

Impact of COVID-19 travel-restriction policies on road traffic accident patterns with emphasis on cyclists: A case study of New York City

Jintai Li, Zhan Zhao*

Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China



ARTICLE INFO

Keywords:
Road traffic accidents
Cyclists
COVID-19
Travel restrictions
Data fusion

ABSTRACT

Since the COVID-19 outbreak, travel-restriction policies widely adopted by cities across the world played a profound role in reshaping urban travel patterns. At the same time, there has been an increase in both cycling trips and traffic accidents involving cyclists. This paper aims to provide new insights and policy guidance regarding the effect of COVID-19 related travel-restriction policies on the road traffic accident patterns, with an emphasis on cyclists' safety. Specifically, by analysing the accidents data in the New York City and estimating three fixed effects logit models on the occurrence of different types of accidents in a given zip code area and time interval, we derived the following findings. First, while the overall number of road traffic accidents plummeted in the NYC after the stay-at-home policy was implemented, the average severity increased. The average number of cyclists killed or injured per accidents more than tripled relative to levels in similar times in previous years. Second, the declaration of the New York State stay-at-home order was significantly associated with a higher risk of accidents resulting in casualties. The number of Citi Bike trips in the area at the time overwhelmingly predicted severe risk for cyclists. Last, we applied the models to detect hot zones for cyclists' severe accidents. We found that these hot zones tend to be spatially and temporally concentrated, making it possible to devise targeted safety measures. This paper contributes to the understanding of the impact of COVID-19 travel-restriction policies on accidents involving cyclists, reveals higher risks for cyclists as an unintended consequence of travel-restriction policies, and provides an analytical tool for road safety impact evaluation should future travel restrictions be considered.

1. Introduction

Since the coronavirus disease 2019 (COVID-19) first started spreading on a large scale in Dec 2019 in Wuhan, China (WHO, 2020), the transmissible disease has changed urban life around the world profoundly almost everywhere it spreads. As of this writing, a total of 195 million cases have been confirmed around the world, and over 4 million deaths have been attributed to COVID-19 (WHO, 2021). In the early days of COVID-19, it seemed the best way to curb transmission was through social distancing and density reduction. Therefore, many cities around the world adopted different forms of policies restricting residents' travel mobility. On Jan 23, 2020, all public transportation services were suspended in Wuhan and residents were not allowed to leave the city without permission from the authorities. In Feb 2020, cities in Lombardy, Italy imposed a set of measures that banned 'non-essential' travel and limited free movement of its residents. Similar measures were later initiated around the world.

These travel-restriction policies fundamentally changed the urban travel patterns for an extended period. Firstly, we saw a drop in travel behaviour in terms of number of trips taken. Some densely populated cities saw a decline of 50% to 62.9% in overall mobility (Saladié et al., 2020; Yabe et al., 2020; Zhang et al., 2021). Secondly, there has been a shift among different modes. Notably, the cycling mode share has significantly increased in several cities. Lastly, the total number of road traffic accidents has decreased. Across cities that have implemented COVID-19 related travel-restriction policies, the number of accidents decreased by 41% to as much as 74% (Aloi et al., 2020; Calderon-Anyosa & Kaufman, 2021; Katrakazas et al., 2020; Saladié et al., 2020).

In the United States, one of the epicentres of transmission has been the New York City (NYC). The first case of COVID-19 in the New York State (NY) was confirmed on Mar 1, 2020 (West, 2020). A state of emergency was declared in NY on Mar 7 (McKinley and Sandoval, 2020). Then on Mar 20, Governor Cuomo declared a state-wide stay-at-home order. All 'non-essential' businesses were ordered to close, and all

* Corresponding author.

E-mail addresses: lijintai09@gmail.com (J. Li), zhanzhao@hku.hk (Z. Zhao).

'non-essential' gatherings cancelled or postponed (Francescani, 2020). Apart from the fact that these plans reduce travel demand substantially, the order also asked residents to 'limit use of public transportation to when absolutely necessary'. Although the state started its reopening on Jun 8, the diminishing effect of the stay-at-home order on travel behaviour proved to be profound and long-lasting. These changes produced some interesting phenomena in road accident patterns. In particular, accidents involving cyclists became much more prominent.

We found two phenomena to be insufficiently discussed in the existing literature. Firstly, COVID-19 related travel restrictions have been widely and continually imposed by government across the world, but their detailed impact on the risk and mechanism of road traffic accidents are not well understood. Secondly, cycling as a travel mode has gradually seen wider adoption and has proven reliable during COVID-19 infections, but we lack understanding of the safety risks to cyclists caused by COVID-19 related travel restrictions. This study is designed to contribute to these two gaps in the literature.

In this paper, we investigated the road accident patterns and Citi Bike usage patterns in the NYC before and after the implementation of the NY stay-at-home order. We then built a set of fixed effects models to interpret the correlation between time-dependent attributes of all NYC zip codes and the likelihood of severe accidents occurring in these zip codes, with an emphasis on accidents involving cyclists. We highlight the following findings and contributions.

Firstly, we found after imposing the stay-at-home order, while the total number of accidents dramatically decreased, the average number of persons killed or injured in each accident saw a large increase. This finding is concerning considering the cost of these accidents as well as policy priorities such as Vision Zero. Particularly, the average number of cyclists killed or injured in each accident more than tripled compared with previous years. At the same time, the number of trips taken on Citi Bike did not exceed similar times in 2019. Therefore, it is likely that cycling in the NYC has become much more dangerous.

Secondly, through the logit models, we found both the stay-at-home order and a higher number of new COVID-19 cases were correlated with significantly lower risk for the occurrence of all accidents, suggesting different rationale for the reduction of accidents. However, the stay-at-home order is also correlated with a higher risk of casualties for each accident. Furthermore, a higher number of Citi Bike trips in the zip code and a higher new COVID-19 case number are associated with the risk of severe accidents involving cyclists.

Lastly, we applied the models to identify high-risk temporal and spatial cross sections (hot zones) for severe accidents for cyclists. We found these hot zones to be relatively heavily concentrated both temporally and spatially. They usually appear on the immediate peripheries of Manhattan during the afternoon rush hour. They can be further identified using proxies such as the trips volume of Citi Bike. The proposed methodology could help city managers to create a safer environment for cyclists and provide guidance in future situations where further travel-restriction policies are being considered for transmissible diseases.

2. Literature review

2.1. Impact of travel-restriction policies on travel patterns

Since the start of worldwide transmission of COVID-19, long-distance modes of transportation suffered some of the largest negative impacts. National governments broadly limited travel from or to other countries, greatly diminishing long-distance travel demand. There was an estimated 59% to 60% decline in world total number of airline passengers in 2020 (Bureau, 2020). Similarly, the global maritime passenger traffic was estimated to have dropped by 20% to 43% (Millefiori et al., 2020), and rail transport experienced a 20% to 30% drop in total number of passengers in Europe, the US and Asia in 2020 (Ding & Zhang, 2021).

At the same time, the urban travel landscape has also profoundly changed. Decline in mobility and travel behaviour was widely observed across the world in cities with travel-restriction policies. In the months of March and April of 2020, overall mobility in Tarragona, Catalonia decreased by 62.9% (Saladié et al., 2020). One week after the state-of-emergency declaration in Tokyo in Apr 2020, human mobility behaviour decreased by around 50% (Yabe et al., 2020). Local travel volume in Hong Kong decreased by 52.3% (Zhang et al., 2021). In particular, public transportation usage plummeted. By the end of Mar 2020, the Metrorail ridership in Washington DC declined by 90% (Authority, 2020). In late April and early May 2020, Metropolitan Transportation Authority ridership in the NYC declined on average 60% compared to baseline (Wang et al., 2021). Even driving behaviour decreased. Vehicle miles driven decreased 35% among adolescents in Alabama (Stavrinou et al., 2020).

While travel behaviour shrunk significantly on most travel modes, cycling activity differed in certain cities. This is possibly in part driven by the finding that transmission of SARS-CoV-2 in outdoor open-air spaces is much rarer than indoor. In Switzerland, cycling increased in mode share drastically, and cycling kilometres saw a large increase that sustained into the summer (Molloy et al., 2021). In Philadelphia, cycling levels increased by about 150% during the COVID-19 outbreak (Barbarossa, 2020). Cycling has proven a resilient transportation mode under the disruption of transmissible diseases such as COVID-19.

As a result of the lowered general travel demand, in some cities there was an increase in the speed of the traffic. For example, the speed of traffic flow in the NYC increased due to lower congestion (Almagro & Orane-Hutchinson, 2020). Similar phenomena were observed in Greece and Saudi Arabia (Katrakazas et al., 2020).

2.2. Impact of travel-restriction policies on road traffic accidents

It is broadly observed that the travel-restriction policies and measures resulted in a significant reduction in the overall frequency of road traffic accidents. In the months of March and April of 2020, the number of road traffic accidents in Tarragona, Catalonia decreased by 74.3% (Saladié et al., 2020). Over a similar period of time, the number of accidents in Santander, Spain also fell by a relative 67% (Aloi et al., 2020). In Peru, the biggest change in the rate of death from external causes after COVID-19 lockdown was observed in traffic accidents, with a reduction of 12.22 deaths per million men per month (Calderon-Anyosa & Kaufman, 2021). In Louisiana, US, the stay-at-home order led to a 47% decrease in road traffic accidents. Greece saw a 41% reduction in overall accidents, with early morning crashes (0:00 – 5:00 AM) having the largest reduction, 81% (Katrakazas et al., 2020).

Apart from the overall number of road traffic accidents, the spatial distribution of accidents also changed during the travel-restriction policies. The concentration of accidents in the metropolitan areas of Tarragona and Reus decreased during lockdown in Tarragona province (Saladié et al., 2020). Over a similar period, the accident hotspots in the Los Angeles area shifted from Hollywood and northern LA to southern LA, while those in New York City shifted from Midtown and Lower Manhattan to Upper East Side, West Bronx and southern Brooklyn (Lin et al., 2020). There has been a consistent increase in the share of accidents in outlying lower density neighbourhoods in metropolitan areas.

At the same time, some researchers discovered that, despite the lower overall number of road traffic accidents, the number of casualties per accident did not decrease after the implementation of travel-restriction policies. In the state of Missouri, researchers found a significant decrease in road traffic accidents resulting in minor or no injuries, but not in those resulting in serious or fatal injuries (Qureshi et al., 2020). In effect, this suggests an increase in the average severity of the accidents. Similarly, using a countrywide dataset in the US, researchers found significant decrease in low severity crashes, but an increase in the category of crashes with the highest severity (Brodeur et al., 2021). The reason of such increase in severity observed in certain regions is most

likely multi-faceted. The severity of accidents tends to increase when vehicles crash at a higher speed (Casado-Sanz et al., 2020; Shefer, 1994). Many regions, such as Greece and Saudi Arabia, saw an increase in traffic speed after the travel-restriction policies took effect (Katrakazas et al., 2020), which could account for the increase in severity. In addition, increased consumption of alcohol and drugs as well as opportunities for speeding and stunt driving might have also played a contributing part (Vingilis et al., 2020).

Despite the rapidly accumulating body of literature on the impact of travel-restriction policies on travel patterns and road traffic accidents, there is still a lack of research focusing on the impact of travel-restriction policies on the frequency and mechanism of accidents involving cyclists. Given the increased significance and proven resilience of cycling as a travel mode during the pandemic, more research in this area will be highly beneficial. This paper sets out to fill in this gap.

2.3. Contributing factors to the frequency of road traffic accidents

There is a very wide range of factors that contribute to the frequency of road traffic accidents. They can be broadly sorted into two categories: environmental and behavioural. Here we use behavioural to refer to the factors that are related to the behaviour and characteristics of the parties involved in the accidents, and environmental to refer to the rest. This paper mainly discusses environmental factors. But behavioural ones play an important role, nonetheless.

The environmental factors further include road characteristics, weather and visibility variables, vehicle characteristics, spatio-demographic factors, etc. Researchers have found a higher local average traffic speed to contribute to lower frequency of accidents, although the effect of speed on accidents casualty is not clear (Golob & Recker, 2003; Milton & Mannerling, 1998; Quddus, 2008). Traffic volume, represented by variables such as annual average daily traffic (AADT), is broadly found to be positively correlated with the frequency of accidents. Some argue the effect of traffic volume is higher than that of speed (Abdel-Aty & Radwan, 2000; Golob & Recker, 2003). In addition, road network density has been found to contribute positively to accidents frequency in some studies (Dumbaugh et al., 2009; Rifaat et al., 2012; Wang et al., 2017), and negatively in some other studies (Marshall & Garrick, 2011). Local socio-demographic characteristics such as total population and economic activity intensity were also found to contribute to more accidents (Kim et al., 2006).

With respect to the relative importance of various risk factors, in a review of risk factors affecting the severity of accidents at intersections, researchers found speed zone, traffic control type, time of day and crash type are significantly associated with the severity of the accident (Chen et al., 2012). A systematic review ranked risk factors related to road infrastructure, and found traffic volume, traffic composition, inadequate friction on road surface, poor visibility and adverse weather to be among the most impactful risk factors (Papadimitriou et al., 2019).

The study of behavioural factors focuses on the roles of the drivers in the accidents, although a few also investigate the characteristics and behaviour of other parties, such as pedestrians and cyclists. Among the drivers' characteristics, their age (Valent et al., 2002; Yau, 2004), driving experience, gender (Regev et al., 2018), safety measures such as wearing of seatbelts (Yau, 2004), alcohol and drug usage, and mental disorder (Alavi et al., 2017) stand out as the main contributing factors in road traffic accidents.

After imposing travel-restriction policies, we can reliably expect that some of the factors reviewed in this section will change significantly. However, there is a lack of literature addressing how these sudden changes as well as the travel restrictions themselves impact the risk of accidents. This paper aims to contribute to better understanding of this relationship.

3. Data

We decided on NYC to study the change in patterns and contributing factors of road traffic accidents after significant travel-restriction measures were taken in relation to COVID-19. This choice is based on four reasons. Firstly, NYC is a city with a large population and international relevance. Its situations and policies are generally of certain referential value to other cities globally. Secondly, NYC has experienced a wide range of situations in terms of number of COVID-19 cases and related travel-restriction policies during an extended period. NYC has seen daily new case numbers as high as 16,715, or about 0.2% of the city population. It has also implemented relatively stringent lockdown rules, as well as different stages of reopening. Thirdly, cycling trips make up a significant portion of trips in NYC. About 1.1% people ride a bicycle to commute to work (Bureau, 2017). Lastly, relevant data regarding NYC during this pandemic are relatively readily available. In the following sections, we will introduce different data sources, and discuss how they were compiled to produce temporal and spatial cross sections to further analyse contributing risk factors to accidents.

3.1. New York City motor vehicle collisions data

The New York City Motor Vehicle Collisions table (Collisions Data) is provided by the New York City Police Department. Police officers are required to record all collisions where somebody is injured or killed, or where there is at least \$1,000 in damage. These records are subsequently transcribed in the Collisions Data. This dataset records qualifying vehicle accidents since 2013 until Jun 2021, as of this writing. It contains detailed information about the time and location of the accident, number of vehicles involved, vehicle types and contributing factors. In addition, the dataset contains types and numbers of persons killed and injured in each accident. In the following sections, we refer to persons either killed or injured as casualties. We also refer to accidents that result in any casualties as a severe accident.

The Collisions Data suggest interesting changes in the total number of road traffic accidents. There was a dramatic drop in the number of collisions since the start of the stay-at-home order. This is shown in Fig. 1. The dotted line indicates declaration of state of emergency in NY on Mar 8, 2020, and the red area indicates the period between the start of the stay-at-home order, Mar 20, 2020, and the start of reopening, Jun 8, 2020. The number of collisions started a sharp decline since the state of emergency was declared. It continued to decline and reached a local minimum shortly after the NY stay-at-home order took effect. Then the number of collisions gradually increased until reaching a relatively stable level, which is still much lower than the level before COVID-19.

3.2. Citi Bike trip histories data

There are different providers of bicycle mobility in the NYC, and Citi Bike accounts for a significant portion of cycling traffic. Citi Bike is a service where the company owns and rents out bikes on a time-dependent pricing scheme to its community of members. Members can access these bicycles from one of the many stations across the NYC and Jersey City, NJ, and must return them to one of the stations after use. Citi Bike publishes disaggregate trip level data every month, which includes the trip origin and destination stations, trip start and end time, etc.

The Citi Bike stations are mostly concentrated in and around Manhattan. This can be seen in Fig. 2. It is common for multiple stations to exist in the same zip code. But there are also zip codes in the NYC without any Citi Bike stations in them.

The Citi Bike trip histories data only include the origin and destination stations, and not the trip trajectories. Therefore, we used the following exponential decay method to approximate the number of trips whose trajectories intersect with a certain zip code. In travel behaviour literature, it is agreed that an exponential specification

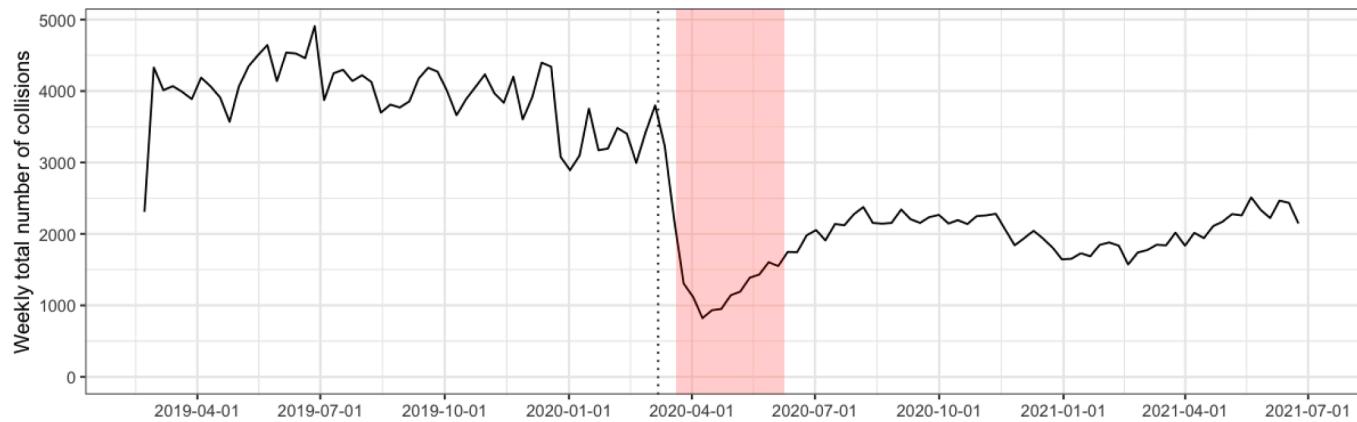


Fig. 1. Weekly Total Number of Collisions Before and After Stay-At-Home Order in NY. The X-axis indicates the date, and the Y-axis indicates the weekly total number of collisions of the following week, starting on Sundays. The horizontal dotted line indicates the declaration of a state of emergency in NY (Mar 8, 2020). The red area indicates the period between the start of the stay-at-home order, Mar 20, 2020, and the start of reopening, Jun 8, 2020. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

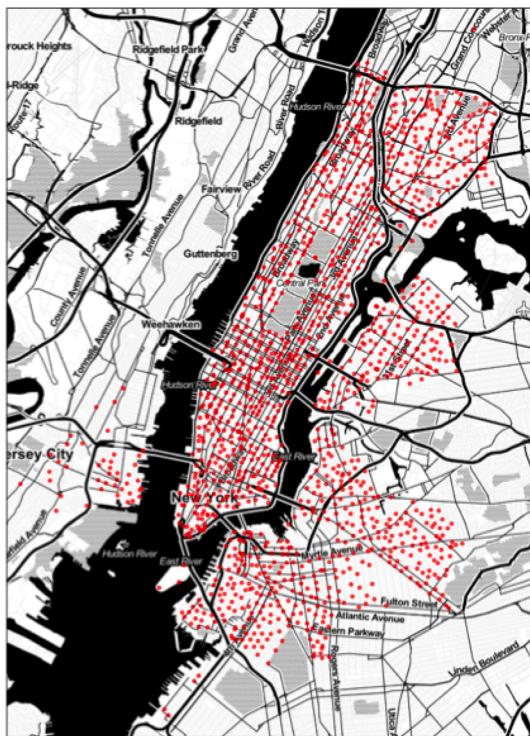


Fig. 2. Locations of Citi Bike stations in the NYC. The locations of Citi Bike stations are represented by red points. This figure is overlaid with a map showing information of general surroundings of the areas with all Citi Bike stations. The black area represents the Hudson River. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$F = \gamma_0 e^{\gamma_1 G}, \quad (1)$$

is in accordance with a utility framework. Here G is the generalized cost of a trip, and F is a decreasing function with respect to G , roughly representing the likelihood of this trip happening. This specification is most prominently used in gravity models (Gochrane, 1975).

Similarly, some studies have used exponential decay to describe the decrease in cyclists' willingness to travel as distance increases. In this study, we use the following specification,

$$P(d) = e^{-\lambda d} \quad (2)$$

where P is the cumulative percentage of cycling trips with a distance of at least d , and λ is a nonnegative parameter representing distance induced decay. This specification has been widely applied, and parameters has been estimated for trips corresponding to different modes and trip purposes (Iacono et al., 2008; Iacono et al., 2010; Wu et al., 2019). Building up on the work of Iacono et al. (2008), which divided cycling trips into work, school, shopping and recreation trips and modelled their distance exponential decay respectively, and empirical cycling trip purpose distribution in New York City from survey (NYCDOT, 2017), we determined a combined λ value for all bike trips should be 8.57×10^{-5} . Here d should be in feet.

Suppose there are totally Z zip codes and S stations. In any time interval, denote the number of trips that start from station s as o_s and the number of trips that end in station s as d_s . The approximated number of Citi Bike trips in zip code z , τ_z is calculated as

$$\tau_z = \sum_s e^{-\lambda D(s,z)} \left(\frac{o_s + d_s}{2} \right), \quad (3)$$

where $D(\bullet)$ calculates the distance between the zip code polygon and the station, and the derivation of λ is as explained in the last paragraph. Notice that because τ_z in Eq. (3) approximates the number of trips whose trajectories intersect with z , one trip can be counted multiple times and in fact even as a fraction. Therefore, the sum of τ_z 's will be greater than the total number of cycling trips. However, this is not a problem in modelling because we just use τ_z to represent the relative frequency of cycling trips in the zip code.

3.3. Other data

Besides these data sets, several other data sources were used to derive characteristics of the spatial-temporal cross sections used in modelling. Despite the different formats they originally come in, different ways of data fusion were applied in the process to map or triangulate these data onto the granularity of the spatial-temporal cross section.

The first relevant dataset is the American Community Survey 2014–2018 5-year estimates published by the Census Bureau. This dataset was used to calculate demographic characteristics of zip codes, such as the total population, gender ratio, median income and the Gini index. As a result, these characteristics vary spatially.

The second relevant dataset is the Annual Average Daily Traffic (AADT) dataset of 2019, published by the New York State Department of Transportation. This dataset contains detailed information about the annual traffic flow on each road segment in NY. These characteristics

were spatially joined with the zip codes to derive characteristics such as total AADT, total roads length, average traffic speed and average percentage of traffic that is car or truck in the zip code. As a result, these characteristics vary spatially.

The third relevant dataset is points of interest dataset published by the New York City Department of Information Technology & Telecommunications. These data were spatially joined to derive the number of points of interest in each zip code. As a result, this characteristic varies spatially.

The fourth relevant dataset is the daily NYC weather data collected from National Centers for Environmental Information. We chose the Central Park station observations to represent weather conditions in NYC, deriving characteristics such as whether the temperature is too hot (highest over 86 degrees Fahrenheit), too cold (lowest below 30 degrees Fahrenheit), there being precipitation, there being snow, there being thunder and the weather being windy (fastest 2-minute wind speed exceeding 20 miles per hour) on the day. As a result, these characteristics vary temporally.

The fifth relevant dataset is the cases by day dataset from NYC Coronavirus Disease 2019 (COVID-19) Data published by the NYC Health Department. This dataset was used to calculate the 7-day moving average of daily new case number. A moving average was chosen instead of the raw value because it represents a more stable long-term trend and is robust to temporary changes in reporting protocols. As a result, this variable varies temporally.

Lastly, we also included data relevant to the NY travel-restriction policies. On Mar 20, 2020, the NY stay-at-home order was issued, and on Jun 8, 2020, Phase 1 of NY reopening started. These policy changes were incorporated in the data.

4. Methods

4.1. Modelling framework

There are two motivations for developing models linking accidents to risk factors beyond descriptive analyses. Firstly, we want to discover important factors that contribute to a place in time having a high risk for road traffic accidents. Furthermore, we want to differentiate the sets of factors that are associated with any accidents occurring, an accident resulting in any casualties and an accident involving cyclists' causalities. An emphasis should be placed on the effect of the spread of COVID-19 and related travel-restriction policies. These models will facilitate the understanding of risk factors and possible measures to curb the probability of a certain type of accident. Secondly, we want to provide city managers with a tool for detection of cyclist's severe accident hot zones. This tool could help them to tailor their actions to specific times and spaces of high risk to cyclists, to improve their safety while minimizing disruption to the mobility system.

In assessing the risk factors of accidents, we found it obvious that many of these factors vary hugely both spatially and temporally. For example, the road network density in downtown Manhattan and Staten Island are very different. Similarly, the weather and visibility conditions at noon and at dusk are different. It is a reasonable assumption that these differences contribute to different levels of risk for accidents. Therefore, in the model, both the risk of accidents and the risk factors should be represented in a way with sufficient spatial and temporal granularity.

In order to achieve this, we implemented spatial and temporal data fusion for both the accident risk and risk factors. Here we use "variables" to refer to both indiscriminately. Each variable comes with certain spatial and temporal granularity, but they do not necessarily match with one another. For example, accidents data could include the exact geographical coordinates of the accident, traffic volume data might only indicate the condition on a specific road segment, and total population can only be represented with respect to an area such as a census block.

To effectively model these variables, we divided New York City before and after imposing the travel-restriction policies into cross

sections in both space and time. Each variable is then mapped or triangulated into these spatial and temporal cross sections. Temporally, variables were aggregated or triangulated based on the frequency the data sources are available in. There are also variables that do not vary with time. Spatially, variables that come in the forms of points, segments and areas go through different spatial joins to conform to the desired span of the chosen spatial cross section. Finally, models were built linking accidents risk to the risk factors based on the cross sections. This process is illustrated in Fig. 3. Categorization of all relevant variables is presented in Table 1.

In terms of the choice of the appropriate spatial and temporal spans of the cross sections, three main factors were considered. Firstly, the spatial and temporal spans should be small enough for the results to be useful. Secondly, the spatial and temporal spans should be large enough to screen out excessive randomness in the data. For example, our data allow us to map the occurrence of accidents onto 1-minute time intervals. But it is difficult to conceive of contributing factors that reliably predicts an accident happening in that very minute as opposed to the minute before or after it. Additionally, the spans should allow us to represent the level of risk in the spatial-temporal cross section using the series of derived Boolean variables. Lastly, we considered the granularities that all variables are available in from data sources so that the mapping process would make sense.

Eventually, we decided to divide time into one-hour intervals between Mar 2019 and Feb 2021. We also decided to divide space in the NYC using zip codes. Subsequently all variables were mapped or triangulated to the spatial-temporal cross sections similarly.

In Table 1, the distinction between time variant and time invariant variables still exists even after they have been spatially joint to zip codes. Therefore, we can divide these variables to two sets. The first set of variables, X_{tz} , is time variant, meaning for the same zip code, the variables take on different values in different time intervals. Examples of this set of variables include weather condition, COVID-19 cases, travel-restriction policies in place, and the number of approximated Citi Bike trips. The second set of variables, W_z , is time invariant, meaning for the same zip code, the variables always take on the same value. Examples of this set of variables include the length of road, the average speed on all roads, and the number of points of interest in the zip code.

To represent the risk of types of accidents in each spatial-temporal cross section, a series of hierarchical variables were generated. Because our interest lies in discovering and differentiating factors that contribute to higher risk of accidents, especially those involving casualties of cyclists, these variables correspond to nested types of accidents happening. Denoting the time interval as t , we generated three variables:

- $\mu_{tz} = 1$ if there were any road traffic accidents in the spatial-temporal cross section, and 0 otherwise.
- $\nu_{tz} = 1$ if any such accidents resulted in any casualties, and 0 otherwise.
- $\xi_{tz} = 1$ if any such accidents resulted in casualties of cyclists, and 0 otherwise.

Notice that these three variables are necessarily hierarchical, i.e., μ_{tz} must be 1 if ν_{tz} is 1 and so forth. The models were subsequently built linking these variables representing accident risks to two sets of independent variables. The hierarchical models' structure is illustrated in Fig. 4. We refer to these three levels of models as Model 1, Model 2, and Model 3.

4.2. Model implementation

In this section we discuss the specific modelling choices based on the framework. As described as the previous subsection, NYC before and after imposing COVID-19 related travel-restriction policies was divided into zip codes of one-hour intervals, between Mar 2019 and Feb 2021. All variables were then mapped or triangulated to these spatial-temporal

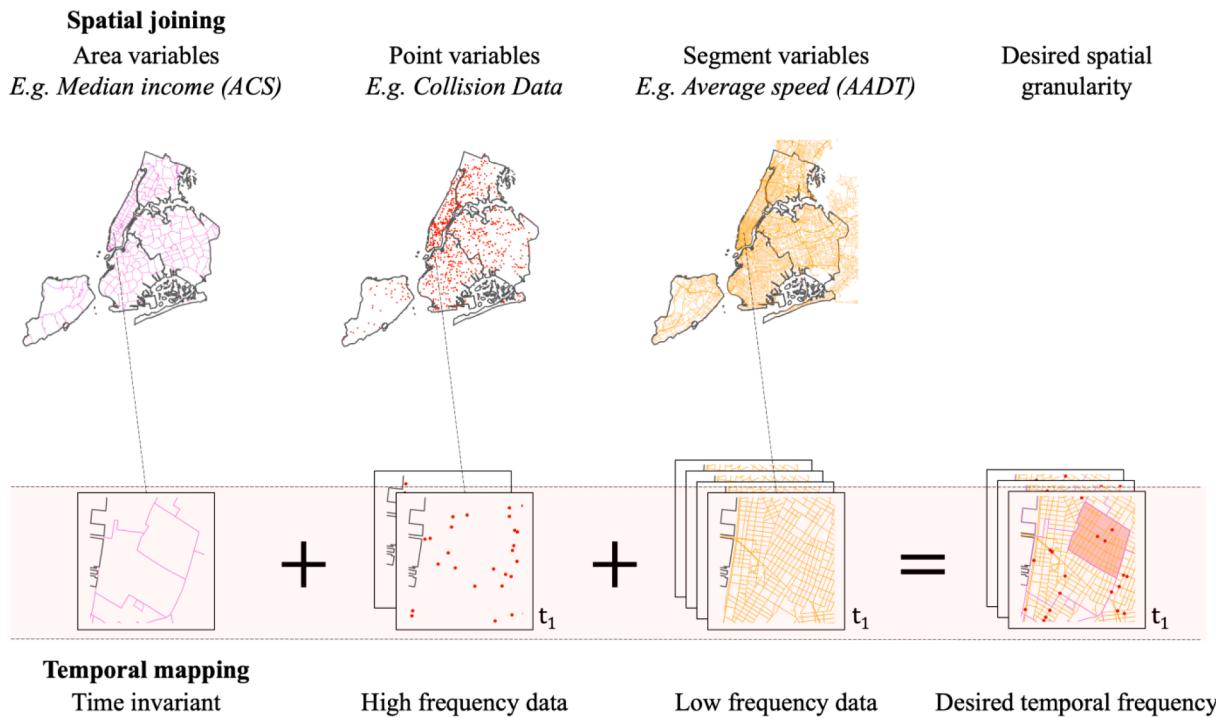


Fig. 3. Spatial and Temporal Data Fusion Process. This figure depicts the fusion process on the spatial and temporal dimensions to derive consequent spatial-temporal cross section. Spatially, variables could be available as points, segments or areas, examples of which are given in figure. They are transformed to match zip codes. Temporally, variables are available in various frequencies. The pink horizontal band represents the process of mapping them to match the desired output frequency, one hour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Variable Categorization based on Spatial and Temporal Formats.

	Point	Segment	Area
Time variant	Collisions Citi bike trips (treatment see subsection 3.2)	–	Stay at home order Reopening order New COVID case number Weather conditions
Time invariant	Points of interest	Total road length Average speed AADT Traffic flow makeup	Total population Male percentage Median income Gini index Population density

cross sections. Different ways of temporal and spatial joins were performed. The specific mapping processes were briefly described in the Data section.

Following the construction of the model dataset, two families of models were built linking accident risk in the spatial-temporal cross section with risk factors: logistic regression (logit) with fixed effects and random forest models. Much to our surprise, the random forest models did not produce significantly better performance on the out-of-sample test set compared to the logit models, measured using metrics such as area under the ROC curve. This is possibly a reflection of a lack of informative data, or of the inherent randomness in the task of predicting road traffic accidents. In the Results section, we will focus on presenting the estimation results of the fixed effects logit models due to their higher interpretability.

We expect strong temporal and spatial autocorrelations in the data. This is especially true given the likely fact that we will omit or fail to include variables that are correlated with space and time. In a simple logit model, this autocorrelation will not be accounted for. Therefore,

fixed effects terms were added in logit models to remedy this issue.

Three fixed effects logit models were built, each corresponding to one generated accident risk variable. They are formulated as follows:

$$P(\mu_{tz} = 1) = \sigma(\beta_X X_{tz} + \beta_W W_z + \beta_\eta \eta_t + \beta_\theta \theta_z) \quad (4)$$

$$P(\nu_{tz} = 1 | \mu_{tz} = 1) = \sigma(\gamma_X X_{tz} + \gamma_W W_z + \gamma_\eta \eta_t + \gamma_\theta \theta_z) \quad (5)$$

$$P(\xi_{tz} = 1 | \mu_{tz} = 1) = \sigma(\omega_X X_{tz} + \omega_W W_z + \omega_\eta \eta_t + \omega_\theta \theta_z) \quad (6)$$

where σ denotes the sigmoid function, η_t denotes a vector of temporal fixed effects variables, and θ_z denotes a vector of spatial fixed effects variables. Specifically, the temporal fixed effects variables include the hour of the day, day of the week, and month of the year of the spatial-temporal cross section. The spatial fixed effects variables include the borough or county that the zip code is in. These fixed effects terms were added to reduce the effect of any unobserved characteristics such as seasonality that would otherwise bias the estimation results. In the following sections, the three models specified by Eqs. (4), (5), and (6) will be referred to as Logit 1, Logit 2 and Logit 3, respectively.

Three random forest models were also estimated parallel to Logit 1, Logit 2, and Logit 3. Each pair share the same dependent and independent variables. These random forest models will be referred to as RF 1, RF 2, and RF 3. For each model, 1000 trees were grown, and the number of variables randomly sampled as candidates at each split was 5.

All models were evaluated using two goodness-of-fit metrics. The first one is McFadden's R squared, defined as one minus the ratio between the log-likelihoods of the model being evaluated and the null model. The second one is the area under the receiver-operator curve (ROC), i.e., AUC. For both metrics, given the same training set, a higher value indicates a better model fit.

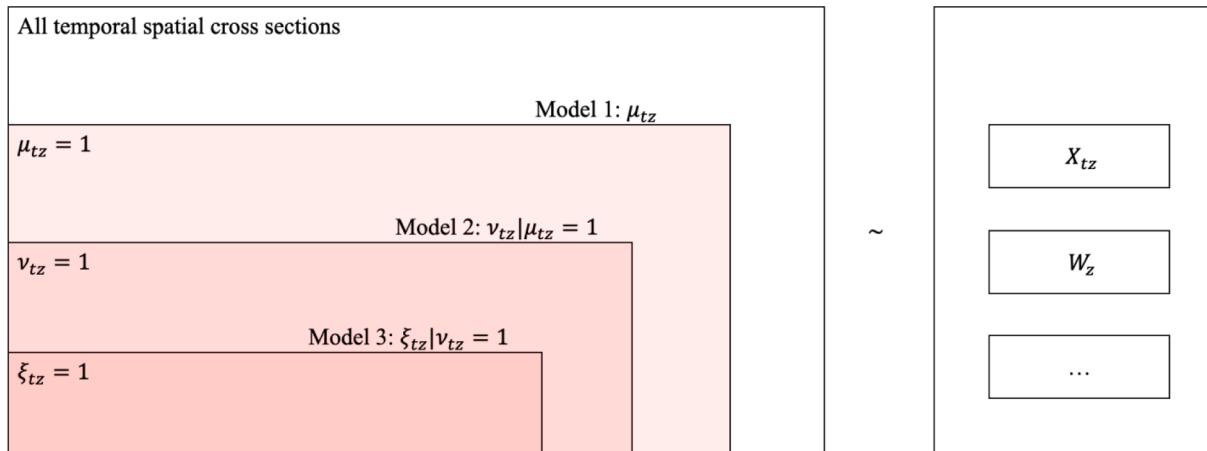


Fig. 4. Hierarchical Models Structure. This figure depicts the hierarchy of models developed in this paper. In the biggest rectangle, Model 1 describes the delineation between accidents happening or not in the spatial-temporal cross section, i.e. μ_{tz} . Inside the pale pink rectangle, where any accidents happen, Model 2 describes the delineation between severe accidents happening or not in the spatial-temporal cross section, i.e. $v_{tz} | \mu_{tz} = 1$. Finally, inside the moderately pink rectangle, Model 3 describes the delineation between any cyclists being involved in the severe accidents that happen in the spatial-temporal cross section, i.e. $\xi_{tz} | v_{tz} = 1$. X_{tz} , W_z and other variables are then used to build these models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Results

5.1. Road traffic accidents and Citi Bike usage pattern shifts

We start by examining the pattern change in road traffic accidents after imposing the stay-at-home order. As presented in the last section, there was a huge decline in the total number of collisions. In addition, the following phenomena are worth noting.

First, the spatial distribution of collisions changed profoundly. This is shown in Fig. 5. To minimize the confounding factor of seasonality, the kernel density of collisions in Aprils of 2019, 2020 and 2021 are plotted. In Apr 2019, there was a marked concentration of all collisions in Manhattan around Midtown. However, in Apr 2020, there was almost no concentration of collisions in Manhattan, but rather moderate concentrations in Bronx and part of Brooklyn. The same trend continued into Apr 2021, although with some reversion to the pattern in 2019. Indeed, the share of all collisions that happened in Manhattan dropped from 20% in Apr 2019 to 13% in Apr 2020 (see Fig. 6). This is coherent with

findings from other studies that accidents shifted spatially to outlying metropolitan areas (Lin et al., 2020; Saladié et al., 2020).

Despite a lower totally number of accidents, the severity of the collisions increased significantly. This can be seen from the average number of casualties caused by each collision. As is shown in Fig. 7, the number of average casualties per collision increased dramatically since the start of the stay-at-home order. It further increased and almost doubled until after the start of the reopening phases. This number of average casualties remained at an elevated level up until the time of writing, although there was a temporary dip in early 2021. This finding is consistent with previous studies that severe accidents did not decrease (Brodeur et al., 2021; Qureshi et al., 2020). However, this study finds the severity increase was still net positive even after reopening.

Most notably, this increase in severity of collisions was the most jarring in accidents involving cyclists. The usage, and therefore casualty pattern of cycling has very strong seasonality. But Fig. 8 shows that, compared with previous years, the average number of cyclist casualties per collision increased significantly after the stay-at-home order and

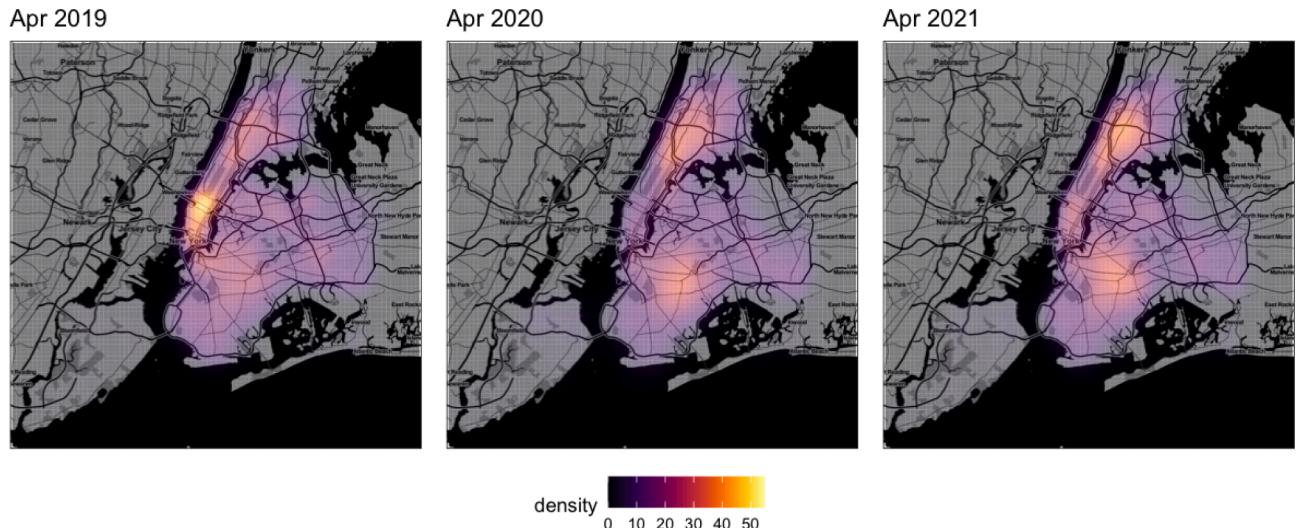


Fig. 5. Kernel Density Plot of Collisions of Various Months in NYC. All collisions during the months of Aprils in 2019, 2020 and 2021 are aggregated and their kernel density plotted in the figure. These plots share one colour code scale, where a bright yellow colour indicates a higher density of collisions, and a dark purple one, a lower. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

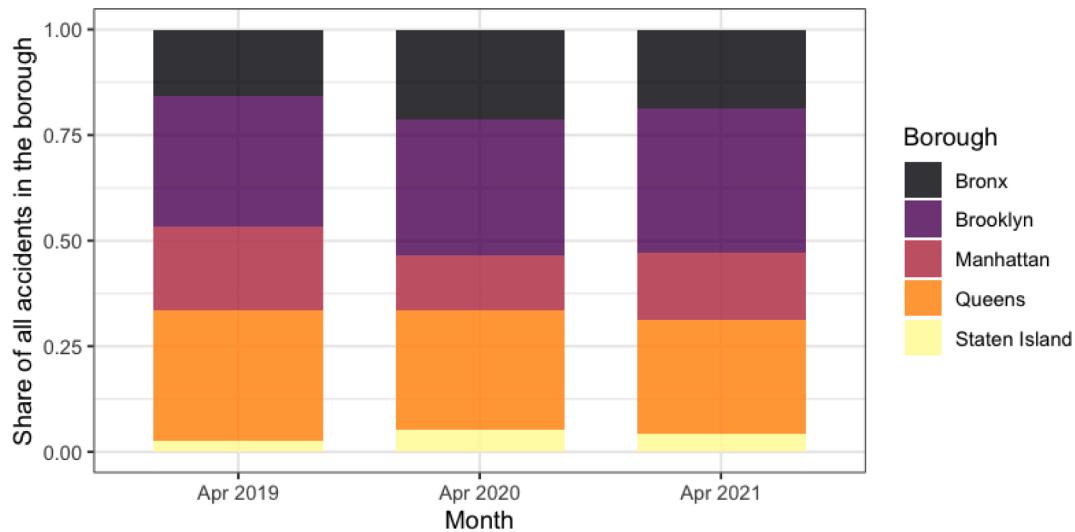


Fig. 6. Share of All Collisions in Five Boroughs in Various Months in NYC. All collisions during the months of Aprils in 2019, 2020 and 2021 are spatially joint to the five boroughs and aggregated by the borough they are in. The percentages of collisions in the boroughs are then plotted in each bar.

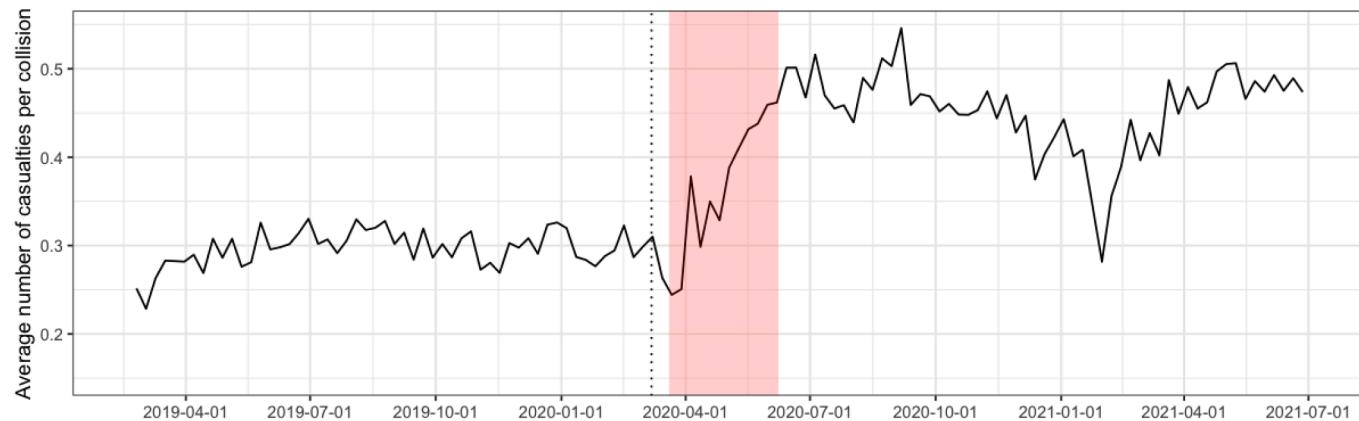


Fig. 7. Weekly Average Number of Casualties per Collision Before and After Stay-At-Home Order in NY. The X-axis indicates the date, and the Y-axis indicates the weekly average number of casualties per collision of the following week, starting on Sundays. The horizontal dotted line indicates the declaration of a state of emergency in NY (Mar 8, 2020). The red area indicates the period between the start of the stay-at-home order, Mar 20, 2020, and the start of reopening, Jun 8, 2020. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

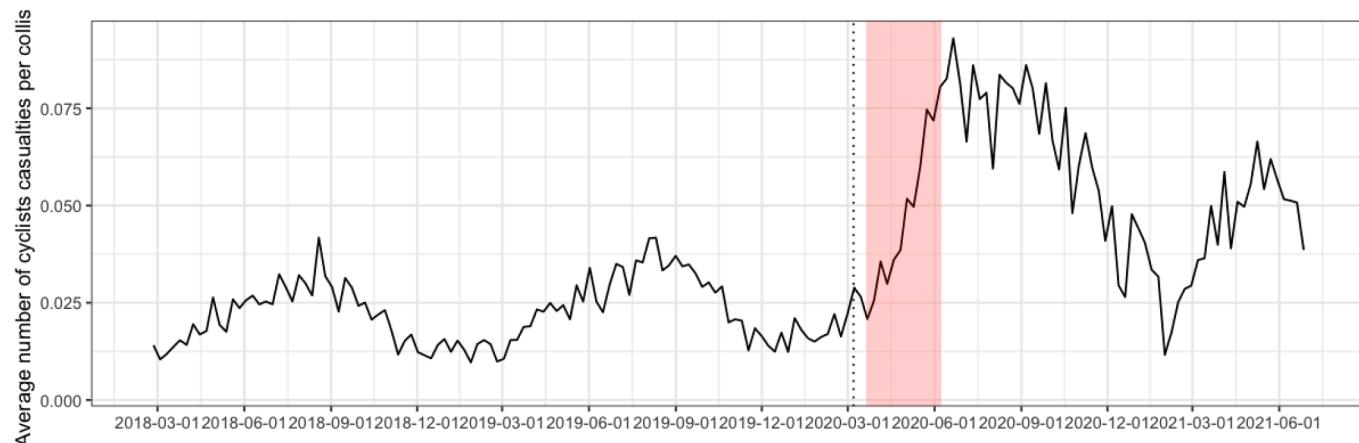


Fig. 8. Weekly Average Number of Casualties Among Cyclists per Collision Before and After Stay-At-Home Order in NY. The X-axis indicates the date, and the Y-axis indicates the weekly average number of casualties among cyclists per collision of the following week, starting on Sundays. The horizontal dotted line indicates the declaration of a state of emergency in NY (Mar 8, 2020). The red area indicates the period between the start of the stay-at-home order, Mar 20, 2020, and the start of reopening, Jun 8, 2020. The range on X-axis is chosen to demonstrate the yearly seasonality of the metric. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

more than tripled at its peak. At the time of writing, this number is still much higher than levels seen in previous years. In addition, the total number of casualties experienced by cyclists in 2020 was higher than in previous years. On the one hand, this increase is likely associated with the rise in the mode share of cycling during the COVID-19 pandemic. On the other hand, this also suggests accidents involving cyclists warrants our attention.

Now we compare collisions patterns with the Citi Bike usage patterns. The number of Citi Bike trips experienced an acute dip after the state of emergency declaration in NY. It then slowly recovered but remained lower than levels at similar times in 2019 after the stay-at-home order started. The number of trips finally recovered fully in the summer. But this level is about the same as 2019 (see Fig. 9). Admittedly, the trip volume of Citi Bike is not a perfect proxy for that of all cycling traffic. But it is very interesting that, despite a higher total number of casualties among cyclists in 2020 relative to previous years, overall bicycle traffic volume might not have increased. This would suggest that riding a bicycle had become more dangerous in the NYC after the declaration of the stay-at-home order.

5.2. Logit model estimation results

In total, 703 days of data were used in the models. Because the data were divided into 1-hour cross sections for each zip code, a total of 4,437,599 data points were used. The definitions of these variables are listed in Table 2, and the summary statistics of these variables are listed in Table 3.

Due to the relatively large number of observations, many variables showed up statistically significant by metrics such as the t-statistic. At the same time, the highly random nature of road traffic accidents means the model fit was not very high. The estimation results are presented in Table 4.

In Logit 1, it is immediately clear that the New York State stay-at-home order taking effect is highly correlated with a much lower number of road traffic accidents occurring. On the contrary, the reopening taking effect is highly correlated with an elevated frequency of road traffic accidents. At the same time, the logarithm of the 7-day moving average of new COVID-19 cases in NYC is also negatively correlated with road traffic accidents risk. These findings suggest that it is likely that people tend to travel less when local COVID-19 cases are on the rise, as well as when there are state actions in place, thus reducing the risk of road traffic accidents happening in general.

Additionally, in Logit 1, a higher average speed in the zip code is correlated with a lower overall risk of road traffic accident occurring. This is consistent with the finding that usually, overall accidents

Table 2
Descriptions of Variables Used in Models.

Variables	Description
(Time variant variables)	
stay at home	True if date is on or after Mar 21, 2020.
reopening	True if date is on or after Jun 9, 2020.
sma7 covid	Logarithm of 7-day simple moving average of daily new COVID-19 cases in New York City
approx. citi trips	Approximated number of cycling trips whose trajectories intersect with the zip code, as calculated according to Eq. (3)
weather very hot	True if the highest temperature is higher than 86F on the day
weather very cold	True if the lowest temperature is lower than 30F on the day
weather precipitation	True if precipitation is strictly positive on the day
weather snow	True if snow is strictly positive on the day
weather thunder	True if there is thunder on the day
weather windy	True if the highest 2-min gust is higher than 20 miles per hour on the day
(Time invariant variables)	
total population	Total population in the zip code
male	Percentage of residents that are male in the zip code
median income	Median annual income in the zip code in US dollars
gini index	Gini index in the zip code
n poi	Number of points of interest in the zip code
length road	Total length of road in the zip code, in feet
avg speed	Average speed on all roads in the zip code, in miles per hour
sum aadt	Sum of annual average daily traffic (AADT) on all roads in the zip code
avg truck pct	Average percentage of AADT that is truck traffic over all roads in the zip code
avg car pct	Average percentage of AADT that is car traffic over all roads in the zip code
pop density	Total population divided by area of the zip code, in persons per square foot
avg speed/avg speed limit	Average speed on all roads in the zip code, divided by average speed limit on all roads in the zip code
length road/AREA	Total length of roads in the zip code, divided by the area of the zip code, in feet ⁻¹

frequency is higher in high density areas such as Manhattan. A lower average speed could indicate more frequent traffic congestions, and therefore a higher likelihood of low severity accidents.

One interesting finding in Logit 2 is that, given that an accident happens in the zip code cross section, a higher average speed as well as a lower total AADT are correlated with a higher likelihood that the accident results in a casualty. This is intuitive and consistent with literature. Accidents that happen in low-density traffic usually involve higher collision speed and therefore, usually result in more casualties per accident. Combined with variables such as total population in the zip code, sociodemographic characteristics, etc., the results indicate lower density



Fig. 9. Weekly Number of Citi Bike Trips Before and After Stay-At-Home Order in NY. The X-axis indicates the date, and the Y-axis indicates the weekly total number of Citi Bike trips of the following week, starting on Sundays. The horizontal dotted line indicates the declaration of a state of emergency in NY (Mar 8, 2020). The red area indicates the period between the start of the stay-at-home order, Mar 20, 2020, and the start of reopening, Jun 8, 2020. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Summary Statistics of Variables Used in Fixed Effects Models.

Variables	Count	Mean	Median	Min	Max
<i>(Time variant variables)</i>					
μ_{tz}	4,437,599	4.96E-03	0	0	1
ν_{tz}	4,437,599	3.54E-04	0	0	1
ξ_{tz}	4,437,599	1.19E-05	0	0	1
stay at home	4,437,599	4.52E-01	0	0	1
reopening	4,437,599	3.39E-01	0	0	1
sma7 covid	4,437,599	7.49E+02	0	0	5.29E+03
approx. citi trips	4,437,599	2.71E+02	4.32E+01	0	5.06E+03
weather very hot	4,437,599	1.22E-01	0	0	1
weather very cold	4,437,599	9.82E-02	0	0	1
weather precipitation	4,437,599	3.73E-01	0	0	1
weather snow	4,437,599	2.00E-02	0	0	1
weather thunder	4,437,599	8.25E-02	0	0	1
weather windy	4,437,599	8.11E-02	0	0	1
<i>(Time invariant variables)</i>					
total population	263	3.33E+04	2.92E+04	0	1.12E+05
male	263	4.79E-01	4.78E-01	3.45E-01	9.49E-01
median income	263	7.32E+04	6.65E+04	2.11E+04	2.50E+05
gini index	263	4.79E-01	4.76E-01	3.63E-01	6.45E-01
n poi	263	7.40E+01	5.00E+01	0	3.73E+02
length road	263	9.89E+04	9.78E+04	0	5.37E+05
avg speed	263	2.13E+01	2.10E+01	1.36E+01	3.94E+01
sum aadt	263	1.20E+06	7.96E+05	0	9.00E+06
avg truck pct	263	6.64E+00	6.56E+00	0	1.57E+01
avg car pct	263	8.64E+01	8.65E+01	7.13E+01	9.90E+01
pop density	263	2.15E-03	9.61E-04	0	7.53E-02
avg speed/avg speed limit	263	1.10E+00	9.97E+01	5.24E-01	5.12E+00
length road/area	263	4.64E-03	3.04E-03	0	1.05E-01

areas outside the city centre have a higher risk of severe accidents.

From the estimation results with respect to time-variant variables in Logit 2, it is remarkable that while the stay-at-home order has a negative association with accidents happening overall, it is also positively associated with such accidents resulting in casualties. Notice this variable is highly significant even though the model includes variables representing road traffic speed and volume such as average speed and total AADT. Admittedly, these variables might not capture the full and real-time effect of COVID-19 and travel-restriction policies on road traffic. Nonetheless, this result suggests the stay-at-home order might have contributed to higher accident severity, beyond the effect of traffic speed and volume. Possible explanations include the reports that consumption of alcohol, reckless driving and drag races on public roads picked up in certain areas in the US after the COVID-19 travel restriction policies took place (Pollard et al., 2020).

In Logit 3, we further focus on whether the spatiotemporal cross

sections where severe accidents happen have cyclists' casualties. Among the time invariant variables, a higher total population, a higher population density and a lower average speed are all correlated with a higher likelihood that a severe accident involves cyclists. This makes sense, as bike traffic volume tends to be higher in dense areas, which would correspond to higher risk of cyclist accidents.

Among the time variant variables, the variable with by far the largest effect size is the logarithm of the approximated number of Citi Bike trips. This is very intuitive and highly advantageous from a policy-making point of view. Both spatially and temporally, roads with a high bike traffic volume are likely to be hot zones for accidents involving cyclists, and therefore, warrant more attention from city managers to protect their safety. Furthermore, we can predict which areas have high cyclists traffic volume from well recorded commercial services such as Citi Bike.

At the same time, imposing the stay-at-home order does not significantly correlate with change in the likelihood of severe accidents involving cyclists in either direction. However, the logarithm of new COVID-19 case numbers is still positively associated with the occurrence of severe accidents involving cyclists. This could indicate that, as COVID-19 cases increase, people are driven to use cycling more to take advantage of the low transmissibility in open-air environments. However, it is difficult to find other evidence supporting that at the time of writing.

In addition to the logit models, three random forest models were also attempted parallel to Logit 1, Logit 2, and Logit 3. The model performance was not better than the logit models. The out-of-sample area under ROC curve metric comparison is presented in Table 5. Despite the slightly inferior performance, the random forest models can provide insight into the likely complex relationship between continuous variables and the risk of different types of accidents.

The random forest models branch the dataset using cut-offs along continuous variables. Therefore, they can be used to interpret the nonlinear effects of these continuous variables. In this study, we are particularly interested in the effects of number of new COVID-19 cases, the approximated number of Citi Bike trips and the average speed in the spatial-temporal cross section. To illustrate the relationship between the risk of accidents and values of these continuous variables from the random forest models, we sampled 1000 spatial-temporal cross sections from the test set, assigned each different values in each continuous variable, and simulated the probability of an accident happening in this cross section using the random forest models. These partial dependence plots with respect to the variables is presented in Fig. 10. The lower and upper limits of the pink bands indicate the 0.1 and 0.9 quantiles of the predicted probabilities.

The predicted probability of accidents changes only slightly with respect to the 7-day moving average of new COVID-19 cases. The directions of change generally agree with estimation results of the logit models, except for a small uptick in predicted probability around low COVID-19 case numbers from RF 2. These findings suggest that the impact of new COVID-19 cases on change in accident patterns might be limited beyond the travel-restriction policies.

By contrast, the predicted probability of severe accidents involving cyclists increases drastically with respect to a higher number of approximated Citi Bike trips in the zip code. This finding makes intuitive sense and is coherent with estimation results of Logit 3. The increase is especially fast when the approximated number of trips is low, and significantly slows down after the approximated number of trips exceeds 1000. The random forest models provide useful granular insight beyond the logit models, while in general agree with the logit estimation results.

Finally, the relationship shown between the average speed and predicted probability of accidents generally agrees with the estimation results of the logit models. A higher average speed is correlated with a lower probability of any accidents, but a slightly higher probability of severe accidents given one already happens, although such trends are rather weak. However, the random forest model predicts that probability of any severe accidents involving cyclists is the highest when

Table 4
Accident Risk Fixed Effects Logit Models Estimation Results.

Variables	Logit 1			Logit 2			Logit 3		
	Estimate	z value	p value	Estimate	z value	p value	Estimate	z value	p value
Intercept	-3.197	-177.729	0.000	-1.171	-30.017	0.000	-2.180	-20.490	0.000
(Time invariant variables)									
total population	0.357	89.153	0.000	-0.004	-0.452	0.651	0.174	7.627	0.000
male	-0.034	-11.928	0.000	-0.010	-1.666	0.096	0.124	8.842	0.000
median income	0.080	24.863	0.000	-0.071	-9.270	0.000	-0.037	-2.086	0.037
gini index	0.184	54.433	0.000	-0.045	-5.946	0.000	0.049	2.621	0.009
n poi	0.003	62.714	0.000	0.000	2.939	0.003	-0.001	-3.833	0.000
length road	0.459	97.320	0.000	0.077	7.432	0.000	-0.092	-3.446	0.001
avg speed	-0.092	-30.480	0.000	0.023	3.509	0.000	-0.187	-10.910	0.000
sum aadt	-0.055	-17.341	0.000	-0.047	-7.062	0.000	0.002	0.137	0.891
avg truck pct	0.099	27.754	0.000	-0.010	-1.208	0.227	-0.087	-4.127	0.000
avg car pct	0.023	6.462	0.000	0.040	4.809	0.000	-0.089	-4.305	0.000
pop density	-0.260	-27.979	0.000	0.096	2.326	0.020	0.402	4.407	0.000
avg speed/avg speed limit	-0.002	-3.531	0.000	0.001	0.831	0.406	0.013	4.433	0.000
length road/area	-0.017	-39.027	0.000	-0.003	-3.002	0.003	0.005	1.961	0.050
(Time variant variables)									
stay at home	-0.877	-38.371	0.000	0.437	8.466	0.000	0.108	0.807	0.420
reopening	0.531	44.375	0.000	0.253	9.905	0.000	-0.005	-0.075	0.940
log(7-d sma covid)	-0.053	-19.211	0.000	-0.030	-4.763	0.000	0.064	3.754	0.000
log(approx. citi trips)	-0.074	-14.696	0.000	0.044	3.848	0.000	0.586	21.236	0.000
weather very hot	0.039	4.565	0.000	0.052	2.942	0.003	0.002	0.057	0.954
weather very cold	0.006	0.734	0.463	-0.109	-5.852	0.000	-0.063	-1.137	0.255
weather precipitation	0.021	4.108	0.000	0.011	1.038	0.299	-0.089	-3.155	0.002
weather snow	0.106	6.650	0.000	-0.037	-1.069	0.285	-0.287	-2.331	0.020
weather thunder	-0.016	-1.777	0.076	-0.022	-1.184	0.237	0.086	1.918	0.055
weather windy	0.005	0.529	0.597	-0.045	-2.390	0.017	-0.028	-0.561	0.575

Table 5
Area Under ROC Curve Comparison between Logit Models and Random Forest Models.

	Model 1	Model 2	Model 3
Logit AUC	0.816	0.597	0.720
Random forest AUC	0.784	0.556	0.683

average speed is lower than 20 mph. This probability climbs slightly again when average speed is higher than 25 mph. This leaves the segment between 20 and 25 mph to be the average speed associated with the lowest probability of any severe accidents involving cyclists. This finding can further help city managers to detect high risk areas for cyclists.

5.3. Severe accidents hot zone detection

Following the estimation of the models, we further explored the application of the models to detect hot zones for accidents. In this application we focus on the shift of hot zones before and after COVID-19 related travel restrictions were imposed. For the comparison to be of practical value, we decided to only compare risks of severe accidents that result in casualties, and especially those involving cyclists. Specifically, we applied the models to different hours on two specific days before and after travel restrictions to find areas of high risk for severe accidents. This was done separately for accidents that result in any casualties and cyclists' casualties.

We chose January 8, 2020, and January 13, 2021, two Wednesdays, to calculate the severe accidents risks. The former date represents a typical weekday respectively before COVID-19 infections became an issue in NYC, and the latter represents one during a period of relatively heightened COVID-19 cases and some travel restrictions in effect. Eight hourly intervals were chosen from 8 AM to 10 PM, and the probability that a severe accident happens in each zip code area was calculated and plotted on a map. These plots are shown in Fig. 11 and Fig. 12.

Both plots suggest that severe accidents hot zones are mostly located outside Manhattan. This is consistent with our findings in the logit models. In addition, these hot zones have strong temporal concentration patterns. Most hot zones have heightened predicted risks in the afternoon, and especially around 4 PM and 6 PM.

Comparing the two dates before and after COVID-19 related travel restrictions were imposed, it is evident that the change in hot zones for severe accidents involving cyclists was much more pronounced. While such hot zones did not shift much geographically or temporally, the predicted risk substantially increased while COVID-19 travel restrictions were in place. By contrast, the risk for severe accidents in general had a muted decrease since the COVID-19 related travel restrictions were imposed. This change should be understood as the combined effect of decreased risk for accidents in general and increased risk for severe accidents conditional on accidents happen at all.

Comparing the two plots, it is notable that the spatial distribution of cyclists' severe accidents hot zones is much closer to the city centre, with high concentrations in parts of Brooklyn, especially Williamsburg, and Queens. There is also a hot spot for cyclists' severe accidents in Lower East Side during the afternoon rush hour. By contrast, hot zones for general severe accidents are much more spread out, with a high-risk area in Staten Island. In addition, the cyclists' severe accidents hot zones are also more temporally concentrated than overall severe accidents hot zones.

6. Conclusion

COVID-19 has brought about dramatic societal shifts. Among these, traffic patterns and road traffic accidents changes are particularly profound and consequential. In this paper, we reviewed the road traffic accident pattern in the NYC before and after imposing the state stay-at-home order and built fixed effects logit models to assess factors that contribute to the occurrence of severe accidents, especially those involving cyclists. Finally, we applied these models to detect severe accidents hot zones. This paper contributes to the limited literature linking travel-restriction policies to patterns of road traffic accidents

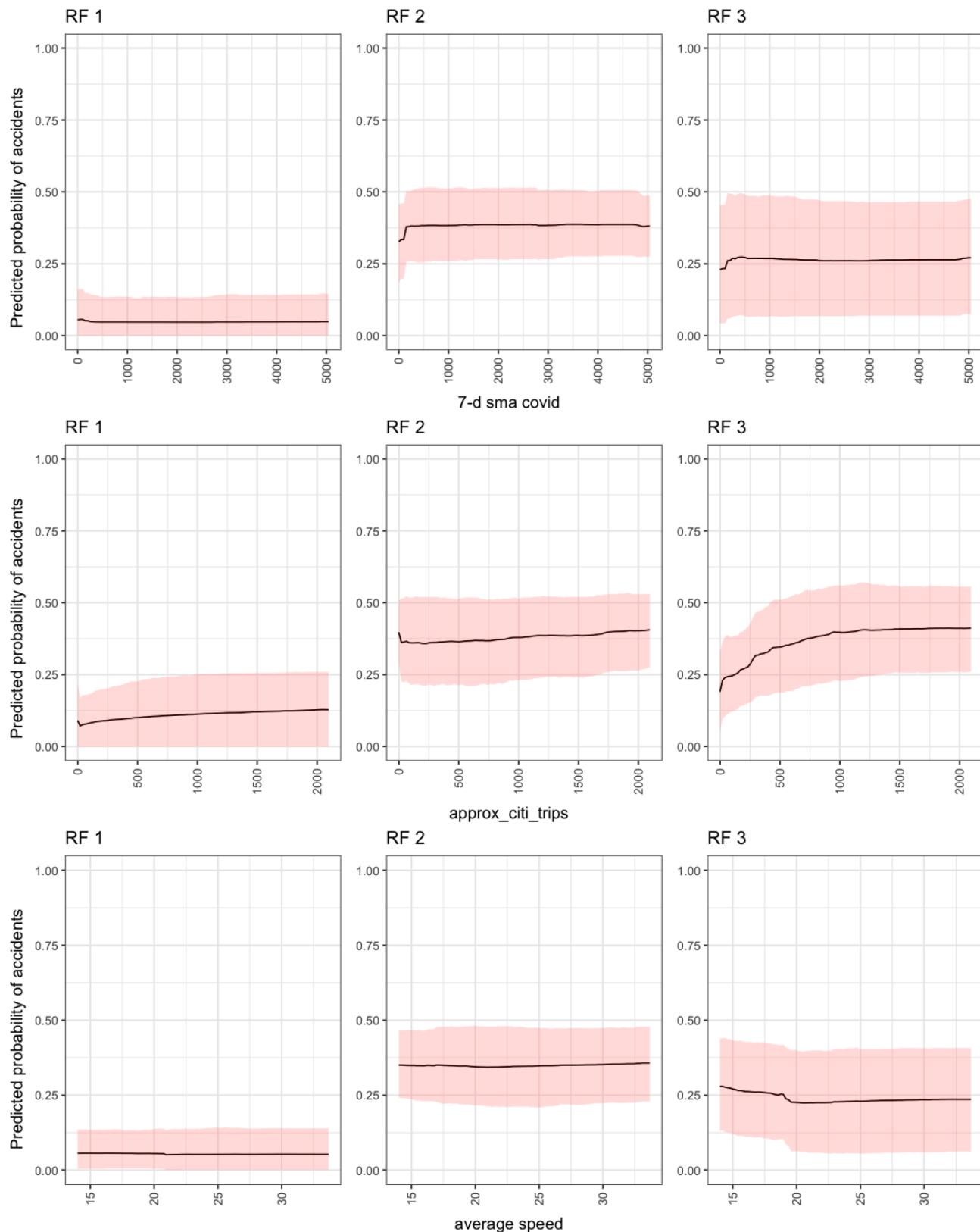


Fig. 10. Partial Dependence Plots With Respect to 7-Day Moving Average of New COVID-19 Cases, Approximated Number of Citi Bike Trips and Average Road Speed in the Zip Code. Each row comprises three partial dependence plots corresponding to RF1, RF2 and RF3 with respect to one variable. The three variables are 7-day moving average of new COVID-19 cases, approximated number of Citi Bike trips and the average road speed in the zip code. In each plot, X-axis indicates the level of the variable of interest. The range is between the 0.01 and 0.99 quantiles of the variable in the dataset. Y-axis indicates the predicted probability of different accidents. The solid line indicates the mean value of the predicted probabilities, and the pink ribbon indicates the 0.1 and 0.9 quantiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

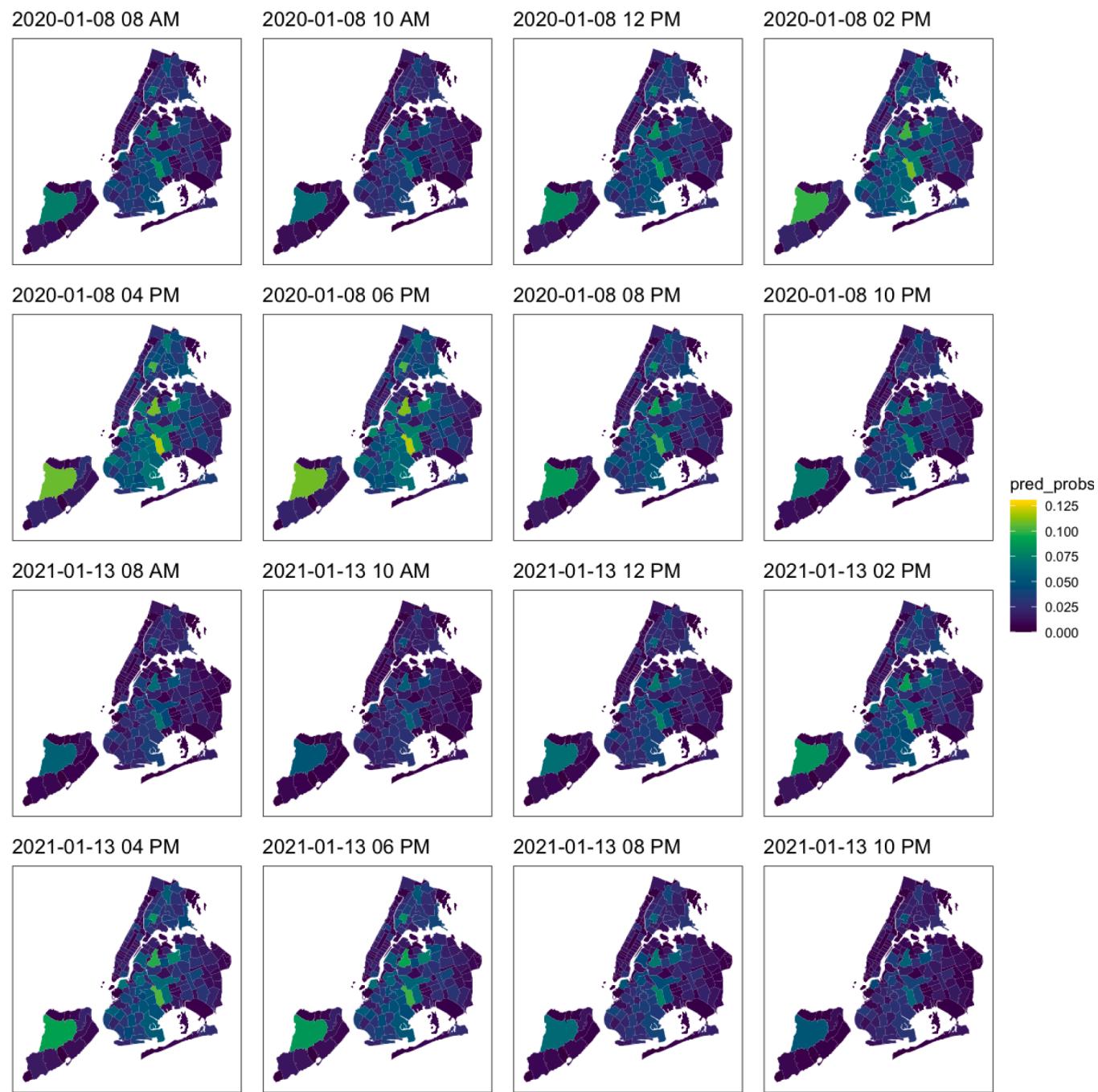


Fig. 11. Maps of Predicted Risk of Accidents Resulting in Any Casualties in Zip Codes on 1/8/2020 and 1/13/2021. The predicted risk of accidents resulting in any casualties every 2 h from 8 AM to 10 PM on 1/8/2020 and 1/13/2021 is plotted. The colour coded spatial units are zip codes. All plots share the same colour code scale, where a bright yellow colour indicates a higher risk and a dark purple one, a lower risk. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

involving cyclists.

We found a huge drop in the total number of accidents in the NYC, but with a sharp increase in the average number of casualties caused by each accident. In particular, the average number of cyclists killed or injured per accident more than tripled relative to similar times in previous years. The total number of cyclist casualties was also higher than previous years, standing out among most accidents counts that saw a decrease. The spatial concentration of accidents inside Manhattan also disappeared.

Using Citi Bike trip volume as a proxy for the overall cycling volume in the NYC, we did not observe an increase in the total number of trips, but rather a temporary dip during the early weeks of the stay-at-home

order. Citi Bike usage later recovered to levels seen in previous years. These observations suggest the NYC possibly became more dangerous for cyclists during COVID-19 lockdown, which goes contrary to Vision Zero policy priorities and is potentially a major unintended negative outcome of travel-restriction policies.

Three accidents risk fixed effects logit models were built and estimated based on public datasets from NYC. We found that while imposing the stay-at-home order in NYC did not contribute to more road traffic accidents happening, it is highly significantly correlated with accidents being more likely to result in casualties. This is antithetical to Vision Zero and related policy priorities. Therefore, the unintended effect of travel-restriction policies in resulting in more severe accidents should be

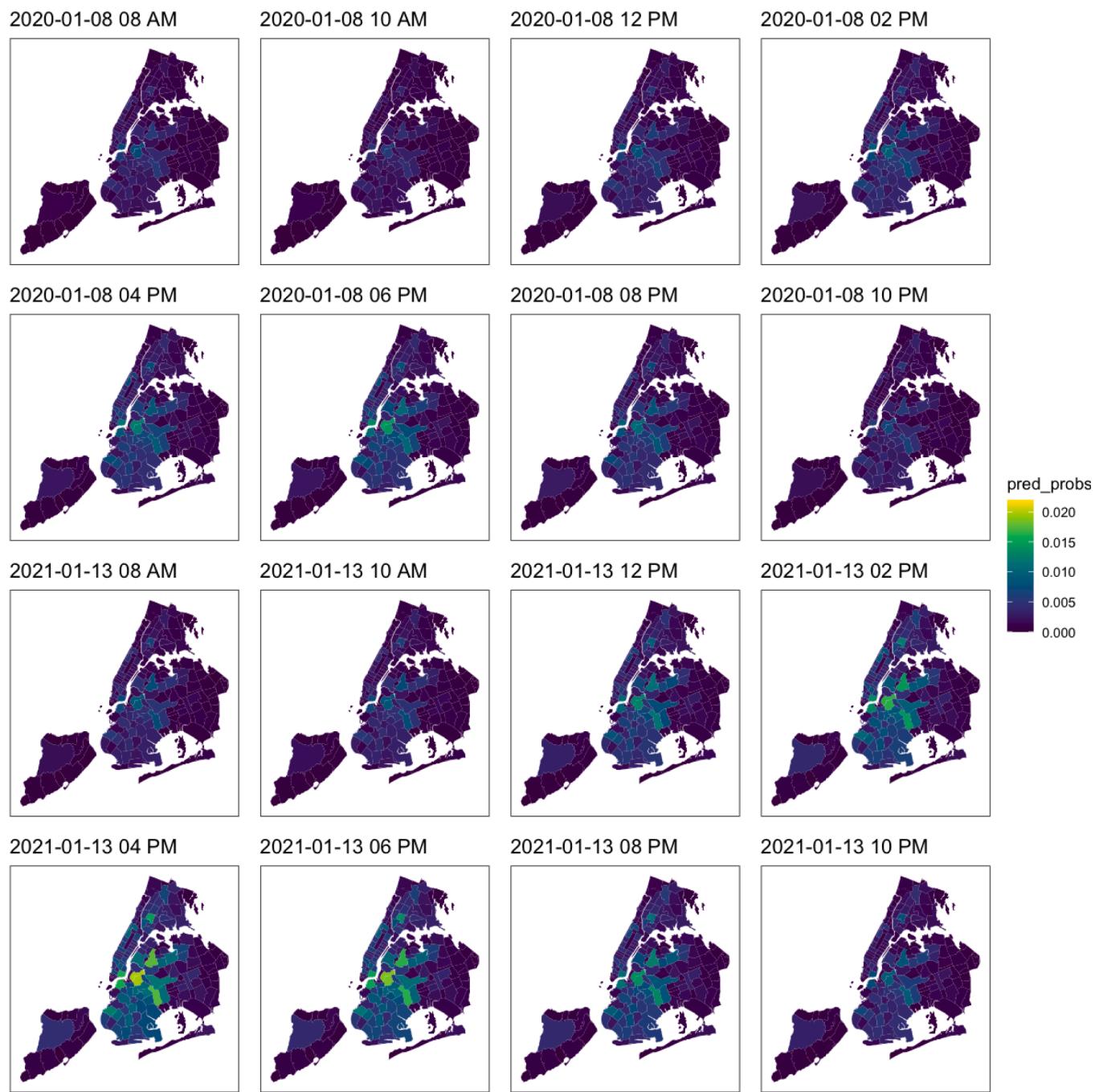


Fig. 12. Maps of Predicted Risk of Accidents Resulting in Cyclists' Casualties in Zip Codes on 1/8/2020 and 1/13/2021. The predicted risk of accidents resulting in cyclists' casualties every 2 h from 8 AM to 10 PM on 1/8/2020 and 1/13/2021 is plotted. The colour coded spatial units are zip codes. All plots share the same colour code scale, where a bright yellow colour indicates a higher risk and a dark purple one, a lower risk. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

taken into account in the future. Should future COVID-19 spread warrant further travel-restriction actions for public health purposes, holistic measures should be taken to minimize their effect on severe accidents.

Among the factors that contribute to the occurrence of an accident resulting in casualties, we found that, apart from characteristics that correspond to a relatively low-density and high-speed environment, the stay-at-home order was highly significant. Among the factors associated with severe accidents involving cyclists, the approximated number of Citi Bike trips had an oversized effect. This greatly helps city managers to identify hot zones for severe accidents involving cyclists. In addition, the number of new COVID-19 cases was found to be correlated with the occurrence of cyclists' accidents, even controlling for travel-restriction

policies.

Lastly, we demonstrated how the models could be used to identify hot zones for severe accidents. We found that hot zones for severe cyclist accidents tend to be spatially and temporally concentrated. They usually appear on the peripheries of Manhattan in the afternoon rush hour. These findings suggest policy makers and city managers should pay more attention to cyclists' safety and take holistic measures in highly concentrated hot zones, in future situations where more travel restriction policies are being considered to curb the spread of transmissible diseases.

There are certain limitations to this study. Firstly, each instance of travel-restriction policy has its own complex set of idiosyncrasies. It is

not clear if the findings in this paper with respect to the Mar 2020 NY stay-at-home order would extrapolate to other jurisdictions or even future travel restrictions in NY. From that perspective, the sample size of travel-restriction policies is small, and we should view such results with a grain of salt. Secondly, due to the inherent random nature of the occurrence of accidents, certain model fits in this paper are relatively poor. This could prevent us from making nuanced interpretations of the results. Thirdly, we were not able to find good quality data with high temporal granularity for some important variables, such as road traffic speed and volume. This has limited our ability to suggest measures that could help to reduce the risk of severe accidents in identified hot zones. We hope to address these limitations in future studies.

CRediT authorship contribution statement

Jintai Li: Conceptualization, Software, Data curation, Writing – original draft, Visualization. **Zhan Zhao:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgements

This research is partly supported by the Seed Fund for Basic Research for New Staff (104006019) at the University of Hong Kong. We thank the anonymous reviewers for their helpful comments.

References

- Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. *Accid. Anal. Prev.* 32 (5), 633–642.
- Alavi, S.S., Mohammadi, M.R., Souri, H., Kalhor, S.M., Jannatifard, F., Sepahbodi, G., 2017. Personality, driving behavior and mental disorders factors as predictors of road traffic accidents based on logistic regression. *Iran. J. Med. Sci.* 42 (1), 24.
- Almagro, M., Orane-Hutchinson, A., 2020. JUE Insight: The determinants of the differential exposure to COVID-19 in New York city and their evolution over time. *J. Urban Econ.* 103293. <https://doi.org/10.1016/j.jue.2020.103293>.
- Aloia, A., Alonso, B., Benavente, J., Cordera, R., Echániz, E., González, F., Ladisa, C., Lezama-Romanelli, R., López-Parra, Á., Mazzei, V., Perrucci, L., Prieto-Quintana, D., Rodríguez, A., Sañudo, R., 2020. Effects of the COVID-19 lockdown on urban mobility: empirical evidence from the city of Santander (Spain). *Sustainability* 12 (9), 3870. <https://doi.org/10.3390/su12093870>.
- Authority, W.M.A.T., 2020. Metro and Covid-19: Steps we've taken. Retrieved on 7 (20), 2020.
- Barbarossa, L., 2020. The post pandemic city: challenges and opportunities for a non-motorized urban environment. An overview of Italian cases. *Sustainability* 12 (17), 7172.
- Brodeur, A., Cook, N., Wright, T., 2021. On the effects of COVID-19 safer-at-home policies on social distancing, car crashes and pollution. *J. Environ. Econ. Manage.* 106, 102427. <https://doi.org/10.1016/j.jeem.2021.102427>.
- Bureau, A.T. (2020). Effects of Novel Coronavirus (COVID-19) on Civil Aviation: Economic Impact Analysis. *International Civil Aviation Organization (ICAO), Montréal, Canada.*
- Bureau, U.S.C. (2017). 2017 American Community Survey. In.
- Calderon-Anyoza, R.J.C., Kaufman, J.S., 2021. Impact of COVID-19 lockdown policy on homicide, suicide, and motor vehicle deaths in Peru. *Prev. Med.* 143, 106331. <https://doi.org/10.1016/j.ypmed.2020.106331>.
- Casado-Sanz, N., Guirao, B., Attard, M., 2020. Analysis of the risk factors affecting the severity of traffic accidents on Spanish crosstown roads: The driver's perspective. *Sustainability* 12 (6), 2237.
- Chen, H., Cao, L., Logan, D.B., 2012. Analysis of risk factors affecting the severity of intersection crashes by logistic regression. *Traffic Inj. Prev.* 13 (3), 300–307.
- Cochrane, R., 1975. A possible economic basis for the gravity model. *J. Transp. Econ. Policy* 34–49.
- Ding, H., Zhang, J., 2021. Dynamic associations between temporal behavior changes caused by the COVID-19 pandemic and subjective assessments of policymaking: A case study in Japan. *Transp. Policy* 110, 58–70.
- Dumbaugh, E., Rae, R., & Wunneberger, D. (2009). Examining the relationship between community design and crash incidence.
- Francescani, C. (2020). Timeline: The First 100 Days of New York Gov. Andrew Cuomo's COVID-19 Response. *ABC News, June, 17.*
- Golob, T.F., Recker, W.W., 2003. Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. *J. Transp. Eng.* 129 (4), 342–353.
- Iacono, M., Krizek, K., & El-Geneidy, A. M. (2008). Access to destinations: How close is close enough? Estimating accurate distance decay functions for multiple modes and different purposes.
- Iacono, M., Krizek, K.J., El-Geneidy, A., 2010. Measuring non-motorized accessibility: issues, alternatives, and execution. *J. Transp. Geogr.* 18 (1), 133–140.
- Katrakazas, C., Michalarakis, E., Sekadakis, M., Yannis, G., 2020. A descriptive analysis of the effect of the COVID-19 pandemic on driving behavior and road safety. *Transp. Res. Interdiscip. Perspect.* 7, 100186. <https://doi.org/10.1016/j.trip.2020.100186>.
- Kim, K., Brunner, I.M., Yamashita, E.Y., 2006. Influence of land use, population, employment, and economic activity on accidents. *Transp. Res. Rec.* 1953 (1), 56–64.
- Lin, L., Shi, F., & Li, W. (2020). Assessing Road Traffic Safety During COVID-19: Inequality, Irregularity, and Severity.
- Marshall, W.E., Garrick, N.W., 2011. Does street network design affect traffic safety? *Accid. Anal. Prev.* 43 (3), 769–781.
- McKinley, J., & Sandoval, E. (2020). Coronavirus in NY: Cuomo declares state of emergency. *New York Times, March, 7, A14.*
- Millefiori, L.M., Braca, P., Zissis, D., Spiliopoulos, G., Marano, S., Willett, P. K., & Carniel, S. (2020). COVID-19 impact on global maritime mobility. *arXiv preprint arXiv:2009.06960*.
- Milton, J., Mannerling, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation* 25 (4), 395–413.
- Molloj, J., Schatzmann, T., Schoeman, B., Tchervenkov, C., Hintermann, B., Axhausen, K.W., 2021. Observed impacts of the Covid-19 first wave on travel behaviour in Switzerland based on a large GPS panel. *Transp. Policy* 104, 43–51.
- NYCDOT. (2017). Citywide Mobility Survey. In.
- Papadimitriou, E., Filtness, A., Theofilatos, A., Ziakopoulos, A., Quigley, C., Yannis, G., 2019. Review and ranking of crash risk factors related to the road infrastructure. *Accid. Anal. Prev.* 125, 85–97.
- Pollard, M.S., Tucker, J.S., Green, H.D., 2020. Changes in adult alcohol use and consequences during the COVID-19 pandemic in the US. *JAMA Network Open* 3 (9), e2022942. <https://doi.org/10.1001/jamanetworkopen.2020.22942>.
- Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accid. Anal. Prev.* 40 (4), 1486–1497.
- Qureshi, A.I., Huang, W., Khan, S., Lobanova, I., Siddiq, F., Gomez, C.R., Suri, M.F.K., 2020. Mandated societal lockdown and road traffic accidents. *Accid. Anal. Prev.* 146, 105747. <https://doi.org/10.1016/j.aap.2020.105747>.
- Regev, S., Rolison, J.J., Moutari, S., 2018. Crash risk by driver age, gender, and time of day using a new exposure methodology. *J. Saf. Res.* 66, 131–140.
- Rifaat, S.M., Tay, R., de Barros, A., 2012. Severity of motorcycle crashes in Calgary. *Accid. Anal. Prev.* 49, 44–49.
- Saladié, C., Bustamante, E., Gutiérrez, A., 2020. COVID-19 lockdown and reduction of traffic accidents in Tarragona province, Spain. *Transp. Res. Interdiscip. Perspect.* 8, 100218. <https://doi.org/10.1016/j.trip.2020.100218>.
- Shefer, D., 1994. Congestion, air pollution, and road fatalities in urban areas. *Accid. Anal. Prev.* 26 (4), 501–509.
- Stavrinos, D., McManus, B., Mrug, S., He, H., Gresham, B., Albright, M.G., Svancara, A. M., Whittington, C., Underhill, A., White, D.M., 2020. Adolescent driving behavior before and during restrictions related to COVID-19. *Accid. Anal. Prev.* 144, 105686. <https://doi.org/10.1016/j.aap.2020.105686>.
- Valent, F., Schiava, F., Savonitto, C., Gallo, T., Brusaferro, S., Barbone, F., 2002. Risk factors for fatal road traffic accidents in Udine, Italy. *Accid. Anal. Prev.* 34 (1), 71–84.
- Vingilis, E., Beirness, D., Boase, P., Byrne, P., Johnson, J., Jonah, B., Mann, R.E., Rapoport, M.J., Seeley, J., Wickens, C.M., Wiesenthal, D.L., 2020. Coronavirus disease 2019: What could be the effects on Road safety? *Accid. Anal. Prev.* 144, 105687. <https://doi.org/10.1016/j.aap.2020.105687>.
- Wang, D., He, B.Y., Gao, J., Chow, J.Y., Ozbay, K., Iyer, S., 2021. Impact of COVID-19 behavioral inertia on reopening strategies for New York City Transit. *Int. J. Transp. Sci. Technol.* 10 (2), 197–211.
- Wang, J., Huang, H., Zeng, Q., 2017. The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian. *Accid. Anal. Prev.* 98, 223–231.
- West, M.G., 2020. First case of coronavirus confirmed in New York State. *Wall Street Journal* 2.
- WHO. (2020). Novel Coronavirus (2019-nCoV) SITUATION REPORT - 1, 21 JANUARY 2020. In.
- WHO. (2021). *WHO Coronavirus (COVID-19) Dashboard* Retrieved Aug 30, 2021 from <https://covid19.who.int/>.
- Wu, X., Lu, Y., Lin, Y., Yang, Y., 2019. Measuring the destination accessibility of cycling transfer trips in metro station areas: A big data approach. *Int. J. Environ. Res. Public Health* 16 (15), 2641.
- Yabe, T., Tsubouchi, K., Fujiwara, N., Wada, T., Sekimoto, Y., Ukkusuri, S.V., 2020. Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic. *Sci. Rep.* 10 (1), 1–9.
- Yau, K.K.W., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. *Accid. Anal. Prev.* 36 (3), 333–340.
- Zhang, N., Jia, W., Wang, P., Dung, C.-H., Zhao, P., Leung, K., Su, B., Cheng, R., Li, Y., 2021. Changes in local travel behaviour before and during the COVID-19 pandemic in Hong Kong. *Cities* 112, 103139. <https://doi.org/10.1016/j.cities.2021.103139>.