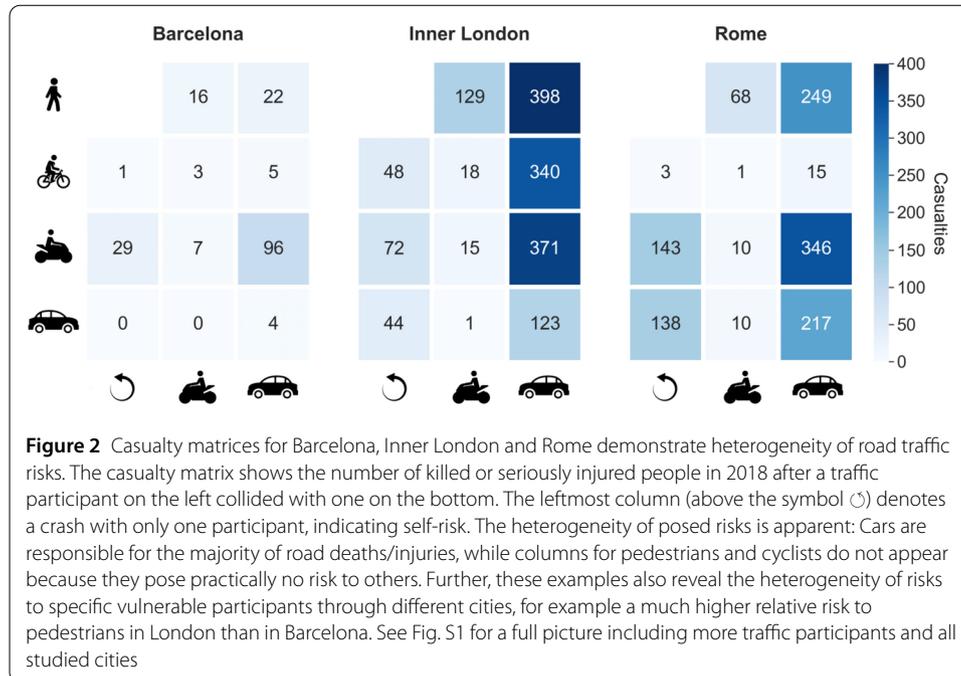
shape, density, speeds, modal share and road user risk”, and (3) to place “an immediate focus [...] on the analysis of casualty matrices to reveal the number of people in each user group who are killed or seriously injured in crashes involving another user group”. By doing so, we adopt an *ecological* study approach that takes into account all traffic modes and casualty matrices across multiple European cities, and that considers the exposure to different population-level urban features as determinants of road safety.

2 Results

2.1 Establishing a road casualty data set with inter-mode impacts

We collected road casualty data from 24 European cities in 5 countries (Spain, Italy, France, UK, and Norway) as shown in Fig. 1 from the year 2018, which was the most recent data available at the time of the study. Of the 24 cities 10 are in France and 10 are in the UK. For more details about the data collection and processing see the Methods section. The data

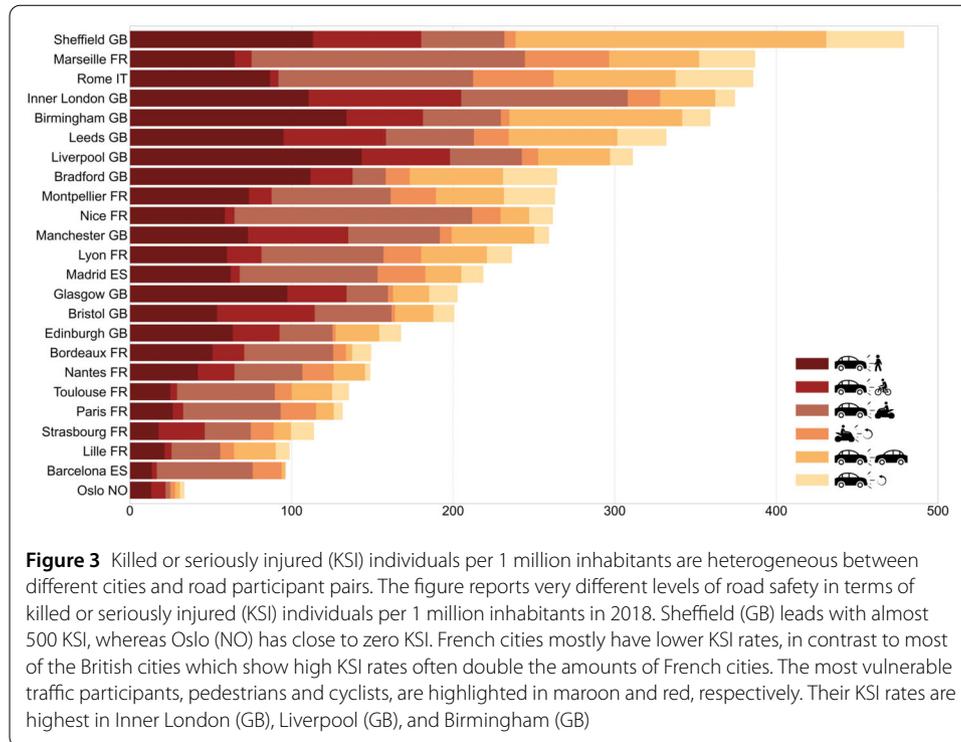




contain records of road crashes in each city, in a line list format, with details about the individuals injured, the severity of the injuries, and the types of vehicles involved. A complete description of the records is reported in the Methods section.

Based on the crash records, we created casualty matrices reporting the number of individuals killed or seriously injured (KSI) caused by the collision of any two pairs of road user types, in each city. Among all road users, we focused in particular on the vulnerable ones, that is pedestrians, cyclists, and powered two-wheelers, apart from cars. As an illustrative example, in Fig. 2, we show the casualty matrices for 3 cities: Barcelona, Inner London and Rome. Casualty matrices for all other cities are shown in the Additional file 1 (Fig. S1). While the highest risk for vulnerable users is expectedly represented by cars in all the cities, the number of KSI varies significantly by user group. For instance, the casualty matrix of Barcelona shows a high level of road safety not only for vulnerable users but for car drivers too, with only 4 KSI reported in car-car collisions in 2018.

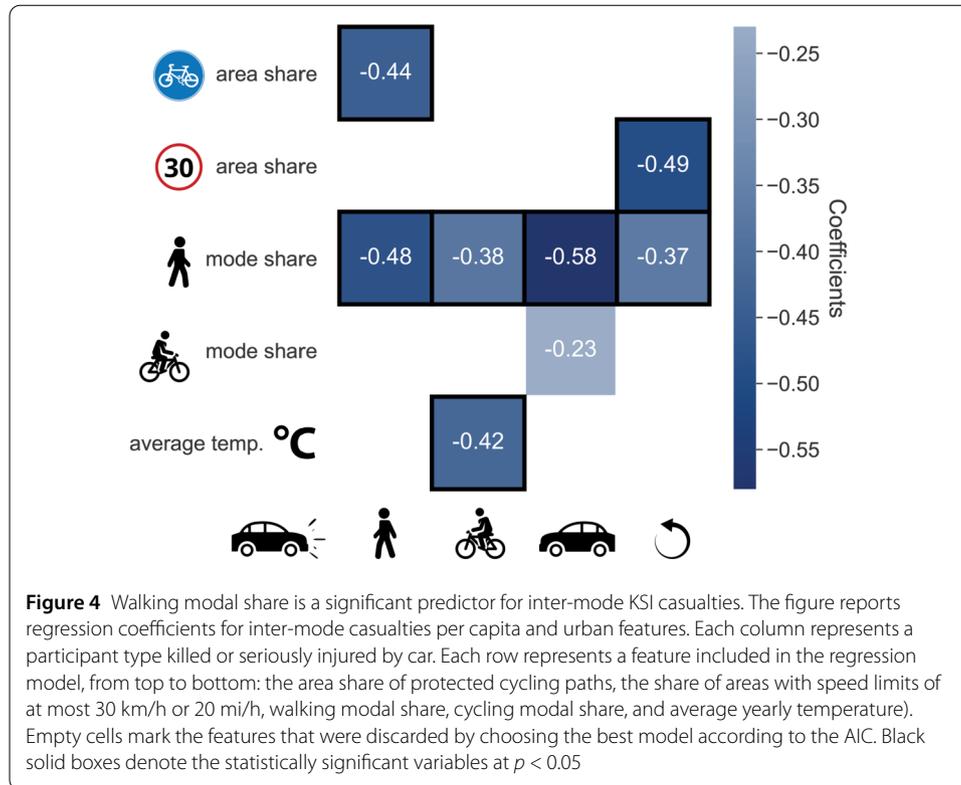
To better compare road safety levels of all cities in our dataset, we normalized the number of KSI, for each type of collision, by population size. Figure 3 shows the number of KSI individuals per 1 Million inhabitants, as a stacked bar chart, where each bar corresponds to a specific type of collision. The chart reveals the high heterogeneity in road safety across the cities under study. On the one hand, we have an extreme case like Sheffield with almost 500 KSI/M, and, at the top of the safety rank, Oslo that is the safest city in our dataset with less than 50 KSI/M in 2018. The highest KSI rates among the most vulnerable road users, pedestrians and cyclists, were recorded in Inner London (308 KSI/M), Liverpool (198 KSI/M), and Birmingham (181 KSI/M), followed by the rest of the British cities. Instead, the highest KSI rates for powered two-wheelers were reported in Marseille, Rome and Nice. British cities were also the least safe for car drivers, with Sheffield leading the rank by KSI rates in car-car crashes, immediately followed by Birmingham. French cities show medium to low rates of KSI individuals across all types of collisions, with the exception of Marseille that ranks as the second least safe city in our dataset (387 KSI/M).



National capitals also show very different levels of road safety, as Rome and Inner London display almost 400 KSI/M. while Paris ranked as the 5th safest city of our dataset, with 132 KSI/M.

2.2 Urban features as determinants of road safety

To explain the observed heterogeneities in road safety across European cities, and in particular for vulnerable users, we examined the relationship between a number of features and the inter-mode KSI rates shown in Fig. 3. We collected data regarding 7 different urban features in the 24 cities using OpenStreetMap (OSM) and from the European Platform on Mobility Management (EPOMM). We also considered climate and economic data, from Eurostat, to take into account possible confounding factors that are not directly related to the urban infrastructure of a city [39–41]. A complete description of the data collection process is reported in the Methods section. The features considered in our study are: population density, the ratio of total cycling area to total driving area, the ratio of total low-speed limited area to total driving area, modal shares for walking, cycling, public transport, and motor vehicles, the yearly average temperature, the yearly average precipitations, and the average GDP per capita. Fig. S2 provides a summary of the frequency distributions of all the features under study. Fig. S3 and Fig. S4 provide an overview of the urban features and the modal shares, respectively, in the 24 cities. All cities displayed a high variability in the urban features and modal shares, also within the same country. Population density ranges from 1417 pop/km² in Oslo to 20,000 pop/km² in Paris. The cycling area share of the total streets is only 3% in Rome but is more than 30% in Strasbourg and Nantes. The speed limited area share varies over more than an order of magnitude across cities, from 2% in Bradford to 87% in Inner London. Modal shares are also very different across the 24 cities. Paris ranks first by walking share (47%) and last by motor vehicle usage (17%). Cy-



cling modal share is generally low, below 4% in all cities, with exception of Bristol (14%), Strasbourg (8%) and Nantes (5%). Public transport leads the modal share of Barcelona (39%) while it is less common in French cities, like Montpellier (8%) and Bordeaux (9%).

For all cities, we examined the relationship between the above features and the inter-mode KSI rates by a multiple linear regression from the sets of all combinations of 2 or 3 variables, as described in the Methods. For each inter-mode KSI rate, we selected the best regression model according to the Akaike Information Criterion (AIC). Each regression coefficient β and its associated 95% confidence interval (CI) quantify the relations of each variable with the inter-mode KSI casualty rates. The main results of the models based on 2 independent variables are summarized by Fig. 4 which shows the association between each urban feature (rows) and the inter-mode KSI rate (columns) of collisions that involved at least one car and pedestrians, cyclists, or other cars. Each entry of the matrix reports the regression coefficient associated with a given feature when predicting the KSI rates of a given collision type. Negative values indicate a reduction of KSI rates and statistically significant values at $p < 0.05$ are highlighted by a solid box. Table S1 in the Additional file 1 reports the full description of the model's coefficients for all KSI rates.

First, let us focus on modal share, i.e. the middle two rows in Fig. 4. In general, larger shares of walking and cycling were most frequently associated with the smallest AIC to predict a reduction in all type of KSI rates, while use of public transport was never selected as a significant regressor. In particular, the share of walking was significantly associated with the inter-mode KSI casualty rates of all collision types. Cities with a higher walking share showed to have lower KSI rates for pedestrians ($\beta = -0.49$, 95% CI $[-0.80, -0.17]$), cyclists ($\beta = -0.38$, 95% CI $[-0.74, -0.01]$) and car/taxi occupants ($\beta = -0.58$, 95% CI $[-0.93, -0.23]$) when injured in a collision with a car or taxi. Walking share was also nega-

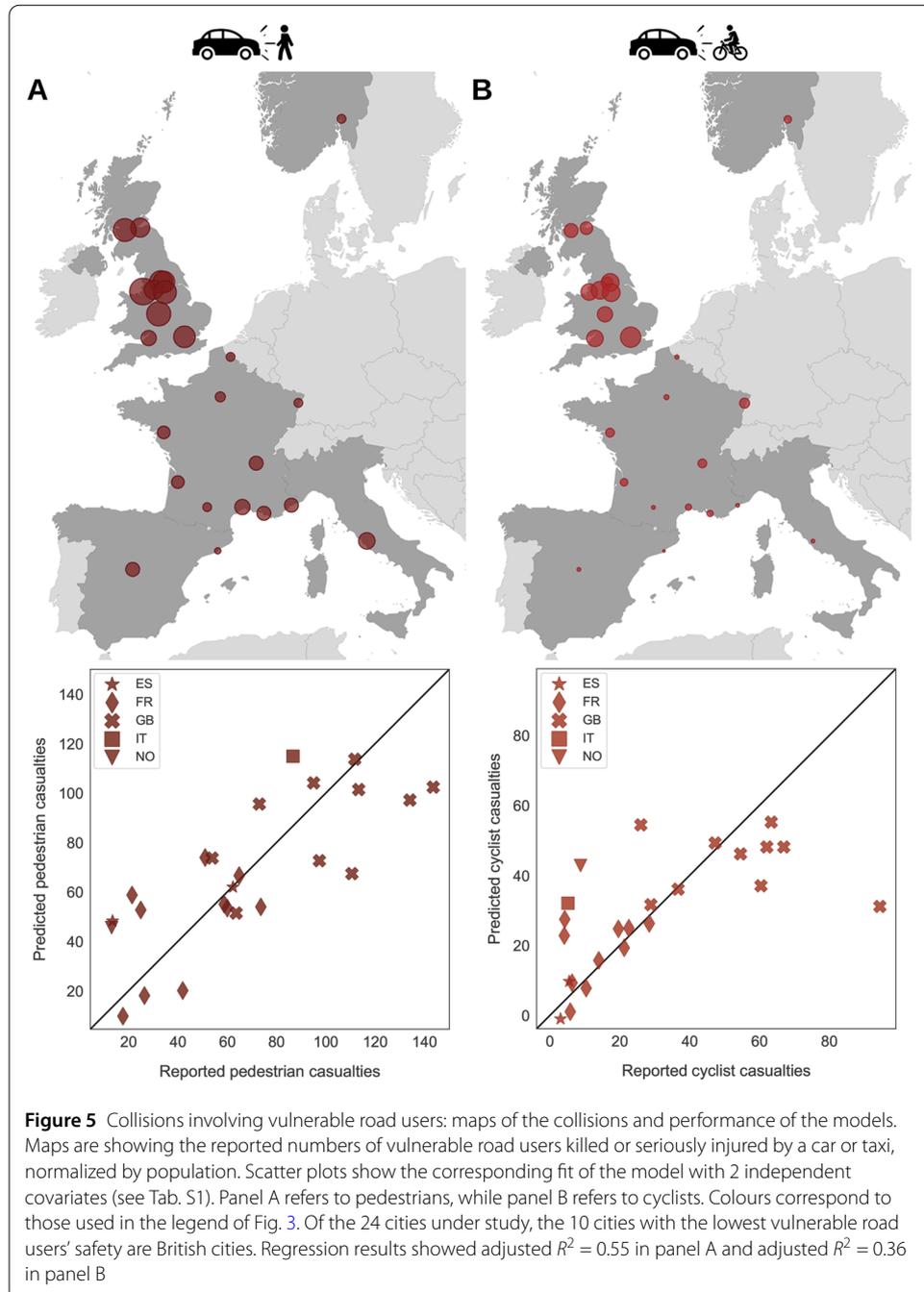
tively associated with single-vehicle car crashes, with a statistically significant coefficient $\beta = -0.37$, 95% CI $[-0.71, -0.02]$. A larger cycling share was associated, although not significantly, with lower KSI rates of car occupants, ($\beta = -0.23$, 95% CI $[-0.58, 0.12]$). Next, let us examine the features related to infrastructure, i.e. the top two rows in Fig. 4. The model showed that cities with a higher proportion of low speed limited streets with respect to the total driving area (second row in Fig. 4), are characterized by lower KSI rates for single-vehicle car crashes ($\beta = -0.49$, 95% CI $[-0.83, -0.14]$, significant). With pedestrian KSI rates, the proportion of low-speed limited streets had no detectable relation. When it comes to the proportion of protected cycling paths (first row in Fig. 4), we found a significant effect: a larger proportion was associated to lower inter-mode KSI casualty rates for pedestrians ($\beta = -0.44$, 95% CI $[-0.75, -0.12]$). Finally, among the climate and economic variables, the only one that leads to the smallest AIC value for one model is the average temperature, which was associated with lower KSI rates for cyclists ($\beta = -0.42$, 95% CI $[-0.78, -0.12]$).

Extending the regression to include 3 different covariates, results were consistent with those observed when using 2 covariates (see Tabs. S2 and S3 in the Additional file 1). Walking modal share was always included as a regressor for lower KSI rates in all collision types. The proportion of speed limited areas appeared more frequently as a regressor, now including car-car collisions and cyclist-car collisions, but not statistically significantly.

2.3 Evaluating model performance on inter-mode KSI rates

We examined to which extent each set of 2 selected covariates explain the variations in KSI rates for each collision type that involved at least one car. Figure 5 shows the results of the regression as predicted vs. reported KSI rates, for collisions between cars and the vulnerable road users of pedestrians and cyclists. In both cases, as shown in the maps, road safety is lowest in British cities, especially for cyclists, when compared to the rest of our sample. Overall, the model reached a good performance in explaining the KSI rates of pedestrians hit by a car or taxi (adjusted $R^2 = 0.55$). The model's performance was lower (adjusted $R^2 = 0.36$) for the KSI of cyclists, as indicated by some outliers in the scatterplot. In particular, the KSI rate of cyclists in Inner London was more than double than what the model could explain, based on the selected features. On the other hand, the model predicted relatively higher KSI rates for cyclists than those reported in Rome, Barcelona and Oslo. Model results for KSI rates of car occupants are shown in Fig. 6. The model's performance was better for collisions involving one car and no other vehicles (adjusted $R^2 = 0.45$) as KSI rates did not differ much between predicted and reported (Fig. 6(D)). The performance of the model was lower in the case of car-car collisions (adjusted $R^2 = 0.36$), mostly due to a single large outlier – Sheffield – where the reported KSI rate was 192 KSI/M but the model predicted a value below 100 KSI/M. On the other hand, the model was better able to explain KSI rates of car occupants in countries characterized by mid to low KSI rates (<50), like France and Spain.

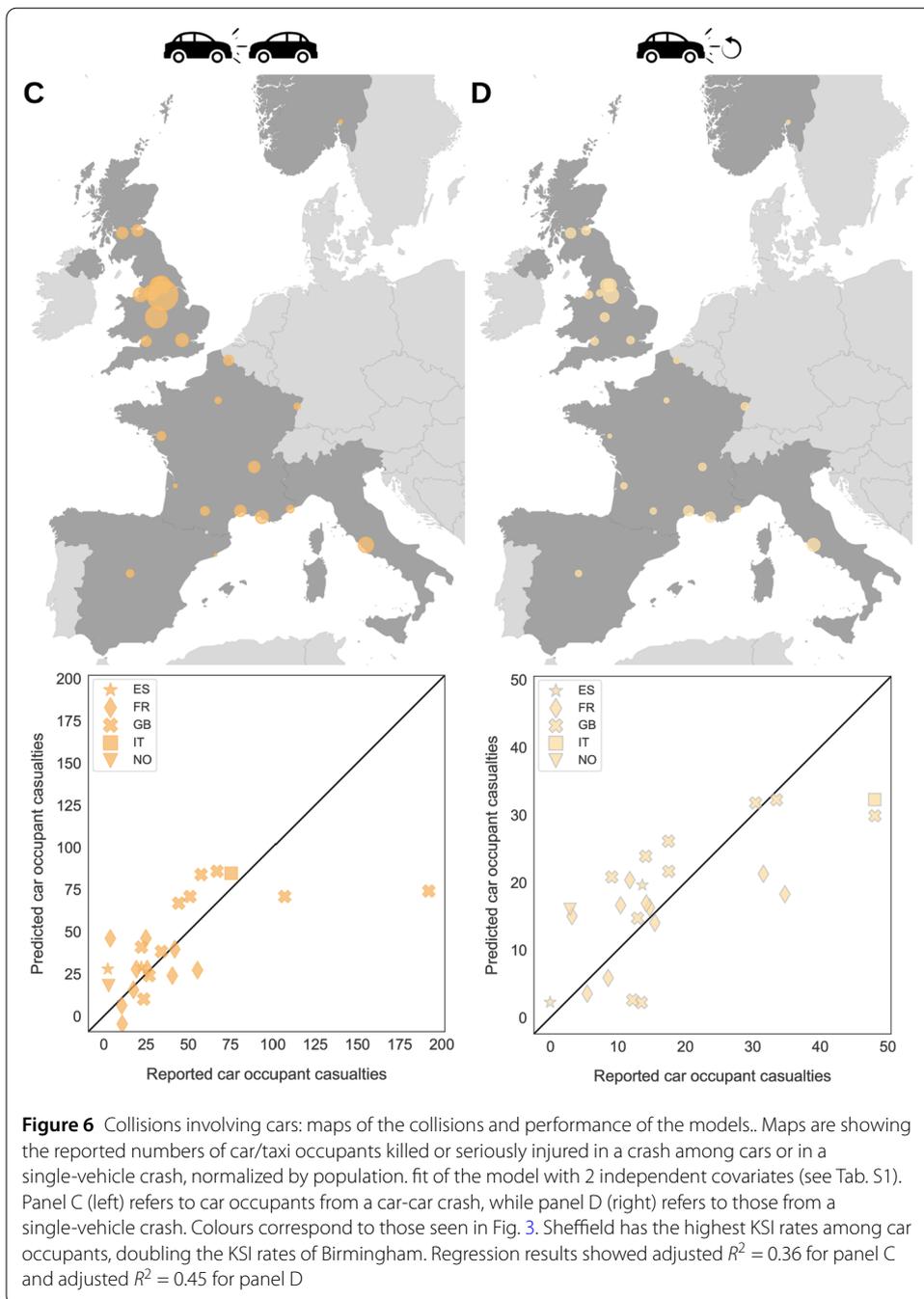
We also investigated the determinants of KSI rates of powered two wheelers (PTW) in collisions involving one car or one single vehicle. In this case, our results consistently showed a higher average temperature to be the most significant predictor of higher KSI rates (see Tabs. S2 and S3, and Fig. S5). This clearly hints at the average temperature to be a proxy for PTW modal share, an information that is missing in our dataset. A higher proportion of speed limited areas and of cycling paths were also associated with lower



PTW KSI rates, leading to an overall good performance of the regression model (adjusted $R^2 = 0.56$).

3 Discussion

In this study, we have shown that cities whose residents are more inclined to walk or cycle in their everyday life are safer for vulnerable road users. Interestingly, the effect of pedestrian modal share extends beyond vulnerable users and such cities also see less deaths or serious injuries among car occupants. Our observation that a high rate of walking and cycling is associated with a smaller number of deaths and serious injuries was already noted



by a seminal study of Jacobsen [42]. Our results confirm that early finding, and extend it by showing that more walkers and cyclists imply more safety for drivers too. Even though there have been significant efforts in recent years to integrate road safety into urban mobility plans of many cities, the incentives to walk or cycle remain among the most promising routes to make cities safer for pedestrians, cyclists and drivers. A notable example is the city of Oslo, which has successfully reached the Vision Zero milestone of zero vulnerable road deaths in 2019, through a concerted effort to turn roadway decision-making from car-centric to people-centric [43]. Another conclusion of our study is the relative impact of low-speed limited roads on vulnerable users. According to our analysis, a larger pro-

portion of speed limited roads is associated with a smaller number of injuries involving car drivers, but there is no clear association with the number of casualties among cyclists and pedestrians.

In the interpretation of the results, it is important to note that our study comes with limitations. We extracted urban features such as city area, protected cycle paths, and low-speed limited zones, from the volunteered geographic information platform OpenStreetMap using OSMnx [7]. Collecting data in this way, we were only able to access the most up-to-date information in each city but we are missing historical records of the urban features under study, thus limiting the investigation of causal effects between the temporal evolution of infrastructures and road injuries.

Nevertheless, these crowdsourced data, which have been shown to be reliable and relatively complete in the Western world [44, 45], allowed us to provide an insightful overview of the relationship between rate of collisions and urban infrastructure. They also have been successfully used in similar urban data science contexts, as in cycling injury analysis [26], in bicycle network analysis [46–48], or in estimating traffic disruption patterns [49]. Apart from novel data sources, also state-of-the-art machine learning methods are currently innovating in road safety research, e.g. with decision trees or neural networks [50–53].

Another limitation of our study lies in the heterogeneity of the data collection process across countries. We focused on the KSI statistics as their definition is rather uniform in Europe, however, the collection of crash data may not be consistent in all countries and in particular deaths or serious injuries of vulnerable road users may go underreported [12, 54]. Several efforts are currently in place to harmonize the collection of KSI numbers in Europe, for instance the maintenance of the CARE database, a community database on road crashes resulting in death or injury for Europe [55].

Further, by definition our findings of statistical associations cannot distinguish cause and effect nor identify possible confounding factors that are not part of the data sets, and we were forced to work with a sample size of 24 cities in no more than 5 countries, due to limitations in publicly available road crash data detailed enough for our ecological analysis approach. In particular, our focus on multiple cities and modes implied restriction of the data to a common denominator, thus excluding possible additional exposure data such as driven kilometers as such data are not publicly available for multiple cities and modes.

Finally, we focused on the potential impact of urban features on the injuries of the most vulnerable road users, however the introduction of additional socioeconomic factors into the model, such as per capita expenditure on alcohol, or age cohorts [56, 57], if available cross-country, could increase its predictive power and better explain the reported KSI rates by user groups in European cities.

Despite these limitations, our results are in line with concrete policy implications. For example, in recent years, several European countries have developed national walking and cycling strategies aimed at improving pedestrian and cyclist safety. However, only six European countries have drafted a national walking strategy and among them, only Finland and Luxembourg have defined a target for increasing the walking modal share [54]. Our results suggest that setting concrete targets for increasing modal shares of walking and cycling represents an effective strategy toward more sustainable and safer cities. Increasing these modal shares could happen through a human-centric mobility space re-allocation, such as pedestrianization or the substantial extension of protected urban cycling infrastructure [58] towards more livable cities, for example following a “Superblock” approach as

pioneered in Barcelona [59]. Our results are fully compatible with policy strategies developed both on the EU and OECD level towards redistributing road space [60] and towards systemic decrease of car-dependence and increase of attractiveness of sustainable modes of transport [61].

4 Methods

4.1 Data collection

We used data from various sources, as shown in Table S1. Data on road crashes were downloaded from national open data portals, with the exception of the data for Oslo, which was provided by the Norwegian Public Roads Administration upon request. Road crash statistics relate to personal injury crashes on public roads that were reported to the police in 2018. Population estimates for the same year were collected from the corresponding National Statistics Office of each country.

Data on urban features were downloaded from OpenStreetMap (OSM), a free, editable map of the world, built by volunteers. We used OSMnx, a Python package for modelling, projecting, visualization, and analysis of real-world street networks from OSM's APIs [7], to collect the following urban features:

- *City area* in km². We selected the administrative surface of a city.
- *Driving area* in km. We selected all the drivable streets by choosing `drive` as network type.
- *Cycling area* in km. We selected all the protected cycling paths by choosing `bike` as network type and by specifying related custom filters.
- *Speed limited area* in km. We selected all the streets with speed limit of ≤ 30 km/h or ≤ 20 mi/h by choosing `drive` as network type and by specifying related custom filters.

Modal share percentages in walking, cycling, public transport and motor vehicles were gathered from the European Platform on Mobility Management (EPOMM), a network of governments in European countries, represented by the Ministries responsible for Mobility Management. They developed The EPOMM Modal Split Tool (TEMS) with comparable modal split data from European cities with more than 100.000 inhabitants.

Climate data (average yearly temperature and average yearly precipitations) were collected from Wikipedia, reporting official measurements from national meteorological institutes. The average GDP per capita of each city, at the NUTS 3 level, is available from the European Statistical Office (Eurostat).

Finally, the full list of features that we use in our analysis is the following:

1. *Population density*. Population per km².
2. *Cycling area share*. The ratio of cycling area and driving area.
3. *Speed limit area share*. The ratio of speed limited area and driving area.
4. *Walking mode share* in percent.
5. *Cycling mode share* in percent.
6. *Public transport mode share* in percent.
7. *Motor vehicles mode share* in percent.
8. *Average yearly temperature* (°C).
9. *Average yearly precipitation* (mm).
10. *Average GDP per capita* (Euros) in the year 2018.

4.2 Casualty matrix

Raw data on road crashes was cleaned and transformed to show only relevant information used for the casualty matrix calculation. Each row of the cleaned data set corresponds to a unique casualty, while columns contain the following details:

- *Crash Index*. Unique index for each crash, used to connect vehicles and casualties to the corresponding crash.
- *Date*.
- *Number of Vehicles*. Total count of vehicles in a crash.
- *Number of Casualties*. Total count of casualties in a crash.
- *Vehicle Reference*. Reference to each vehicle in a crash, used to connect vehicles with the corresponding casualty.
- *Vehicle Type*. Options: Bicycle, Powered Two-Wheeler (PTW), Car/Taxi, Bus/Coach, Goods Vehicle or Other Vehicles.
- *Casualty Reference*. Reference to each casualty in a crash, used to connect casualties with the corresponding vehicle.
- *Casualty Class*. Options: Driver, Passenger, Pedestrian.
- *Casualty Type*. Options: Pedestrian, Cyclist, PTW occupant, Car/Taxi occupant, Bus/Coach occupant, Goods Vehicle occupant or Other Vehicles occupant.
- *Casualty Severity*. Options: killed (on spot or died within 30 days of the crash), seriously injured (hospitalized for >24 hours) or slightly injured (hospitalized for ≤ 24 hours).

Casualty Type information was available only in the UK data set which made the casualty matrix calculation easier, so we also formed this column in the rest of the data sets based on the *Casualty Class* and *Vehicle Type* columns. This enabled us to base our analysis on the number of inter-mode casualties, instead of the common approach focusing on the total number of casualties per each type [37]. For example, a pedestrian casualty from a crash between two cars and a pedestrian was counted as a pedestrian injured in a pedestrian-car crash. Similarly, an injured car occupant from a crash with four cars was counted as a car occupant injured in an car-car crash. Regarding the casualty severity levels, casualties with slight injuries were removed from the data set and only killed or seriously injured (KSI) people were observed. We eliminated casualties from crashes with >2 different parties involved (including pedestrians), as they represented $\leq 2\%$ of total KSI casualties in each city, which aligns with previous research [50]. Also, all the crashes with missing relevant data (mentioned above) were not taken into account.

From the newly created data set, we formed two pivot tables, one with *Vehicle Type* counts as columns, and another one with *Casualty Type* counts as columns. This time, each row of both tables corresponded to a unique crash. These two tables were joined into a single table based on *Crash Index* and we queried them twice for all possible casualty-vehicle pairs – at first for only fatal casualties and then for the seriously injured ones. These counts were used to create the KSI casualty matrix for each city. Rows of the matrix represent casualty types, while columns represent vehicle types. Finally, each matrix cell represents the number of casualties from one casualty-vehicle pair. For the next steps, we observed only the following six casualty-vehicle pairs from the casualty matrix (we chose the pairs with median value > 5):

- pedestrian – car (pedestrians killed or seriously injured in a crash between pedestrians and cars/taxis).

- cyclist – car (cyclists killed or seriously injured in a crash between bicycles and cars/taxis).
- PTW – itself (PTW occupants killed or seriously injured in a single-vehicle crash).
- PTW – car (PTW occupants killed or seriously injured in a crash between PTWs and cars/taxis).
- car – itself (car/taxi occupants killed or seriously injured in a single-vehicle crash).
- car – car (car/taxi occupants killed or seriously injured in a crash between two or more cars/taxis).

4.3 Linear regression models

To explain the potential relations between the independent features (10 input variables) and the number of inter-mode casualties (6 target variables), we used a multilinear regression model. More specifically, we fit through Ordinary Least Squares a regression of the form:

$$\mathbf{y} = \beta X, \quad (1)$$

where the response vector \mathbf{y} represents one of the inter-mode casualty rates and X represents the matrix of predictors, and β is a vector of regression coefficients. The input variables were standardized by scaling variance to one and centering mean to zero. The target variables were firstly normalized by population (per 1 million inhabitants) and then standardized the same way as the input variables. Given the limited number of observations, 24 in total, for each inter-mode KSI rate, we compared linear models with all combinations of 2 and 3 different response variables, to have an adequate number of observations per covariate estimated. We selected the best model using the Akaike Information Criterion (AIC). Smaller values of AIC indicate better quality of the model, and we identified the best model as the one with the smallest AIC value by examining all possible linear combinations of 2 and 3 regressors.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-022-00339-5>.

Additional file 1. Supplementary information (PDF 5.6 MB)

Acknowledgements

We thank Monica Knoblauch Brathaug from the Norwegian Public Roads Administration for the data provided. We also thank Geoff Boeing for helpful comments concerning OSMnx, and Anastassia Vybornova for acquisition of meteorological and economic data. Transportation icons designed by Freepik.

Funding

We acknowledge the support of the Lagrange Project funded by the CRT Foundation.

Availability of data and materials

Computed inter-mode road crashes data and urban features for 24 cities, used in this study, are available at: <https://github.com/klmrrn/urbanroadsafety>. Other data sources are reported in Table S1. The Python code developed for the data analysis is available at: <https://github.com/klmrrn/urbanroadsafety>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MK acquired and processed the data. All authors contributed to the conception and design of the work, analysis and interpretation of the data, and drafted the work. All authors read and approved the final manuscript.

Author details

¹ISI Foundation, Via Chisola 5, 10126, Torino, Italy. ²LINKS Foundation, Via Pier Carlo Boggio 61, 10138, Torino, Italy.

³Computer Science Department, IT University of Copenhagen, Rued Langgaards Vej 7, 2300, København, Denmark.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 11 June 2021 Accepted: 21 March 2022 Published online: 28 April 2022

References

1. Global Status Report on Road Safety 2018 (2018) World Health Organization, Paris
2. United Nations General Assembly (2015) Transforming our world: the 2030 agenda for sustainable development. a/res/70/1. Technical report, United Nations General Assembly
3. European Commission (2019) EU road safety policy framework 2021-2030 – next steps towards “vision zero”. Technical report, European Commission
4. Khayesi M (2020) Vulnerable road users or vulnerable transport planning? *Front Sustain Cities* 2:25
5. Klopp JM, Petretta DL (2017) The urban sustainable development goal: indicators, complexity and the politics of measuring cities. *Cities* 63:92–97
6. Utrianen R, Pöllänen M, Liimatainen H (2018) Road safety comparisons with international data on seriously injured. *Transp Policy* 66:138–145
7. Boeing G (2017) OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput Environ Urban Syst* 65:126–139
8. Resch B, Szell M (2019) Human-centric data science for urban studies. *ISPRS Int J Geo-Inf* 8(12):584
9. Walton D, Jenkins D, Thoreau R, Kingham S, Keall M (2020) Why is the rate of annual road fatalities increasing? A unit record analysis of New Zealand data (2010–2017). *J Saf Res* 72:67–74
10. Cantill V, Márquez L, Díaz CJ (2020) An exploratory analysis of factors associated with traffic crashes severity in Cartagena, Colombia. *Accid Anal Prev* 146:105749
11. Fountas G, Fonzone A, Gharavi N, Ryez T (2020) The joint effect of weather and lighting conditions on injury severities of single-vehicle accidents. *Anal Methods Accid Res* 27:100124
12. Olszewski P, Szagala P, Rabczenko D, Zielińska A (2019) Investigating safety of vulnerable road users in selected EU countries. *J Saf Res* 68:49–57
13. Branion-Calles M, Götschi T, Nelson T, Anaya-Boig E, Avila-Palencia I, Castro A, Cole-Hunter T, de Nazelle A, Dons E, Gaupp-Berghausen M, Gerike R, Panis LI, Kahlmeier S, Nieuwenhuijsen M, Rojas-Rueda D, Winters M (2020) Cyclist crash rates and risk factors in a prospective cohort in seven European cities. *Accid Anal Prev* 141:105540
14. den Berghe WV, Schachner M, Sgarra V, Christie N (2020) The association between national culture, road safety performance and support for policy measures. *IATSS Res* 44:197–211
15. Safarpour H, Khorasani-Zavareh D, Mohammadi R (2020) The common road safety approaches: a scoping review and thematic analysis. *Chin J Traumatol* 23:113–121
16. Mohan D, Tiwari G, Varghese M, Bhalla K, John D, Saran A, White H (2020) PROTOCOL: effectiveness of road safety interventions: an evidence and gap map. *Campbell Syst Rev* 16:e1077
17. Chen F, Lyu J, Wang T (2020) Benchmarking road safety development across OECD countries: an empirical analysis for a decade. *Accid Anal Prev* 147:105752
18. Blaizot S, Papon F, Haddak MM, 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
19. Batouli G, Guo M, Janson B, Marshall W (2020) Analysis of pedestrian-vehicle crash injury severity factors in Colorado 2006–2016. *Accid Anal Prev* 148:105782
20. Yannis G, Nikolaou D, Laiou A, Stürmer YA, Buttler I, Jankowska-Karpa D (2020) Vulnerable road users: cross-cultural perspectives on performance and attitudes. *IATSS Res* 44:220–229
21. Värnild A, Tillgren P, Larm P (2020) What types of injuries did seriously injured pedestrians and cyclists receive in a Swedish urban region in the time period 2003–2017 when Vision Zero was implemented? *Publ Health* 181:59–64
22. Meuleners LB, Fraser M, Johnson M, Stevenson M, Rose G, Oxley J (2020) Characteristics of the road infrastructure and injurious cyclist crashes resulting in a hospitalisation. *Accid Anal Prev* 136:105407
23. Vilaca M, Silva N, Coelho MC (2017) Statistical analysis of the occurrence and severity of crashes involving vulnerable road users. *Transp Res Proc* 20:1113–1120
24. Te Brömmelstroet M (2020) Framing systemic traffic violence: media coverage of Dutch traffic crashes. *Transp Res Interdiscip Perspect* 5:100109
25. Verkade T, Te Brömmelstroet M (2020) Het Recht van de Snelste: Hoe Ons Verkeer Steeds Asocialer Wordt. De Correspondent, Amsterdam
26. Aldred R, Goodman A, Gulliver J, Woodcock J (2018) Cycling injury risk in London: a case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits. *Accid Anal Prev* 117:75–84
27. Aldred R, García-Herrero S, Anaya E, Herrera S, Mariscal MÁ (2020) Cyclist injury severity in Spain: a Bayesian analysis of police road injury data focusing on involved vehicles and route environment. *Int J Environ Res Public Health* 17(1):96
28. von Stülpnagel R, Lucas J (2020) Crash risk and subjective risk perception during urban cycling: evidence for congruent and incongruent sources. *Accid Anal Prev* 142:105584

29. Ding H, Sze NN, Li H, Guo Y (2020) Roles of infrastructure and land use in bicycle crash exposure and frequency: T A case study using Greater London bike sharing data. *Accid Anal Prev* 144:105652
30. Fischer J, Nelson T, Laberee K, Winters M (2020) What does crowdsourced data tell us about bicycling injury? A case study in a mid-sized Canadian city. *Accid Anal Prev* 145:105695
31. Carvajal GA, Sarmiento OL, Medaglia AL, Cabrales S, Rodríguez DA, Quistberg DA, López S (2020) Bicycle safety in Bogotá: a seven-year analysis of bicyclists' collisions and fatalities. *Accid Anal Prev* 144:105596
32. Ling R, Rothman L, Cloutier M-S, Macarthur C, Howard A (2020) Cyclist-motor vehicle collisions before and after implementation of cycle tracks in Toronto, Canada. *Accid Anal Prev* 135:105360
33. Bahrololoom S, Young W, Logan D (2020) Modelling injury severity of bicyclists in bicycle-car crashes at intersections. *Accid Anal Prev* 144:105597
34. Marshall WE, Ferenchak NN (2019) Why cities with high bicycling rates are safer for all road users. *J Transp Health* 13:285–301
35. Pedroso FE, Angriman F, Bellows AL, Taylor K (2016) Bicycle use and cyclist safety following Boston's bicycle infrastructure expansion, 2009–2012. *Am J Publ Health* 106:2171–2177
36. Cicchino JB, McCarthy ML, Newgard CD, Wall SP, Maggio CJD, Kulie PE, Arnold BN, Zuby DS (2020) Not all protected bike lanes are the same: infrastructure and risk of cyclist collisions and falls leading to emergency department visits in three U.S. cities. *Accid Anal Prev* 141:105490
37. Santacreu A (2018) Cycling safety: summary and conclusions of the ITF roundtable on cycling safety, 29–30 January 2018, Paris
38. ITF (2019) Road safety in European cities: performance indicators and governance solutions. OECD Publishing, Paris. Technical report, International Transport Forum Policy Papers, No. 67
39. Spencer P, Watts R, Vivanco L, Flynn B (2013) The effect of environmental factors on bicycle commuters in Vermont: influences of a northern climate. *J Transp Geogr* 31:11–17
40. Behnood A, Mannering F (2017) Determinants of bicyclist injury severities in bicycle-vehicle crashes: a random parameters approach with heterogeneity in means and variances. *Anal Methods Accid Res* 16:35–47
41. van Beeck EF, Borsboom GJ, Mackenbach JP (2000) Economic development and traffic accident mortality in the industrialized world, 1962–1990. *Int J Epidemiol* 29(3):503–509
42. Jacobsen PL (2003) Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Inj Prev* 9(3):205–209
43. Hartmann A, Abel S (2020) How Oslo achieved zero. *ITE J* 90(5):32–38
44. Haklay M (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ Plan B, Plan Des* 37(4):682–703
45. Barrington-Leigh C, Millard-Ball A (2017) The world's user-generated road map is more than 80. *PLoS ONE* 12(8):0180698
46. Olmos LE, Tadeo MS, Vlachogiannis D, Alhasoun F, Alegre XE, Ochoa C, Targa F, González MC (2020) A data science framework for planning the growth of bicycle infrastructures. *Transp Res, Part C* 115:102640
47. Natera Orozco LG, Battiston F, Iñiguez G, Szell M (2020) Extracting the multimodal fingerprint of urban transportation networks. *Transport Findings* 13171
48. Natera Orozco LG, Battiston F, Iñiguez G, Szell M (2020) Data-driven strategies for optimal bicycle network growth. *R Soc Open Sci* 7:201130
49. Camargo CQ, Bright J, McNeill G, Raman S, Hale SA (2020) Estimating traffic disruption patterns with volunteered geographic information. *Sci Rep* 10(1):1–8
50. Prati G, Pietrantonio L, Fraboni F (2017) Using data mining techniques to predict the severity of bicycle crashes. *Accid Anal Prev* 101:44–54
51. Montella A, de Oña R, Mauriello F, Riccardi MR, Silvestro G (2020) A data mining approach to investigate patterns of powered two-wheeler crashes in Spain. *Accid Anal Prev* 134:105251
52. Yu L, Du B, Hu X, Sun L, Han L, Lv W (2021) Deep spatio-temporal graph convolutional network for traffic accident prediction. *Neurocomputing* 423:135–147
53. Roland J, Way PD, Firat C, Doan T-N, Sartipi M (2021) Modeling and predicting vehicle accident occurrence in Chattanooga, Tennessee. *Accid Anal Prev* 149:105860
54. Admainité-Fodor D, Jost G (2020) How safe is walking and cycling in Europe? PIN Flash Report 38, European Transport Safety Council
55. Bauer R, Machata K, Brandstaetter C, Yannis G, Laiou A, Folla K (2016) Road traffic accidents in European urban areas. In: Proceedings of the 1st European Road Infrastructure Congress, Leeds, pp 18–20
56. Noland RB, Quddus MA (2004) Analysis of pedestrian and bicycle casualties with regional panel data. *Transp Res Rec* 1897(1):28–33
57. Nikolaou P, Folla K, Dimitriou L, Yannis G (2021) European countries' road safety evaluation by taking into account multiple classes of fatalities. *Transp Res Proc* 52:284–291
58. Szell M, Mimar S, Perlman T, Ghoshal G, Sinatra R (2022, in print) Growing urban bicycle networks. *Sci Rep*
59. Nieuwenhuijsen MJ (2020) Urban and transport planning pathways to carbon neutral, liveable and healthy cities; a review of the current evidence. *Environ Int* 140:105661
60. European Commission, Directorate-General for the Environment (2004) Reclaiming city streets for people: chaos or quality of life? Technical report
61. (2021) Transport strategies for net-zero systems by design. Technical report, OECD Publishing