Sustainable Cities and Society 89 (2023) 104318

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Exploring the Impact of Bike Lanes on Transportation Mode Choice: A simulation-based, route-level impact analysis

Uijeong Hwang, Subhrajit Guhathakurta *

School of City and Regional Planning, Georgia Institute of Technology, Atlanta, GA 30332, USA

ARTICLE INFO

Keywords:
Bike lanes
Active mobility
Mode choice
Bike Route Simulation

ABSTRACT

Many U.S. cities are investing in making a more bike-friendly environment in hopes of reducing auto-dependency. Studies have shown that improving bike lanes enhance bike users' perceived safety and comfort, but whether it also shifts mode choice towards more biking remains largely unaddressed. This study proposes a model to examine whether and how bike lanes shift the mode choice towards various non-automobile modes. With the help of more than 110,000 sample trips from travel surveys, hypothetical bike routes are obtained to assess each route's level of bike-friendliness. A mode choice model with four modes – walking, biking, driving, and transit – is developed using this data. The model results suggest that bike lanes increase not only biking trips but also walking and public transit trips. However, the impact on biking trips alone was marginal, suggesting a more comprehensive strategy may be necessary to make a significant transition towards non-auto trips. When the data are segmented by poverty level, model results show that the mode choice of the low-income population is as significantly affected by bike lanes as the general population. In addition, the study results indicate that frequent transit users can greatly benefit from bike infrastructure since walking, biking, and public transit trips are closely associated with each other.

1. Introduction

There has been a concerted effort in the U.S. to advance active mobility, particularly biking, to reduce car-dependence for short-distance trips, which also benefits the environment, public health, and the economy. Accordingly, a considerable amount of local, state, and federal budgets is being put into bike infrastructure development. For example, Atlanta, one of the most auto-dependent cities in the U.S., has invested multi-million dollars in extending the protected bike lane network in recent years (City of Atlanta, 2018, 2019). Besides the city government, many other organizations — such as Georgia DOT, Renew Atlanta (Atlanta transportation SPLOST), the PATH foundation, and the Atlanta BeltLine — are making significant investments in bike lanes.

A better bike lane network is expected to enhance the safety and comfort of bike users and encourage more people to ride a bike. Many earlier studies have examined people's stated preferences and supported the hypothesis that a better bike lane network is associated with more willingness to ride a bike (Akar & Clifton, 2009; Aldred & Dales, 2017; Clark, Mokhtarian, Circella & Watkins, 2019; Dill & Voros, 2007; Garrard, Rose & Lo, 2008; Moudon et al., 2005). However, we know far less about whether better bike lanes have influenced people's actual travel behavior and increased the number of bike users (Aziz et al., 2018; Buehler & Pucher, 2012; Dill & Voros, 2007; Krizek, Barnes & Thompson, 2009; Moudon et al., 2005; NACTO, 2016; Pedroso, Angriman, Bellows & Taylor, 2016; Schoner & Levinson, 2014; Zahabi, Chang, Miranda-Moreno & Patterson, 2016). Most previous studies have used cross-sectional analyses to show correlations between biking and bike lanes but cannot guarantee the direction of causality. For example, do people ride a bike more because there are bike lanes in their neighborhood? or does the city build more bike lanes in a neighborhood where people already choose to ride more?

Bike lanes not only support bike users and, perhaps, induce more bike riding, but it also accommodates micro-mobility users, which helps ensure a safer environment for pedestrians. In addition, a better bike lane network can enhance first/last-mile accessibility, thereby facilitating transit use. Therefore, bike lanes can play an essential role in creating a better environment for alternative modes of transportation and reducing car dependency. Improving the bike lanes would have greater implications for the underserved population who often have fewer transportation options. In other words, bike lanes can contribute

* Corresponding author.
E-mail addresses: uhwang3@gatech.edu (U. Hwang), subhro.guha@design.gatech.edu (S. Guhathakurta).

https://doi.org/10.1016/j.scs.2022.104318
Received 19 April 2022; Received in revised form 1 November 2022; Accepted 20 November 2022
Available online 23 November 2022
2210-6707/

Table 1
Variables for the Hierarchical Cluster Analysis.

Variable	Data Source
Population Density Employment Density Ratio of Employment to Population Per Capita Personal Income	Census Bureau & Bureau of Economic Analysis
Commute Mode Share by Private Vehicle / Public Transit / Bike / Walk Mean Commuting Time	Census Bureau
Density of Highway / Primary & Secondary Road / Local Road / Subway Line / Bike Lane	OpenStreetMap
Fuel Price	American Automobile Association
Bikeability Score	PeopleforBikes
5-year Average Temperature 5-year Total Precipitation	National Centers for Environmental Information
Terrain	Google Maps Elevation API

to promoting transportation equity depending on how it is implemented. However, we find limited research on the relationship between bike lanes and mode choice that focus on underserved populations. Our study is designed to address this gap.

This study examines whether and to what extent the bike lanes affect mode choice for a short-distance trip *at the route level*. Route-level mode choice research has not been explored simply due to a lack of data: that is, without first knowing the routes people have taken, we cannot examine the mode choice at the route level. This study takes a novel approach in which bike routes are *simulated*. Based on a path-finding algorithm and actual origin-destination (O-D) coordinates from travel surveys, this study simulates hundreds of thousands of bike routes. A mode choice model then employs the information from the simulated routes (e.g., the proportion of bike lanes in a route) to examine how it affects mode choices – walking, biking, driving, and public transit. The estimation in the model uses two types of samples—a whole sample and a low-income sample—so that we can infer the impact of bike lanes on the mode choice of the general population as compared to the low-income population. Through this analysis, we can have a better picture of how people's mode choices might change as they encounter a street environment that ensures safety and comfort for non-motorized mode users such as pedestrians or cyclists.

2. Literature review

Previous studies that examined stated preference surveys have shown that the type and quality of bike lanes substantially influence people's perception of biking (Hull & O'Holleran, 2014; Johns, 2012) as well as the level of safety (Marshall & Ferenchak, 2019; Pedroso et al., 2016). For example, a survey conducted by Clark et al. (2019) found that people reported higher levels of safety and comfort for biking when protected bike lanes and multi-use paths are presented. This research also concluded that perceived safety and comfort are significantly related to people's willingness to try. Similarly, a survey conducted at the University of Maryland-College Park campus showed that both bike users and non-bike users stated that improved bike lanes and trails would encourage them to use bikes for commuting to the campus (Akar et al., 2008). Studies based on actual biking behavior reinforce those findings. Garrard et al. (2008) observed 6589 bike users in Melbourne, Australia, and found that there are gender differences in risk aversion by noting that female bike users prefer to use bike routes that are well separated from motorized traffic. A study based on 4360 bike users in London, UK, drew similar conclusions: women and older people are more likely to ride on protected bike lanes than on typical roads. In addition, people who ride on protected bike lanes are less likely to wear helmets, which shows bike lanes increase their perceived level of safety (Aldred & Dales, 2017).

While empirical studies on the impact of bike infrastructure on bike mode choice have been conducted, these are mostly cross-sectional studies that are limited in showing the direction of causality. The authors found only one study (Xu & Chow, 2020) that conducted a longitudinal analysis of the impact of bike lanes on the daily trip count of shared bikes in New York City. The study found that improving bike lanes has a meaningful impact on bike ridership within Manhattan but it was not as effective outside of the area.

Most of the cross-sectional studies undertaken at the city level noted a significantly positive correlation between bike lanes and bike ridership (Buehler & Pucher, 2012; NACTO, 2016; Pedroso et al., 2016; Schoner & Levinson, 2014). For example, a study based on 90 large U.S. cities showed a positive correlation between bike lanes/paths and bike commuting rates (Buehler & Pucher, 2012). On the other hand, Krizek et al. (2009) found that the relationship between proximity to bike facilities and bike mode share was not significant in Minneapolis, Minnesota. It is difficult to tease out from these city-level studies whether a better bike lane network leads to higher bike mode share or whether cities spend more on bike lanes and paths as the number of bike users surges.

Results from individual-level studies are also divergent. According to Aziz et al. (2018), who analyzed travel survey data from five counties around New York City, a 1% increase in bike lanes (in the home and work census tracts) increases the probability of bike commuting by 1.13%. A similar study based on travel survey data from Montreal, Canada, also identified a significant association between bike lane accessibility from home and bike mode choice (Zahabi et al., 2016). Also, a study based on households in Bogota, Colombia, found that the proximity from home to bike lanes has a positive but marginal impact on the probability of making bike trips (Rodriguez-Valencia, Rosas-Satizábal, Gordo & Ochoa, 2019). In contrast, Moudon et al. (2005) and Dill and Voros (2007) drew an opposite conclusion: the presence/density of bike lanes near the home location was not associated with bike ridership. These individual-level studies are more able to examine the direction of causality when compared to city-level studies. However, these studies include independent variables, such as proximity to or density of bike infrastructure near the respondents' home location, which may not reflect people's trip-level experience of the bike infrastructure. This study addresses these limitations by quantifying bike lanes along individual routes (from an origin to a destination) to capture the trip-level experience more accurately.

3. Conceptual framework

3.1. Obtaining bike route information through network simulation

This project seeks to answer the following research question: Does the availability of bike lanes affect mode choice? To answer this question, we design a model in which an individual's mode choice is largely affected by characteristics of 1) the person, 2) the trip, and 3) the route. The personal characteristics include a traveler's socio-demographic

Table 2
Traffic Stress in terms of MRS.

Number of Lanes	2	2	2–3	4–5	2–3	6+	4–5	6+	2–3	4–5	6+
Speed Limit	25	30	25	25	30	25	30	30	35+	35+	35+
Traffic Stress (in MRS)	10%	15%	20%	35%	40%	67%	70%	80%	100%	120%	140%

Note: Data are from Lowry et al., 2016.

Table 3
Model Specification and Data Source.

Category	Variable	Data Source
Dependent variable	Mode choice (Walking / Biking / Transit / Driving)	Travel Surveys
Trip Characteristics	Travel Time	Google Directions API
	Distance to the Nearest Subway Station (mile)	–
	Weekend/Weekday	Travel Surveys
	Feels-like Temperature at Departure Time	Visual Crossing API
	Weather of the Day	
	Purpose	Travel Surveys
	Intra-city Trip	–
	Region (Atlanta / Chicago / the Twin Cities)	–
Socio-demographic Characteristics	Age	Travel Surveys
	Gender	
	Race	
	Disability	
	Driver's License	
	Household Income	
	Frequent Transit User	
Bike-friendliness of the route	Average Traffic Stress (%)	Network Simulation
	Proportion of Bike Lanes (%)	
	Average Slope (%)	

conditions such as age, gender, race, income level, or physical ability. The trip characteristics are about a specific trip such as purpose, distance, cost, date and time, and weather. The route characteristics can be considered as a part of trip characteristics, but it is more associated with conditions along a particular route such as the level of congestion, speed limit, quality of sidewalks, the presence of bike lanes, slope, and aesthetics of the surrounding environment. While previous studies have only focused on the first two, route characteristics are also fundamental factors that affect an individual's travel behavior. Particularly for active transportation modes, travelers are much more sensitive to the condition along the routes.

The primary reason that route characteristics have not been considered in previous studies is that data on actual routes selected are rarely available. In addition, analyzing the impact of bike lanes on mode choice requires not just bike routes for bike users but also for other mode users. Our approach, therefore, is to 'simulate' a bike travel route for every trip using the O-D coordinates, regardless of the respondents' actual modes. Through the simulation, we obtain information on the 'bike-friendliness' of the hypothetical route. This approach is based on counterfactual thinking that estimates what route would be chosen if the respondent had taken a bike for the trip. The bike-friendliness information includes the following three variables: *Proportion of Bike Lanes, Average Traffic Stress*, and *Average Slope*. The next section discusses how the simulated bike trips are derived and how the bike-friendliness variables are measured.

3.2. Selection of the study areas

This study employs individual trips in travel surveys as observations for estimating a mode-choice model. Building a discrete choice model with modes of transportation including bike and public transit requires a significantly large sample size because trips made by either of the two modes are relatively rare. However, the travel survey data in our initial area of interest, Atlanta, Georgia, does not consist of a sizable sample, which necessitated securing more trip samples. Therefore, we included additional travel survey data from two other cities: Chicago, Illinois, and the Twin Cities (Minneapolis + St. Paul), Minnesota. The two cities are chosen since (1) they are closest to Atlanta in terms of demographic, economic, transportation, and environmental characteristics according to a hierarchical cluster analysis (described later) and (2) have recently completed a region-wide travel survey with O-D coordinates and were willing to share this data.

3.3. Modeling mode choices

To quantify the impact of improving bike lanes on each mode of transportation, the mode choice model includes information about bike-friendliness among other explanatory variables. The model includes four modes—walking, biking, driving, and public transit—among which driving is the base category. Therefore, the significance of the bike-friendliness variables as well as the other covariates show the likelihood of choosing walking, biking, and public transit compared to driving. The model will be estimated using two types of samples: a complete sample and a separate sample of individuals with low income. The two model estimation results will provide insights into how the bike-friendliness variables affect the mode choice of the general population and the low-income population in particular.

4. Methods and data

4.1. Hierarchical cluster analysis

This study uses hierarchical cluster analysis to select cities that have similar characteristics to Atlanta. The hierarchical clustering algorithm first calculates the distance between data points. The dataset is normalized so that each variable is equally considered in calculating the distance. Euclidean distance is used for this analysis. From this distance matrix, the algorithm builds a hierarchy of clusters. There are multiple clustering methods available at this step and this study uses the complete linkage method. Complete linkage is one of the agglomerative clustering schemes that iteratively combine two clusters until all the inputs become a single cluster. At each step, the two clusters with the shortest distance are combined, where the complete linkage method defines the shortest distance based on the two most distant elements. This method is known to be more robust when the data has outliers. Compared to non-hierarchical cluster analysis, hierarchical clustering has the advantage that the number of clusters need not be assigned *a priori*. In addition, the dendrogram derived from this analysis allows us to visually check the distance between observations and determine how the sequence of cluster fusion has taken place.

The unit of analysis for the clustering method is the county. To include only counties with similar contexts to Atlanta (Fulton County) in the analysis, we first selected all U.S. Counties that meet the following three criteria: 1) its county seat is the principal city within a Metropolitan Statistical Area, 2) it has a population of more than 400,000, and 3) it has a metro/light rail system. These criteria resulted in 28 counties, which were clustered based on the variables listed in Table 1. Among them, the density-related variables were computed by using the urbanized area as the denominator. For the 'bikeability score', we used Peopleforbikes' Bicycle Network Analysis score which measures accessibility via a low-stress bike network. 'Terrain' represents how flat an urbanized area is; a value of 0 means the area is completely flat. It is calculated as the average standard deviation of elevations (extracted at intervals of 500$\times$500 feet) in each neighborhood (3 $\times$ 3 mile area). In other words, it is a measure of the average elevation change that occurs when a person makes a short-distance trip.

4.2. Bike route simulation

In defining an optimal bike route, this study assumes that bike users seek a route with the lowest overall stress, where the stress is induced by 1) the biking distance, 2) the traffic, and 3) the slope. The stress from each of the three factors would vary by individual – some people may want to take the shortest route, while others prefer a safer, flatter, but longer route. Since we cannot personalize those preferences by the sampled individuals, this study is based on reasonable logic that estimates how much stress the general population would feel from biking

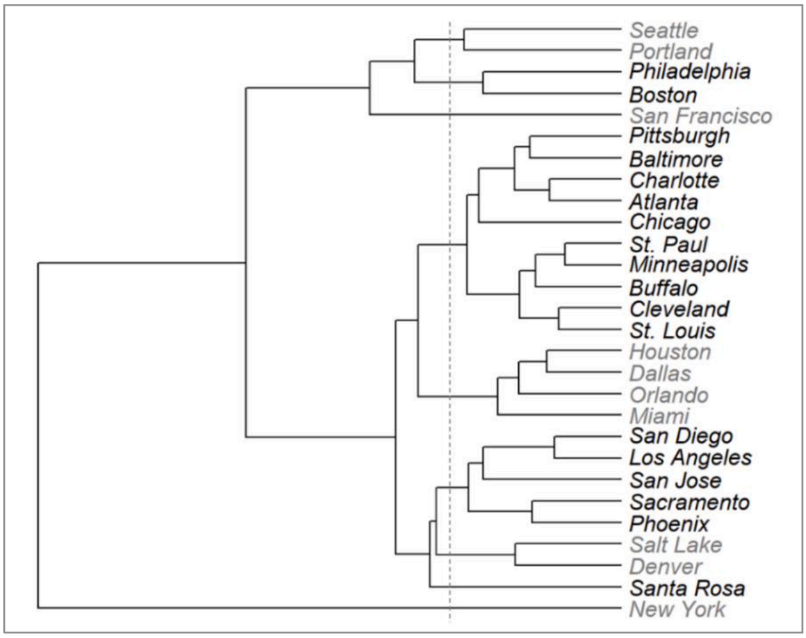

Fig. 1. Hierarchical Cluster Analysis Result.

distance relative to traffic or slope. We borrow insights from Hood, Sall and Charlton (2011) and Broach, Dill and Gliebe (2012) that quantified the factors that affect biking stress in terms of the concept of the marginal rate of substitution (MRS). In a later study, Lowry, Furth and Hadden-Loh (2016) leveraged the findings from the two studies and summarized the stress values in MRS based on the number of lanes, speed limit, slope, and the presence of dedicated bike lanes. MRS, in this context, can be thought of as the distance bike users are willing to substitute with the stress from a given traffic/biking condition. For example, 65% of traffic stress in terms of MRS means an individual is willing to travel 65% more distance for a zero-traffic-stress detour. In other words, MRS quantifies the trade-off between the stress from traffic/slope and the stress from prolonged biking and is, therefore, a good proxy to show how much the traffic stress would reduce by implementing bike lanes.

Table 2, based on the work by Lowry et al. (2016), summarizes traffic stress induced by the number of lanes and the speed limit. According to this study, striped bike lanes, buffered bike lanes, and protected bike lanes can reduce stress by 50%, 65%, and 75%, respectively. Using these three factors, it calculates the traffic stress of road segments. For example, the traffic stress of a street with four lanes, a 35mph speed limit, and protected bike lanes would be: 120% * (100%–75%) = 30%.

The slope of the route is also a crucial stress factor for biking. The stress induced by the slope is as follows: 37% in MRS if the slope is 2–4%; 120% if the slope is 4–6%; 320% if the slope is greater than 6% (Broach et al., 2012). This study measures 'biking stress' by adding the stress induced by the slope to the 'traffic stress.' We made the distinction between the two to highlight their different uses: the 'traffic stress' is used in the mode choice model, while 'biking stress' is used when simulating bike trip routes.

For the network simulation, we employ the A-star shortest path search algorithm. The A-star algorithm is an advanced form of Dijkstra's algorithm. While Dijkstra's algorithm compares every possible route in the network, the A-star algorithm includes a heuristic function that guides its search direction. Thus, it requires much less computation than Dijkstra's. The algorithm searches for the shortest path but depending on how we define the objective to optimize for each segment (e.g., distance to find the shortest path), it can be used for other purposes. To find an optimal bike route, the distance of each road segment is weighted by the biking stress, so that the algorithm can find the stress-minimizing route from the cyclists' perspective.

Using the algorithm, we find a bike route for every trip whose travel distance is shorter than 3 miles which accounts for 90 percent or more of the bike trips and is consistent with existing studies (Bearn, Mingus & Watkins, 2018; Martens, 2004). From these trips, we extract the values of the bike-friendliness variables for the route. The *Proportion of Bike Lanes* is the ratio of the distance of streets that have a dedicated bike lane (protected or not) to the total distance of the route. The *Average Traffic Stress* is the distance-weighted average of the traffic stress – measured by the number of vehicular lanes, speed limit, and the presence of dedicated bike lanes – at the road segment level. Similarly, the *Average Slope* is the distance-weighted average of the slope of each road segment. Therefore, if a route has a high proportion of bike lanes, low average traffic stress, and a low average slope, the route can be considered bike-friendly.

All the data analytical tasks in this study – such as data cleaning and manipulation, network simulation, and logistic regressions – were done with the programming language R. The A-star search algorithm is implemented using an R package called 'astar-r'. Since the network simulation requires a huge amount of computing resources, this study utilized a high-performance computing environment (or supercomputer) called the PACE cluster at the Georgia Institute of Technology. The network simulation took 4,631 CPU hours[1] in total.

4.3. Mode choice model

Multinomial logistic regression is employed to build a mode choice model with the four modes: walking, biking, private vehicle, and public transit. Private vehicle is set as a comparative baseline. We can simply think of this multinomial logistic regression as three separate binomial logistic regression models: walking vs. private vehicle, biking vs. private vehicle, and public transit vs. private vehicle.

As shown in Eqs. (1)–(3), multinomial logistic regression can be expressed as a series of linear predictor functions that construct a logit value from a set of coefficients that are linearly combined with the independent variables. That means, from the set of coefficients and given

[1] 4,631 CPU hours are calculated based on 44 batch jobs, 3,947 seconds per job on average, 4 nodes per job, and 24 processors per node: 44 * 3,947/3,600 * 4 * 24 = 4,631 CPU hours.

Table 4
Descriptive Statistics of the Three Study Areas.

City (County)		Atlanta(Fulton County)	Chicago(Cook County)	The Twin Cities	
				Minneapolis(Hennepin County)	St. Paul(Ramsey County)
Population		1,063,937	5,150,233	1,265,843	550,321
Employment		1,201,235	3,600,552	1,204,282	424,967
Urbanized area (mi^2)		424	906	404	428
Population Density		2,507	5,684	3,137	3,328
Employment Density		2,830	3,973	2,984	2,570
Ratio of Employment to Population		1.13	0.7	0.95	0.77
Per Capita Income ($)		88,832	65,306	76,552	55,583
Commuting Mode Share (%)	Private Vehicle	78.5	69.1	80.4	83.2
	Public Transit	7.5	19.1	7.1	6.7
	Walking	2.7	4.3	3.5	2.7
	Biking	0.5	1.1	1.7	1
	Other Modes	10.8	6.4	7.3	6.4
Mean Commuting Time (minute)	28.8	33.4	23.7	24.2	
Infrastructure Density (mile/urbanized area (mi^2))	Highway	0.55	0.61	0.95	1.06
	Primary & Secondary Road	1.44	2.47	1.91	1.45
	Residential Road	8.43	10.69	8.97	8.17
	Railway	0.20	0.32	0.09	0.11
	Bike Lane	0.27	0.29	0.59	1.02
Length of Bike Lanes (mile)		113.7	262.6	238.1	169.4
Gas Price ($)		2.37	2.76	2.30	2.31
Bikeability Score		23	37	35	42
5-year Precipitation (inch)		283	213	168	171
Average Temperature		63°F (17 °C)	51°F (11 °C)	46°F (8 °C)	46°F (8 °C)
Terrain (Std. Dev. in meters)		14.6	4.5	7.6	12.6

values, we can calculate a logit for a certain category of observation *i*, from which we can then calculate the probability for *i* to choose the category.

$$ln\frac{\Pr(Y_i = walk)}{\Pr(Y_I = car)} = \beta_{walk} * X_i \tag{1}$$

$$ln\frac{\Pr(Y_i = bike)}{\Pr(Y_i = car)} = \beta_{bike} * X_i \tag{2}$$

$$ln\frac{\Pr(Y_i = transit)}{\Pr(Y_i = car)} = \beta_{transit} * X_i \tag{3}$$

Assuming that there are only four modes available, the probability of choosing a car equals the probability of not choosing any other three modes, which can be expressed as Eq. (4). Therefore, as shown in Eqs. (5)–(7), probabilities of either walking, biking, or public transit can be calculated using the set of coefficients from each of the three models and given values of the explanatory variables.

$$\begin{aligned}\Pr(Y_i = car) &= 1 - \Pr(Y_i = walk) - \Pr(Y_i = bike) - \Pr(Y_i = transit)\\ &= \frac{1}{1 + e^{\beta_{walk}*X_i} + e^{\beta_{bike}*X_i} + e^{\beta_{transit}*X_i}}\end{aligned} \tag{4}$$

$$\Pr(Y_i = walk) = \frac{e^{\beta_{walk}*X_i}}{1 + e^{\beta_{walk}*X_i} + e^{\beta_{bike}*X_i} + e^{\beta_{transit}*X_i}} \tag{5}$$

$$\Pr(Y_i = bike) = \frac{e^{\beta_{bike}*X_i}}{1 + e^{\beta_{walk}*X_i} + e^{\beta_{bike}*X_i} + e^{\beta_{transit}*X_i}} \tag{6}$$

$$\Pr(Y_i = transit) = \frac{e^{\beta_{transit}*X_i}}{1 + e^{\beta_{walk}*X_i} + e^{\beta_{bike}*X_i} + e^{\beta_{transit}*X_i}} \tag{7}$$

Table 3 shows the list of explanatory variables and their data source. All variables except for 'travel time' are choice-situation-specific variables. Those variables provide a set of coefficients which represent an impact on the mode shift from a private vehicle to another mode. It means those variables return coefficients for the three modes: walking, biking, and public transit. On the other hand, 'travel time' is an alternative-specific variable. This variable can have a different value for each alternative. For example, if a certain O-D trip takes 15 min by driving, it would take at least 30 min by walking. Unlike the choice-situation-specific variables, alternative-specific variables return coefficients for all categories. In this case, the coefficient represents the variable's impact on choosing that specific mode. We implemented this model specification using the R package 'mlogit' (Croissant, 2012).

The primary data source in this study is household travel surveys in each of the three regions: 1) '2017 National Household Travel Survey (NHTS)' Georgia add-on survey conducted by the Federal Highway Administration, 2) '2018–19 My Daily Travel' survey from Chicago Metropolitan Agency for Planning, and 3) '2018–19 Travel Behavior Inventory' survey from Metropolitan Council in the Twin Cities. Those travel survey data consist of 1) respondents' personal and household information, 2) their trip information in a day by each purpose and mode, and 3) O-D information of each trip at the coordinates level.[2] The O-D coordinates are used for the network simulation. Data for most covariates—such as trip purpose and traveler's age, gender, race, disability, and income—are from travel surveys.

Data for constructing the network such as streets, bike lanes, and railways are from OpenStreetMap. Since it is crowd-sourced map data, we cannot say OpenStreetMap is 100% accurate. Regardless, it provides data reliable enough for major cities like Atlanta, Chicago, and the Twin Cities. Since the raw data in OpenStreetMap is not in a suitable form for network simulation, we converted it into graph data using a Python library 'OSMnx' (Boeing, 2017). The number of lanes which is required to calculate traffic stress is also collected from OpenStreetMap.

Most of the variables describing trip characteristics and socio-demographic characteristics are self-explanatory. Table A3 in Appendix shows categories and descriptive statistics for all the variables. 'Feels-like temperature at departure time' and 'weather of the day' are collected using Visual Crossing Weather API. From their historical weather database, we obtained the temperature and weather information of each trip at its departing location and time. 'Intra-city trip' indicates a trip that has both its origin and destination inside the city boundary. 'Frequent transit user' is a dummy variable indicating those

[2] In the case of the Twin Cities data, due to concerns about data privacy issues, we received coordinates data that was trimmed to three decimal places. That decreases the accuracy of the coordinates approximately by ±160 feet, but it is not a serious issue for the purpose of this study.

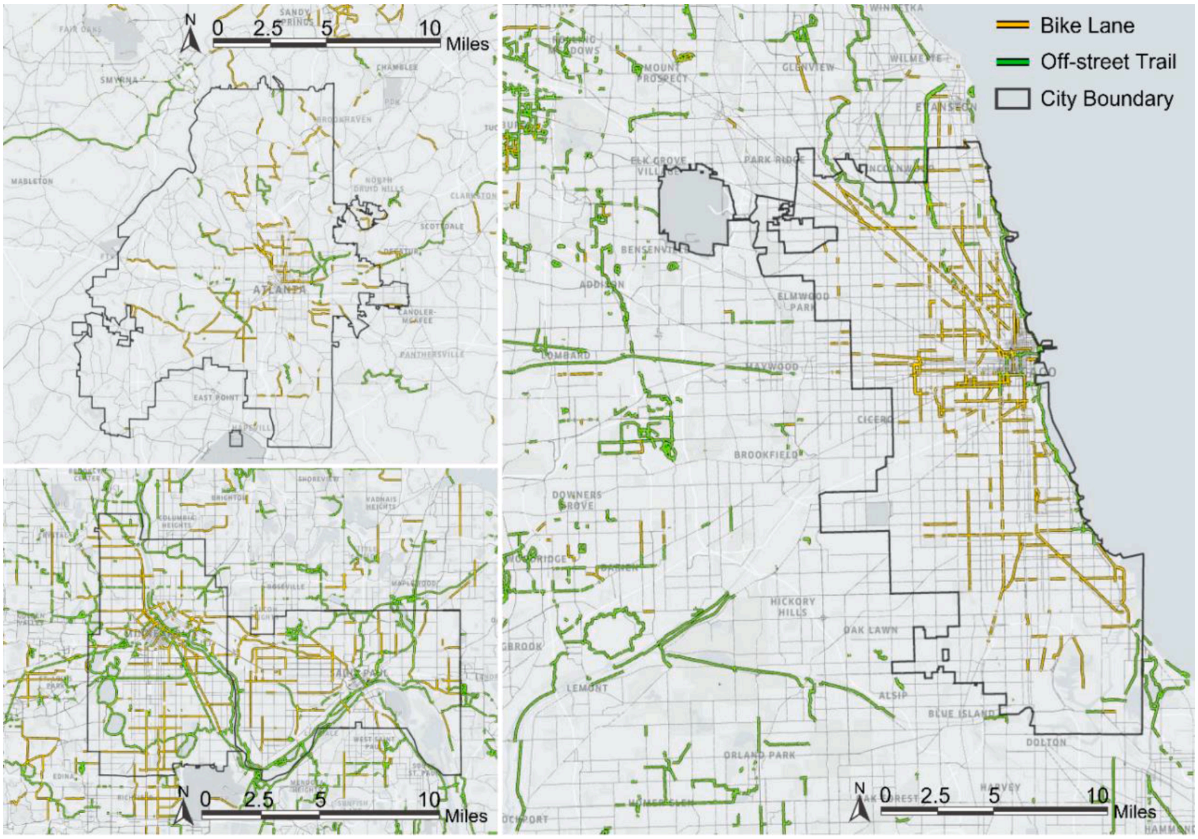

Fig. 2. Bike Networks in the Study Areas (clockwise from top-left: Atlanta, Chicago, and the Twin Cities).

who use public transit at least once a week. Variables in the Bike-friendliness category – 'average traffic stress', 'proportion of bike lanes', and 'average slope' – are explained in Section 4.2 *Bike route simulation*.

Note that, since 'travel time' is an alternative-specific variable in the mode choice model, we need travel time for each mode for every trip regardless of its actual mode. Fortunately, Google Directions API provides route information between locations for all four modes: walking, biking, driving, and public transit. Routes using the first three modes are available for every trip even though it may seem unfeasible: we do not want to walk for 3 h, but at least the route exists. However, a lot of O-Ds do not have public transit options: it turned out that 58.5% of trips in the sample cannot be made by public transit. Therefore, in the mode choice model, 58.5% of the trips have three mode choices and the rest 41.5% have four mode choices. Furthermore, since all three cities in this study have both bus and train systems, the public transit trip can be made by either bus, train, or a combination of the two modes. This study lumped the bus trip, train trip, and bus + train trip into a single mode category in the mode choice model.

5. Results

5.1. Hierarchical cluster analysis

Fig. 1 shows the dendrogram of the hierarchical clusters.

We can notice that cities in each cluster are geographically close as well. It implies that, from the perspective of the whole nation, geographically adjacent cities are likely to share similar properties in terms of demography, infrastructure, travel behaviors as well as climate. Although how the cities are grouped in the hierarchical clustering is a matter of selecting the section from which the tree segment is extracted, the dendrogram suggests a generous number of cities that are closest to the travel behavior characteristics of Atlanta. There are nine cities: Charlotte NC, Pittsburgh PA, Baltimore MD, Chicago MI, St. Paul and Minneapolis MN, Buffalo NY, Cleveland OH, and St. Louis MO. Minneapolis and St. Paul are considered as one city region: the Twin Cities. Based on the availability of travel survey data, we finally selected Chicago and the Twin Cities to include in the model.

Table 5
Descriptive Statistics of the Bike-friendliness Variables.

Category		Walking	Biking	Transit	Driving
Average Traffic Stress	Atlanta	24%	20%	34%	34%
	Chicago	16%	14%	18%	15%
	Twin Cities	15%	14%	16%	17%
Proportion of Bike Lanes	Atlanta	15.5%	36.0%	9.5%	21.3%
	Chicago	12.1%	15.8%	9.1%	19.8%
	Twin Cities	18.2%	26.6%	19.5%	22.3%
Average Slope	Atlanta	2.38%	2.35%	2.27%	2.05%
	Chicago	0.81%	0.70%	0.76%	0.72%
	Twin Cities	1.30%	1.26%	1.29%	1.25%

Table 4 summarizes the descriptive statistics of the three study areas.[3] Although they are clustered in a group, the table demonstrates that the three cities are different in many ways. It would be appropriate to say that the hierarchical cluster analysis yielded the 'least dissimilar' cities to Atlanta among the 28 cities. In terms of travel behaviors, Chicago shows a significantly higher commuting mode share of public transit while the Twin Cities shows a higher bike mode share in commuting. Interestingly, this trend is associated with the infrastructure density: Chicago has the highest railway density and the Twin Cities have the highest bike lane density. Combining data from the three cities

[3] The descriptive statistics of the other six cities in the cluster is in Table A2 in Appendix.

Table 6
Multinomial Logistic Regression Result: Whole Population Model.

Variable		Walking Coef.	p	Biking Coef.	p	Public Transit Coef.	p	Driving Coef.	p
Travel Time		−0.16***	0.000	−0.12***	0.000	−0.07***	0.000	−0.11***	0.000
(Intercept)		4.91***	0.000	2.24***	0.000	−4.72***	0.000		
Distance to the Nearest Subway Station		−0.06***	0.000	−0.20***	0.000	−0.37***	0.000	*** significant at	
Weekend/Weekday (Weekend = 1)		−0.23***	0.000	−0.11*	0.050	−0.66***	0.000	$p = 0.001$ or better	
Feels-like Temperature at Departure Time	Below 30°F (∼ −1 °C)	−0.64***	0.000	−1.88***	0.000	0.09	0.224	** significant at	
	30∼39°F (−1∼4 °C)	−0.35***	0.000	−1.02***	0.000	−0.07	0.401	$p = 0.01$ or better	
	40∼49°F (4∼9 °C)	−0.22***	0.000	−0.66***	0.000	−0.01	0.880	* significant at	
	50∼59°F (10∼15 °C)	−0.02	0.607	−0.25***	0.000	0.00	0.956	$p = 0.05$ or better	
	60∼69°F (16∼21 °C)	(base)						ˆ significant at	
	70∼79°F (21∼26 °C)	0.02	0.604	0.08	0.166	−0.06	0.407	$p = 0.1$ or better	
	80∼89°F (27∼32 °C)	−0.12*	0.011	0.10	0.183	0.01	0.941		
	Above 90°F (32 °C∼)	0.04	0.673	0.39*	0.011	−0.12	0.608		
Weather	Sunny	0.16***	0.000	0.28***	0.000	0.10	0.122		
	Cloudy	(base)							
	Rain	−0.16***	0.000	−0.18***	0.000	0.05	0.367		
	Snow	−0.08*	0.043	−0.29*	0.043	0.06	0.457		
Purpose	Home	−1.22***	0.000	−0.65***	0.000	−0.30***	0.000		
	Work	−0.97***	0.000	−0.10	0.159	0.39***	0.000		
	School	−0.82***	0.000	0.26*	0.017	0.68***	0.000		
	Social/Recreational	(base)						$N = 118{,}423$	
	Shopping	−2.35***	0.000	−1.35***	0.000	−0.89***	0.000	Pseudo-$R^2 = 0.44$	
	Errand	−2.49***	0.000	−1.49***	0.000	−1.25***	0.000		
	Meal	−1.66***	0.000	−1.41***	0.000	−0.87***	0.000		
	Drop-off/Pick-up	−2.52***	0.000	−2.14***	0.000	−1.61***	0.000		
	Other Purposes	−1.64***	0.000	−0.52***	0.000	0.44***	0.000		
Region	Atlanta	(base)							
	Chicago	−0.18*	0.011	0.34ˆ	0.057	0.90***	0.000		
	Twin Cities	−0.21**	0.001	0.34*	0.047	0.45*	0.040		
Intra-city Trip		1.24***	0.000	1.23***	0.000	1.31***	0.000		
Age	Under 15	−1.48***	0.000	−1.64***	0.000	−1.01***	0.000		
	16–24	0.40***	0.000	−0.03	0.761	0.29***	0.000		
	25–34	0.29***	0.000	0.12*	0.023	0.24***	0.000		
	35–44	(base)							
	45–54	0.00	0.928	−0.51***	0.000	0.26***	0.001		
	55–64	0.10**	0.006	−0.29***	0.000	0.36***	0.000		
	65 or older	0.01	0.712	−0.49***	0.000	0.42***	0.000		
Gender (Female = 1)		−0.13***	0.000	−0.56***	0.000	−0.11**	0.007		
Race	Asian	−0.28***	0.000	−0.34***	0.000	0.25**	0.005		
	Black	−0.55***	0.000	−0.91***	0.000	0.49***	0.000		
	Other Races	−0.24***	0.000	−0.19*	0.026	0.38***	0.000		
	White	(base)							
Disability		−0.39***	0.000	−0.18	0.180	0.51***	0.000		
Driver's License		−1.55***	0.000	−1.34***	0.000	−2.01***	0.000		
Household Income	Less than $25,000	0.50***	0.000	1.01***	0.000	1.04***	0.000		
	$25,000–50,000	0.08*	0.039	0.31***	0.000	0.34***	0.000		
	$50,000–75,000	(base)							
	$75,000–100,000	0.11**	0.001	0.15*	0.040	−0.20**	0.009		
	$100,000–150,000	0.15***	0.000	0.21**	0.001	−0.30***	0.000		
	$150,000 or more	0.25***	0.000	0.19**	0.005	−0.48***	0.000		
Frequent Transit User		1.15***	0.000	1.02***	0.000	2.74***	0.000		
Average Traffic Stress		−1.14***	0.000	−2.62***	0.000	2.08***	0.000		
Proportion of Bike Lanes	**Atlanta**	0.47*	0.050	1.87***	0.000	0.84	0.287		
	Chicago	0.74***	0.000	0.65***	0.000	1.26***	0.000		
	Twin Cities	0.26***	0.000	0.53***	0.000	0.13	0.265		
Average Slope		−1.50	0.285	−12.1***	0.000	−10.7**	0.008		

would increase variations in independent variables. For example, the Twin Cities data allow us to model how cold, snowy weather would affect mode choices, which would be very difficult to achieve with Atlanta samples only.

Fig. 2 shows the bike networks in the three study areas. The designated paths described as 'Bike lanes' (yellow) indicate dedicated bike lanes along vehicular streets, including striped bike lanes, buffered bike lanes, and protected bike lanes. Most of these are striped lanes.[4] The bike network of Atlanta mostly consists of dedicated bike lanes and is less dense than the other two cities. Chicago has provided dedicated bike lanes mostly within the city boundary, while there are many off-street trails (green) outside the city. The Twin Cities has the densest bike network. Unlike Atlanta and Chicago, the Twin Cities have a high proportion of off-street trails, attributed to the extensive network along the Mississippi river.

5.2. Descriptive statistics

Based on the three cities' household travel survey data, we created a sample of 118,423 short-distance (less than 3 miles) trips. The sample includes trips of the following four modes: walking, biking, driving, and public transit. Out of the 118,423 trips, 4,285 trips are from Atlanta, 38,814 trips are from Chicago, and 75,324 trips are from the Twin Cities.

[4] Among the on-street dedicated bike lanes, striped bike lanes are 93% (in length) in Atlanta, 71% in Chicago, 95% in the Twin Cities. Chicago has a relatively higher proportion of buffered lanes (17.5%) and protected lanes (11.4%) than the other two cities

Table 7
Multinomial Logistic Regression Result: Low-income Population Model.

Variable		Walking Coef.	p	Biking Coef.	p	Public Transit Coef.	p	Driving Coef.	p
Travel Time		**−0.15*****	0.000	**−0.22*****	0.000	**−0.08*****	0.000	**−0.15*****	0.000
(Intercept)		**5.08*****	0.000	**5.19*****	0.003	**−1.96ˆ**	0.094	*** significant at	
Distance to the Nearest Subway Station		**−0.10*****	0.000	**−0.18*****	0.001	**−0.23*****	0.000	$p = 0.001$ or better	
Weekend/Weekday (Weekend = 1)		**−0.40*****	0.000	**−0.37ˆ**	0.051	**−0.64*****	0.001	** significant at	
Feels-like Temperature at Departure Time	Below 30°F (∼ −1 °C)	**−0.46*****	0.000	**−1.95*****	0.000	0.04	0.827	$p = 0.01$ or better	
	30∼39°F (−1∼4 °C)	**−0.34***	0.019	**−0.83*****	0.000	0.07	0.725	* significant at	
	40∼49°F (4∼9 °C)	0.24	0.150	0.00	0.988	0.13	0.544	$p = 0.05$ or better	
	50∼59°F (10∼15 °C)	0.10	0.485	−0.09	0.673	0.21	0.279	ˆ significant at	
	60∼69°F (16∼21 °C)	(base)						$p = 0.1$ or better	
	70∼79°F (21∼26 °C)	0.10	0.417	0.26	0.155	0.07	0.696		
	80∼89°F (27∼32 °C)	−0.05	0.758	0.38	0.113	**0.45ˆ**	0.074		
	Above 90°F (32 °C∼)	**−0.92****	0.007	**1.33****	0.004	−0.47	0.489		
Weather	Sunny	**0.61*****	0.000	**1.20*****	0.000	**0.34***	0.042		
	Cloudy	(base)							
	Rain	−0.03	0.774	0.08	0.553	0.04	0.741		
	Snow	−0.15	0.322	−0.63	0.172	−0.14	0.448		
Purpose	Home	**−0.70*****	0.000	**−0.54****	0.004	−0.27	0.119		
	Work	**−0.38***	0.021	0.39	0.102	−0.01	0.955	$N = 8077$	
	School	−0.15	0.411	**0.74****	0.007	**0.61****	0.009	Pseudo-$R^2 = 0.45$	
	Social/Recreational								
	Shopping	**−1.39*****	0.000	**−0.99*****	0.000	**−0.61****	0.002		
	Errand	**−1.55*****	0.000	**−1.59*****	0.000	**−0.93*****	0.000		
	Meal	**−1.49*****	0.000	**−1.22*****	0.000	**−0.96*****	0.000		
	Drop-off/Pick-up	**−2.06*****	0.000	**−2.33*****	0.000	**−1.87*****	0.000		
	Other Purposes	**−0.77*****	0.000	0.13	0.614	0.14	0.533		
Region	Atlanta	(base)							
	Chicago	**−0.49***	0.025	**1.37ˆ**	0.074	**0.63ˆ**	0.098		
	Twin Cities	**−0.84*****	0.000	1.06	0.158	0.17	0.631		
Intra-city Trip		**0.74*****	0.000	**0.88*****	0.000	**1.37*****	0.000		
Age	Under 15	**−1.17*****	0.000	**−2.04*****	0.000	**−0.87*****	0.000		
	16–24	**0.42****	0.003	**0.52***	0.023	**0.47***	0.012		
	25–34	**0.44*****	0.001	0.10	0.667	0.22	0.245		
	35–44	(base)							
	45–54	**−0.31***	0.048	0.27	0.267	**0.65*****	0.001		
	55–64	0.19	0.196	0.35	0.157	**0.58****	0.004		
	65 or older	**−0.51*****	0.001	**−0.72***	0.011	**0.68****	0.002		
Gender (Female = 1)		−0.05	0.507	**−0.62*****	0.000	0.00	0.997		
Race	Asian	0.17	0.271	−0.37	0.183	**0.61****	0.009		
	Black	**−0.36*****	0.000	**−0.89*****	0.000	**0.77*****	0.000		
	Other Races	**0.36****	0.003	−0.34	0.158	**0.47****	0.005		
	White	(base)							
Disability		**−0.35****	0.010	**−0.42ˆ**	0.078	**0.56*****	0.000		
Driver's License		**−1.73*****	0.000	**−1.42*****	0.000	**−2.10*****	0.000		
Frequent Transit User		**1.48*****	0.000	**1.56*****	0.000	**2.50*****	0.000		
Average Traffic Stress		**−0.96***	0.028	**−5.02*****	0.000	0.66	0.384		
Proportion of Bike Lanes		**0.42****	0.008	**0.43ˆ**	0.091	0.35	0.113		
Average Slope		**−11.2***	0.043	**−44.7*****	0.000	**−37.2*****	0.000		

The descriptive statistics of the sample are in Table A3 in the Appendix – the variables other than the bike-friendliness variables will not be discussed here.

The O-D coordinates of the 118,423 trips were used as inputs for the bike route simulation from which we generated the data for the bike-friendliness variables. Table 5 shows the descriptive statistics of the bike-friendliness variables. The modes in each column indicate the actual trip modes in the travel surveys, while the values in each row are from the bike route simulation. It means, as for the trips that are not made by bike, the values are based on a counterfactual scenario: what if the trips were made by bike instead of their actual modes?

Average traffic stress is lowest in biking routes of actual bike trips and highest in biking routes (simulated) of actual driving trips. The gap in the traffic stress between modes is large in Atlanta while it does not seem significant in the other two cities. In Atlanta and the Twin Cities, the proportion of bike lanes is the highest in actual bike trip routes. However, in Chicago, simulated bike routes of transit trips show the highest proportion of bike lanes. The average slope does not seem to have a significant impact on mode choice in any city.

5.3. Result of mode choice models

Tables 6 and 7 show the results of the multinomial logistic regression of the total population model and the low-income population model respectively. McFadden's R^2 values of the two models are 0.44 and 0.45 respectively, which are reasonably high goodness-of-fit statistics for discrete choice models. The available modes in the model are walking, biking, public transit, and private vehicle with the private vehicle being the base mode. Thus, the coefficients and p-values in each mode column represent the variables' impact on the probability of choosing each mode over driving.

As shown in Table 6, the coefficients for all variables are mostly in line with expectations. 'Travel time' shows the highest absolute coefficient value in the walking trip and is followed by bike, drive, and public transit trips. This result indicates that travel time is a constraint affecting walking trips most and the public transit trip least, which follows our expectations. The coefficients of the 'distance to the nearest subway station' show that increasing distance from a subway station is associated with a lower chance of choosing walking and bike trips as well as transit trips. Coefficients and p-values of the 'feels-like temperature at departure time' and 'weather of the day' imply that those two are crucial

Table 8
Mode Share Prediction by Bike Lane Extension Scenarios: 20% to 80%; whole Sample.

Mode	City	Proportion of Bike Lanes in the Route	Non-frequent transit user sample (N = 97,219) Mode share	Δ(20→80%)	Frequent transit user sample (N = 21,204) Mode share	Δ(20→80%)
Walking	Atlanta	20%	12.9%	1.7%	34.3%	1.4%
		80%	14.6%		35.7%	
	Chicago	20%	16.1%	3.7%	45.3%	2.6%
		80%	17.3%		47.9%	
	Twin Cities	20%	18.5%	1.3%	45.7%	1.2%
		80%	19.8%		46.9%	
Biking	Atlanta	20%	0.7%	1.3%	2.3%	3.2%
		80%	2.0%		5.6%	
	Chicago	20%	2.0%	0.5%	5.0%	0.2%
		80%	2.5%		5.2%	
	Twin Cities	20%	2.0%	0.6%	5.6%	1.3%
		80%	0.6%		7.0%	
Public Transit	Atlanta	20%	0.2%	0.2%	8.6%	2.0%
		80%	0.4%		10.6%	
	Chicago	20%	2.1%	1.2%	17.0%	6.2%
		80%	3.3%		23.2%	
	Twin Cities	20%	0.6%	0.0%	11.1%	0.1%
		80%	0.6%		11.2%	
Non-automobile (Walk + Bike + Transit)	Atlanta	20%	13.9%	3.1%	45.2%	6.6%
		80%	17.0%		51.8%	
	Chicago	20%	20.2%	5.5%	67.3%	9.1%
		80%	25.6%		76.4%	
	Twin Cities	20%	21.0%	1.9%	62.4%	2.6%
		80%	22.9%		65.1%	

conditions for active mobility, particularly for the bike trip, while public transit trips are not significantly affected by weather conditions. The variable 'purpose' also emerges as a significant factor for mode choice: people are more likely to walk or bike for social/recreational purpose trips while public transit is preferred for commuting. The variable 'region' shows that, compared to Atlanta, people in Chicago and the Twin Cities are less likely to walk, but more likely to bike and use public transit with Chicago showing a considerably higher coefficient in public transit. 'Intra-city trip' shows that people who are traveling within a city area are much more likely to use a mode other than driving compared to those from/to outside the city.

Socio-demographic variables are all significant and intuitive as well. The coefficient of 'age' implies that being too young or old can be a constraint for bike trips while being old is not negatively affecting walking and public transit trips. The variable "gender" indicates that women are less likely to choose any of the three modes, particularly biking, than men. In terms of race, compared to White population, people of any other race are less likely to walk or bike, but more likely to use public transit, which we think is largely influenced by the neighborhood environment. It is not a surprise that White residents are less likely to take public transit than any other racial group given that in most U.S. cities, neighborhoods close to the city center (where public transit is well supplied) have a relatively lower proportion of White households compared to racial minorities. In addition, White residents are more likely to walk or bike because their neighborhoods have better community infrastructure, thereby encouraging more physical activity (Kelly, Schootman, Baker, Barnidge & Lemes, 2007).

Having a disability is correlated negatively with walking trips and positively with public transit trips. Having a driver's license is a strong factor for driving trips. Our results also indicate that travelers' household income and mode choices are closely related. For public transit users, there is a clear negative relationship between income and the probability of public transit usage. This finding reinforces earlier studies showing that as income rises, individuals are less likely to take public transit and more likely to use a private vehicle for travel. For walking and biking, on the other hand, the relationship with income is U-shaped: individuals with household incomes between $50,000 and $75,000 are least likely to walk or bike; this probability increases as the income either increases or decreases. Low-income populations are more likely *captive users* of walking/biking modes for transport purposes (Murakami & Young, 1999). On the other hand, high-income populations are *choice users* of walking/biking mode and seek out more walkable and bike-able neighborhoods with better mental/physical quality of life (Sallis et al., 2009). The coefficient of 'frequent transit user' indicates that people who regularly take public transit are more likely to walk and bike.

Finally, the bike-friendliness variables show interesting results. As mentioned earlier, values of the three bike-friendliness variables are based on bike route simulations for all O-D pairs of each respondent regardless of their actual modes. Therefore, in the case of bike trips, we consider their actual routes and values but in the case of non-bike trips, they are hypothetical values based on simulations designed to obtain counterfactual insights. The variable 'average traffic stress' shows that when the bike route has high traffic stress, people are less likely to walk or bike, but more likely to take public transit. Note that the 'proportion of bike lanes' is segmented by the three cities to differentiate the impacts (and the predictions in the following section) by cities. Although the coefficients vary by city, the significant positive values support the hypothesis of this study. That is, <u>the proportion of bike laneswould be positively associated with the probability of biking trips as well as walking trips and, in some cases, public transit trips</u>. The coefficients for walking and biking are significant in all three cities and the coefficients for public transit are only significant in Chicago.

Table 7 presents the result of the model for the low-income population segment only. In general, the coefficients suggest that this group of the population is not significantly different from the whole population. Below we discuss the variables showing noteworthy differences.

First, 'feels-like temperature at departure time' is less significant in the low-income population model and their signs are different from the same variable in the full-population model. We can surmise that low temperatures make people reluctant to walk or bike. In the whole population model, it is statistically significant that people start to walk less when the temperature goes below 50°F (10°C) and bike less when the temperature is below 60°F (16°C). However, in the low-income population model, the likelihood of walking or riding a bike at a temperature between 60°F (16°C) and 69°F (21°C) does not change statistically significantly until it reaches 40°F (4°C). In addition, the 'weather of the day' variable shows a similar difference: while the whole population model indicates that people are significantly less likely to walk or bike on a rainy/snowy day, the low-income population model says that the probability of walking or biking on a rainy/snowy day is not statistically

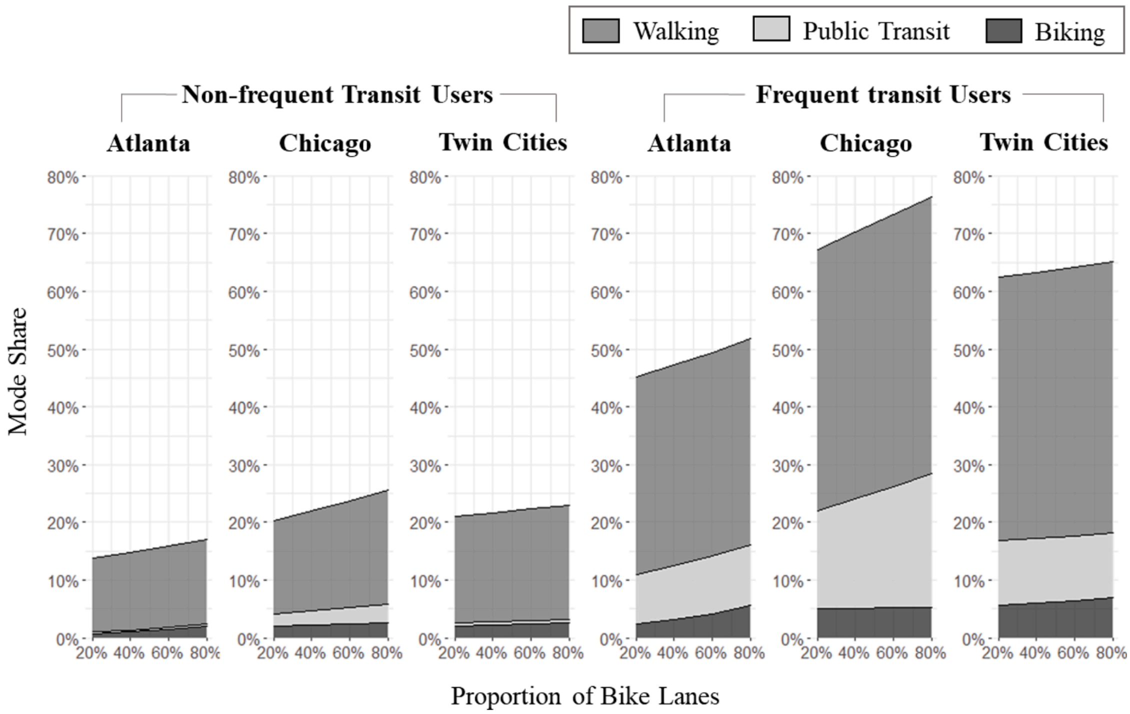

Fig. 3. Mode Share Prediction by Bike Lane Extension Scenarios: 20%, 40%, 60%, and 80%.

different from that on a cloudy day. Based on those results, we can presume that the low-income population is likely to have limited options for mode choice or to avoid travel, compared to people in other income groups.

Second, 'frequent transit user' in the low-income population model shows the same signs as the whole population model, albeit with larger coefficients for walking and biking trips. That is, frequent transit users in the low-income population are much more likely to have walking or biking trips in addition to public transit trips. It implies the possibility that they are more open to making multi-modal trips between active mobility and public transportation.

Lastly but most importantly, the bike-friendliness variables in the low-income population model show that walking and biking trips are significantly affected by the average traffic stress and the proportion of bike lanes,[5] but public transit trips are not strongly affected by either. Also, the influence of the average slope is much greater for all three modes compared to the whole population model.

5.4. Mode share prediction by bike lane investment scenario

Coefficients from the multinomial logistic regression are based on a logit transformation, which makes their interpretability less intuitive than, say, ordinary least square regression results. One way to directly compare coefficients from the mode choice model is to predict a mode share. Here, an average probability of choosing a certain mode can be considered equivalent to the mode share. By testing various values of a particular variable, we can examine how much impact the variable has on the mode share.

Table 8 and Fig. 3 are the mode share prediction results based on the whole population model results shown in Table 6. The predicted mode shares in Table 8 are based on a scenario in which the proportion of bike lanes increases from 20% (which is close to the current average) to 80%.[6] The non-frequent transit user sample column shows the prediction result based on the 97,219 observations who reported that they use transit less than once a week and the frequent transit user sample column is based on the 21,204 observations who use transit once a week or more. The prediction was done separately by each city using the segmented coefficients of the variable 'proportion of bike lanes.'

First, the non-frequent transit user sample column shows that, if the proportion of bike lanes increases to 80%—in other words, if it quadruples—, the bike mode share would increase by 1.3% in Atlanta, 0.5% in Chicago, and 0.6% in the Twin Cities. The increases in the walking mode share are also noticeable: 1.7% in Atlanta, 3.7% in Chicago, and 1.3% in the Twin Cities, which are greater than the increases in the biking mode share. The non-automobile mode share in total would increase by 3.1% in Atlanta, 5.5% in Chicago, and 1.9% in the Twin Cities.

The impact is more evident in the frequent transit user sample whose current non-automobile mode share is roughly three times higher than the non-frequent transit user sample in all three cities. Accordingly, the increase in the mode share is also substantially higher. For example, if the proportion of bike lanes quadruples, the bike mode share will increase by 3.2% in Atlanta, 0.2% in Chicago, and 1.3% in the Twin Cities. For Atlanta and the Twin Cities, the predicted increase in the bike mode share of the frequent transit user sample was more than twice greater

5 Contrary to the whole sample model, the "Proportion of bike lanes" in the low-income sample model was not segmented by the three cities for two reasons: 1) the low-income sample is not large enough to get a significant result from the segmented model and 2) the point of running the low-income model is to see how the overall trend differs from the whole sample model but not necessarily by each city.

6 It is difficult to estimate how much bike lanes should be provided to achieve 80% of the proportion of bike lanes on average because it would vary depending on how the bike network is designed. One reasonable scenario for covering most part of bike trip routes by bike lanes would be to provide all secondary and tertiary roads with dedicated bike lanes. Secondary and tertiary roads are where most of bike trips are happening and most of dedicated bike lanes exist: 76.3% of the total length of dedicated bike lanes in Atlanta, 85.9% in Chicago, and 90.9% in the Twin Cities are in either secondary or tertiary roads. Thus, it is plausible that, if all the secondary and tertiary roads are covered by dedicated bike lanes, the proportion of bike lanes in a route would be above 80%. The length of new bike lanes required for that scenario would be 348 miles in Atlanta, 888 miles in Chicago, and 570 miles in the Twin Cities.

Table 9
Mode Share Prediction by Bike Lane Extension Scenarios: 20% to 80%; low-income sample.

Mode	Proportion of Bike Lanes in the Route	Low-income, non-frequent transit user sample (N = 5580)		Low-income, frequent transit user sample(N = 2497)	
		Mode Share	Δ(20→80%)	Mode Share	Δ(20→80%)
Walking	20%	24.2%	2.6%	48.2%	1.9%
	80%	26.8%		50.1%	
Biking	20%	3.4%	0.3%	8.0%	0.5%
	80%	3.7%		8.5%	
Public Transit	20%	7.8%	0.5%	24.2%	1.1%
	80%	8.3%		25.3%	
Non-automobile	20%	35.3%	3.4%	80.5%	3.5%
(Walk + Bike + Transit)	80%	38.7%		84.0%	

Table A1
List of Acronyms.

Acronym	Definition
DOT	Department of transportation
SPLOST	Special-purpose local-option sales tax
O-D	Origin and destination
MRS	Marginal rate of substitution
LTS	Level of traffic stress
NACTO	National association of city transportation officials
NHTS	National household travel survey

than that of the non-frequent transit user sample. The total non-automobile mode share would increase by 6.6% in Atlanta, 9.1% in Chicago, and 2.6% in the Twin Cities, which is a considerably greater increase compared to the non-frequent transit user sample.

Table 9 is the mode share prediction result based on the low-income population model in Table 7. Unlike Table 8, the prediction results in Table 9 are not segmented by cities as the variable 'proportion of bike lanes' in the low-income population model in Table 7 was not segmented. A noticeable difference between Table 8 and Table 9 is that the low-income sample (either frequent or non-frequent transit users) shows much higher current mode shares for all non-automobile modes. For example, the bike mode share of non-frequent transit users is 3.4% in the low-income sample while it is between 0.7% and 2.0% in the whole sample. The difference is more apparent in the frequent transit user group: 8.0% in the low-income sample and 2.3 - 5.6% in the whole sample.

In general, the increases in mode shares in the low-income sample are similar to those of the whole sample. Bike mode share in the low-income sample shows a bit smaller increase than the whole sample: if the proportion of bike lanes quadruples, the bike mode share would increase by only 0.3% in the low-income, non-frequent transit user sample and 0.5% in the low-income, frequent transit user sample.

The differences in the mode share change between frequent transit users and non-frequent transit users in the low-income sample are not as great as they are in the whole sample: the non-automobile mode share increases by 3.4% in the non-frequent transit user sample and 3.5% in the frequent transit user sample.

Taken together, we can draw three conclusions from these results. First, providing more bike lanes does affect the mode choice of all non-automobile modes. Second, frequent transit users are considerably more affected by bike lane network improvement in terms of higher shares of non-automobile modes compared to non-frequent transit users. Third, income does not make a significant difference in terms of the impact of bike lanes on mode choice; although the low-income population is less

Table A2
Descriptive Statistics of Six Other Cities that are Similar to Atlanta According to the Cluster Analysis.

City(County)		Charlotte (Mecklenburg)	Baltimore (Baltimore)	Pittsburgh (Allegheny)	St. Louis(St. Louis)	Buffalo (Erie)	Cleveland (Cuyahoga)
Population		1.11 M	1.42 M	1.22 M	1.29 M	0.92 M	1.24 M
Employment		0.96 M	0.98 M	0.93 M	1.11 M	0.59 M	0.95 M
Urbanized area (mi^2)		468	356	538	447	332	417
Population Density		2,374	3,991	2,262	2,900	2,763	2,960
Employment Density		2,045	2,748	1,732	2,486	1,771	2,287
Employment / Population		0.86	0.69	0.77	0.86	0.64	0.77
Per Capita Income ($)		62,890	60,730	65,784	69,480	53,498	56,502
Commuting Mode Share (%)	Private Vehicle	85.8	80.4	79.5	87.4	89.3	87.2
	Public Transit	2.9	9.8	9.5	3.9	3.7	4.6
	Walking	1.9	3.5	4.1	2.2	2.6	2.7
	Biking	0.1	0.4	0.6	0.3	0.4	0.4
	Other Modes	9.3	6.0	6.3	6.3	4.0	5.1
Mean Commuting Time (minute)		26.4	30.4	27.2	24.4	21.5	24.3
Infrastructure Density (mile/urbanized area (mi^2))	Highway	0.61	0.95	0.63	0.77	0.87	0.78
	Primary & Secondary Road	1.59	2.55	1.48	1.55	2.08	1.60
	Residential Road	7.76	8.77	8.13	10.47	9.10	7.76
	Railway	0.11	0.25	0.11	0.14	0.04	0.19
	Bike Lane	0.14	0.04	0.06	0.17	0.16	0.08
Length of Bike Lanes (mile)		63.5	13.5	31.4	76.9	54.9	34.5
Gas Price ($)		2	2	3	2	2	2
Bikeability Score		12	24	21	32	6	23
5-year Precipitation (inch)		245	255	240	219	215	207
Average Temperature		62°F (17 °C)	56°F (13 °C)	53°F (12 °C)	57°F (14 °C)	49°F (9 °C)	52°F (11 °C)
Terrain (Std. Dev. in meters)		10.4	16.3	32.8	9.2	8.0	16.0

Table A3
Descriptive Statistics of Variables by Type of Sample and Mode.

Category		Whole sample (N = 118,423)				Low-income sample (N = 8,077)			
		Walk	Bike	Transit	Drive	Walk	Bike	Transit	Drive
	Travel time (minute)	10.8	7.8	16.2	5.6	10.8	6.4	15.9	5.7
	Distance to the nearest subway station (mile)	1.58	1.26	0.66	3.43	1.26	1.12	0.77	2.83
	Weekend/Weekday (% weekday)	84.8%	84.0%	93.3%	82.5%	88.9%	87.6%	93.7%	85.9%
Feels-like temperature at departure time	Below 30°F (~ −1 °C)	26.0%	9.5%	35.2%	38.6%	29.4%	7.7%	36.5%	29.5%
	30~39°F (−1~4 °C)	10.0%	7.0%	12.1%	11.5%	10.8%	9.5%	16.7%	12.2%
	40~49°F (4~9 °C)	7.1%	6.1%	7.2%	6.6%	7.9%	8.2%	7.8%	6.3%
	50~59°F (10~15 °C)	11.9%	13.6%	10.9%	9.7%	13.2%	14.2%	10.6%	11.5%
	60~69°F (16~21 °C)	18.5%	24.4%	14.7%	13.4%	14.2%	20.9%	11.9%	16.1%
	70~79°F (21~26 °C)	20.0%	28.2%	14.3%	14.2%	17.6%	26.3%	11.6%	15.5%
	80~89°F (27~32 °C)	5.4%	9.3%	4.8%	4.8%	6.1%	10.3%	4.7%	6.7%
	Above 90°F (32 °C~)	1.2%	1.9%	0.7%	1.1%	0.8%	2.8%	0.3%	2.0%
Weather of the day	Sunny	17.1%	21.6%	13.5%	13.9%	16.3%	23.5%	10.0%	12.2%
	Cloudy	49.5%	46.2%	49.7%	52.7%	49.4%	42.0%	51.8%	53.4%
	Rainy	26.8%	30.2%	23.6%	27.2%	26.7%	33.0%	28.7%	26.7%
	Snowy	6.6%	2.0%	9.7%	9.7%	7.7%	1.5%	9.5%	7.8%
Purpose	Home	26.8%	34.1%	31.5%	31.4%	30.9%	30.4%	34.0%	31.8%
	Work	14.9%	16.7%	20.8%	8.1%	9.1%	13.7%	7.3%	5.6%
	School	3.7%	4.7%	5.5%	2.7%	8.7%	9.5%	8.6%	2.8%
	Social/recreational	27.3%	18.3%	9.2%	10.0%	13.5%	11.9%	14.9%	17.6%
	Shopping	6.9%	8.2%	10.2%	15.3%	4.4%	2.8%	6.5%	9.4%
	Errand	3.1%	3.7%	3.7%	8.4%	18.4%	16.5%	9.4%	9.6%
	Meal	10.6%	6.0%	6.0%	9.9%	7.5%	5.7%	6.6%	9.1%
	Drop-off/pick-up	2.6%	2.2%	2.3%	9.2%	2.6%	1.0%	2.4%	9.1%
	Other purposes	4.1%	6.0%	10.8%	5.1%	4.8%	8.5%	10.4%	5.1%
Intra-city trip	71.4%	76.3%	90.6%	26.9%	74.4%	78.9%	89.8%	39.5%	
Age	Under 15	9.1%	8.4%	5.0%	13.9%	8.2%	2.8%	6.6%	11.3%
	16–24	8.6%	8.4%	13.9%	5.0%	23.3%	25.8%	20.7%	10.8%
	25–34	28.5%	33.2%	31.5%	16.6%	27.0%	25.0%	17.6%	18.4%
	35–44	19.1%	23.4%	17.3%	21.0%	11.5%	10.8%	11.3%	12.1%
	45–54	12.7%	8.7%	13.2%	14.6%	9.5%	14.4%	18.0%	11.4%
	55–64	13.3%	10.9%	11.8%	15.4%	11.7%	14.2%	14.7%	18.0%
	65 or older	8.7%	7.0%	7.3%	13.5%	8.8%	7.0%	11.2%	18.0%
Gender (% of female)	51.8%	42.5%	54.1%	55.7%	57.1%	45.9%	60.6%	64.5%	
Race	Asian	4.7%	4.8%	6.0%	4.7%	6.2%	4.9%	4.4%	5.1%
	Black	5.7%	4.0%	20.9%	5.5%	23.0%	11.1%	44.7%	23.7%
	Other Races	5.2%	5.8%	9.0%	4.9%	11.9%	6.4%	11.6%	7.7%
	White	84.5%	85.4%	64.1%	84.9%	59.0%	77.6%	39.2%	63.5%
Disability (% of disabled)		2.1%	2.4%	8.7%	2.0%	8.0%	7.7%	20.0%	9.3%
Driver's license (% of licensee)		84.3%	85.9%	71.9%	84.7%	60.2%	74.2%	42.7%	81.7%
Household income	Less than $25,000	9.6%	12.7%	27.6%	4.8%				
	$25,000–50,000	11.0%	12.1%	19.4%	11.0%				
	$50,000–75,000	15.1%	13.0%	15.9%	15.7%				
	$75,000–100,000	15.9%	15.6%	11.8%	17.0%				
	$100,000–150,000	24.8%	25.2%	15.1%	27.6%				
	$150,000 or more	23.5%	21.5%	10.2%	23.8%				
Frequent transit user		36.4%	37.4%	74.3%	9.0%	47.4%	51.8%	58.4%	11.8%
Average traffic stress		15%	14%	16%	18%	16%	14%	16%	18%
Proportion of bike lanes		16.3%	23.2%	21.0%	15.7%	14.8%	21.4%	16.4%	15.1%
Average slope		1.17%	1.1%	1.0%	1.16%	1.13%	1.0%	0.9%	1.23%
Number of trips		26,312 (22.2%)	3,067 (2.6%)	3,757 (3.2%)	85,287 (72%)	2,536 (31.4%)	388 (4.8%)	1,037 (12.8%)	4,116 (51%)

sensitive to the effect of the bike lanes on travel mode choices compared to the whole population.

6. Conclusions

Biking is a notably underutilized mode of transportation in the majority of U.S. cities. Many local governments are spending a sizeable amount of their budget on improving the bike lanes to ensure a safer travel environment for both pedestrians and bike users and to offer better access to public transit. Ultimately, the bike infrastructure is expected to reduce car-centricity and bring other benefits to the economy and the environment. Yet, previous studies have not adequately assessed what impact the bike infrastructure has on mode choice. This study sought to understand that relationship at an individual route level based on simulated routes. The analysis was conducted for three cities—Atlanta, Chicago, and the Twin Cities. Altogether 118,423 individual trips from the travel surveys of the three cities were used as the sample for simulation and the mode choice model.

The simulation result showed that bike routes of actual bike trips have lower average traffic stress and a higher proportion of bike lanes compared to simulated bike routes of trips made using other modes. The multinomial logistic regression results based on the entire 118,423 trips confirmed that all non-automobile modes—walk, bike, and public transit—are found to be positively affected by bike-friendliness variables: the average traffic stress, the proportion of bike lanes, and the average slope. Another multinomial regression model based on the low-income population sample also demonstrated significant correlations between the bike-friendliness of routes and mode choice.

The analyses gave us three valuable insights. First, investing in bike lanes does have a positive impact on promoting non-automobile modes. Although the magnitude varies by city, the proportion of bike lanes is found to be significantly associated with walking and biking in all three cities. If bike lanes cover 80% of an average route, non-automobile mode share is predicted to increase by 2–5% for non-frequent transit users. The impact on bike trips is, however, relatively small – the predicted increase in the bike mode share is 0.5–1.3% for non-frequent transit users and 0.2–3.2% for frequent transit users. Considering that achieving 80% of bike lanes on average would require hundreds of miles of new bike lanes, the effect is not very promising. In fact, it tells us that providing bike lanes alone cannot bring an appreciable effect on promoting bike usage. Therefore, the investment in bike lanes should be accompanied by other types of effort, such as providing bike parking facilities (particularly in transit hubs), introducing a bike-sharing system and making public bike stations, organizing public bike events or campaigns, and creating bike education programs.

Second, the non-automobile mode choice of the low-income population is as significantly affected by the bike lanes as the general population. Also, the magnitudes are similar. Based on model predictions, if the proportion of bike lanes quadruples (from the current 20% to 80%), non-automobile mode share will increase by around 2–5% in the general population and 3% in the low-income population. Therefore, the low-income population who are residing in historically underserved areas with low transportation services can greatly benefit from the new bike lanes, which can ultimately help reduce social inequities.

Third, the mode shares of the non-automobile modes are positively associated with each other. Whether a respondent is a frequent transit user affects not only the probability of making public transit trips but also the probability of making walking/biking trips. The impact of providing a better biking environment is also evident for all non-automobile modes. The prediction results of the whole population model demonstrate that the increases in non-automobile mode share are significantly higher among frequent transit users (2.6–9.1%) compared to non-frequent transit users (1.9–5.5%). Therefore, investing in bike lanes would benefit not just bike users but all active travelers and public transit users.

This study adds to the literature by empirically demonstrating that, although the impact of bike lanes on promoting bike usage may not seem great, their impact on the non-automobile modes as a whole would be significant because walking, biking, and public transit are positively associated with each other. This study also introduces methodological novelty by demonstrating the validity of estimations using a mode choice model based on simulated routes. The individual route-level analysis lets us understand how the linear infrastructure influences our travel experience and the mode we choose, which was not attainable in previous studies.

Making people drive less is the utmost goal in the major U.S. cities to make themselves economically and environmentally sustainable. The model results and predictions in this study provide valuable information on how improving the bike lane network can serve the goal. Transportation planners should carefully consider the potential linkages between active modes of travel and public transit to correctly estimate the efficiency of implementing bike lanes, which can ultimately contribute to reducing auto-dependency and creating a sustainable city.

7. Limitations

The research framework and methods in this study have multiple limitations. First, the bike routes of each trip are based on simulations, which may differ from the real route choice. Second, although we believe that our method is more elaborate than any previous study, even individual route-level analysis cannot guarantee the direction of causality by the nature of the cross-sectional aspect of this study. Third, the sample size between the three cities is unbalanced. Among the 118,423 observations, only 4% are from Atlanta while more than 70% are from the Twin Cities. This large variation in sample size is partly the reason why we segment the data by each city for the variable 'proportion of bike lanes.' Otherwise, the coefficients would largely be influenced by the Twin Cities data. Lastly, the model did not consider other planning factors that may affect the bike mode choice other than bike lanes. We acknowledge that future studies are necessary to determine how much synergetic effect can be achieved from various types of physical/non-physical strategies for encouraging biking such as public bike sharing.

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.

Subhrajit Guhathakurta reports financial support was provided by National Science Foundation.

Data Availability

The authors do not have permission to share data.

Acknowledgement

This research is sponsored by National Science Foundation (grant number CMMI-1854684).

Appendix

Table A1, A2 and A3

References

Akar, G., & Clifton, K. J. (2009). Influence of individual perceptions and bicycle infrastructure on decision to bike. *Transportation research record, 2140*(1), 165–172.

Aldred, R., & Dales, J. (2017). Diversifying and normalising cycling in London, UK: An exploratory study on the influence of infrastructure. *Journal of Transport & Health, 4*, 348–362.

Aziz, H. M., Nagle, N. N., Morton, A. M., Hilliard, M. R., White, D. A., & Stewart, R. N. (2018). Exploring the impact of walk–bike infrastructure, safety perception, and built-environment on active transportation mode choice: A random parameter model using New York City commuter data. *Transportation, 45*(5), 1207–1229.

Bearn, C., Mingus, C., & Watkins, K. (2018). An adaption of the level of traffic stress based on evidence from the literature and widely available data. *Research in Transportation Business & Management, 29*, 50–62.

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems,, 65*, 126–139.

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice, 46*(10), 1730–1740.

Buehler, R., & Pucher, J. (2012). Cycling to work in 90 large American cities: New evidence on the role of bike paths and lanes. *Transportation, 39*(2), 409–432.

City of Atlanta, (2018). Atlanta's transportation plan. Atlanta, Georgia.

City of Atlanta, (2019). One atlanta: Strategic transportation plan. Atlanta, Georgia.

Clark, C., Mokhtarian, P., Circella, G., & Watkins, K. (2019). User preferences for bicycle infrastructure in communities with emerging cycling cultures. *Transportation research record, 2673*(12), 89–102.

Croissant, Y. (2012). Estimation of multinomial logit models in R: The mlogit Packages. R package version 0.2-2. URL: Http://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf.

Dill, J., & Voros, K. (2007). Factors affecting bicycling demand: Initial survey findings from the Portland, Oregon, region. *Transportation Research Record, 2031*(1), 9–17.

Garrard, J., Rose, G., & Lo, S. K. (2008). Promoting transportation cycling for women: The role of bicycle infrastructure. *Preventive medicine, 46*(1), 55–59.

Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco. *California. Transportation letters, 3*(1), 63–75.

Hull, A., & O'Holleran, C. (2014). Bicycle infrastructure: Can good design encourage cycling? *Urban, Planning and Transport Research, 2*(1), 369–406.

Jones, T. (2012). Getting the British back on bicycles—The effects of urban traffic-free paths on everyday cycling. *Transport Policy, 20*, 138–149.

Kelly, C. M., Schootman, M., Baker, E. A., Barnidge, E. K., & Lemes, A. (2007). The association of sidewalk walkability and physical disorder with area-level race and poverty. *Journal of Epidemiology & Community Health, 61*(11), 978–983.

Krizek, K. J., Barnes, G., & Thompson, K. (2009). Analyzing the effect of bicycle facilities on commute mode share over time. *Journal of urban planning and development, 135*(2), 66–73.

Lowry, M. B., Furth, P., & Hadden-Loh, T. (2016). Prioritizing new bicycle facilities to improve low-stress network connectivity. *Transportation Research Part A: Policy and Practice, 86*, 124–140.

Marshall, W. E., & Ferenchak, N. N. (2019). Why cities with high bicycling rates are safer for all road users. *Journal of Transport & Health, 13*, 285–301, 100539.

Martens, K. (2004). The bicycle as a feedering mode: Experiences from three European countries. *Transportation Research Part D: Transport and Environment, 9*(4), 281–294.

Moudon, A. V., Lee, C., Cheadle, A. D., Collier, C. W., Johnson, D., Schmid, T. L., et al. (2005). Cycling and the built environment, a US perspective. *Transportation Research Part D: Transport and Environment, 10*(3), 245–261.

Murakami, E., & Young, J. (1999). Daily Travel by Persons with Low Income. In *Report No. FHWA-PL-99-003*. Silver Spring, MD: Federal Highway Administration.

National Association of City Transportation Officials. (2016). Equitable Bike Share Means Building Better Places for People to Ride (Nacto Bike Share Equity Practitioners' Paper #3). URL: Https://nacto.org/wp-content/uploads/2016/07/NACTO_Equitable_Bikeshare_Means_Bike_Lanes.pdf.

Pedroso, F. E., Angriman, F., Bellows, A. L., & Taylor, K. (2016). Bicycle use and cyclist safety following Boston's bicycle infrastructure expansion, 2009–2012. *American journal of public health, 106*(12), 2171–2177.

Rodriguez-Valencia, A., Rosas-Satizábal, D., Gordo, D., & Ochoa, A. (2019). Impact of household proximity to the cycling network on bicycle ridership: The case of Bogotá. *Journal of Transport geography, 79*, Article 102480.

Sallis, J. F., Saelens, B. E., Frank, L. D., Conway, T. L., Slymen, D. J., Cain, K. L., et al. (2009). Neighborhood built environment and income: Examining multiple health outcomes. *Social science & medicine, 68*(7), 1285–1293.

Schoner, J. E., & Levinson, D. M. (2014). The missing link: Bicycle infrastructure networks and ridership in 74 US cities. *Transportation, 41*(6), 1187–1204.

Xu, S. J., & Chow, J. Y. (2020). A longitudinal study of bike infrastructure impact on bikesharing system performance in New York City. *International journal of sustainable transportation, 14*(11), 886–902.

Zahabi, S. A. H., Chang, A., Miranda-Moreno, L. F., & Patterson, Z. (2016). Exploring the link between the neighborhood typologies, bicycle infrastructure and commuting cycling over time and the potential impact on commuter GHG emissions. *Transportation research part D: Transport and environment, 47*, 89–103.