Issue 24(4), 2024, pp. 89-110
ISSN: 1567-7141
TU Delft OPEN

EJTIR

Research paper
https://ejtir.tudelft.nl
DOI: 10.59490/ejtir.2024.24.4.6965

The role of the (e-)bike: a mode choice model for short distances

Chantal Huurman[1], Adam Pel[2], Winnie Daamen[3], Kees Maat[4]

[1] corresponding author ch.huurman@gmail.com, Witteveen+Bos, The Netherlands; https://orcid.org/0009-0009-1969-6822
[2] Department of Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, The Netherlands; https://orcid.org/0000-0003-3754-5779
[3] Department of Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, The Netherlands; https://orcid.org/0000-0003-4518-4184
[4] Department of Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, The Netherlands; https://orcid.org/0000-0002-7832-7017

Keywords
Mode choice
E-bike
Logit models
Urban areas

Publishing history
Submitted: 30 March 2023
Revised date(s): 23 October 2023, 30 June 2024,
Accepted: 11 October 2024
Published: 28 December 2024

Cite as
Huurman, C.H., Pel, A.J., Daamen, W., & Maat, K. (2024). The role of the (e-)bike: a mode choice model for short distances. *European Journal of Transport and Infrastructure Research, 24*(4), 89-110.

Abstract

The bicycle is a very important mode for travel in various countries, particularly in the Netherlands. However, it is in practice often modelled with less detail than other urban modes, such as the car and public transport. Moreover, the increasing use of e-bikes and the differences with conventional bikes show that more research into this transport mode is needed. E-bikes require less physical effort and allow higher speeds, making the e-bike suitable for longer distances. The goals of this research are to (1) create a mode choice model that predicts an accurate modal split for urban areas in the Netherlands and this model is used to (2) find significant factors that influence the modal split, in order to support municipalities of Dutch urban areas to stimulate the use of the (e-)bike. Within both goals, potential differences between conventional bikes and e-bikes are considered. A conceptual model, following from the literature, describes the assumed modal choice including factors relevant to cycling. Data was used mainly from the Dutch National Travel Survey (ODiN). Discrete choice models, a multinomial logit and a nested logit, are estimated to identify significant influencing factors. Results show that a nested logit model is the most explanatory one compared to the other models, with a rho-square-bar of 0.469. The model includes 15 main variables, 3 quadratic components and 4 interaction effects. The nested structure is formed by a correlation between the bike and the e-bike. The factors that show to be generally highly influential for the bike and the e-bike are the travel distance, owning a driver's license and street density. The model is practically applicable for municipalities to form expectations in the modal shift for changes in their networks or policies. However, modelling these changes has not been validated and thus needs further research.

1 Introduction

Cycling is increasingly viewed as an ideal mode of transport for urban travel. Compared with cars, bicycles are more sustainable, consume less space, and contribute less to traffic congestion. Moreover, with the rapid rise in electric bicycles, the share of cycling is increasing. The substitution of car traffic for bicycle traffic has many societal benefits and is, therefore, encouraged by municipalities.

Municipalities apply traffic models to gain insight into traffic flows and the effects of travel-related measures on these flows. However, compared to cars and public transport, objective ways to substantiate bicycle policies are lacking. Proper prediction of bicycle traffic flows can be more challenging than that of car traffic flows. Although much research has been conducted on the determinants of bicycle use (see references in Section 2), this does not alter the fact that, to the best of our knowledge, this expertise is not used in traffic models with cycling as a travel mode. One reason for this could be that, until one or two decades ago, cycling caused almost no congestion anywhere and was therefore hardly seen as a planning problem. Today, however, this is rapidly changing, especially in bigger cities, student cities, and recreational areas, where bicycle traffic often exceeds infrastructure capacity.

To accurately predict bicycle traffic flows, the prediction of bicycle traffic share must be sufficient. There are several important aspects to consider in the prediction of the bicycle travel share. First, cycling competes with both cars and (urban) public transport over short and medium distances. Second, the factors affecting bicycle trips differ from the factors affecting cars, so the traditional mode choice model cannot be applied. Moreover, behavior and influential factors also differ between conventional bicycles and e-bikes, so a model reflecting bike and e-bike choice as one mode might not be sufficient.

Further developments of traffic models require correct estimation of the modal split between all modes, that is, bicycle, e-bike, urban public transport (bus, tram, metro), train, and car, particularly for short and medium distances where these modes of transport compete (from a travel behavior perspective) and where municipalities aim to stimulate active modes (from a policy perspective).

This paper describes the estimation of a choice model based on traditional factors, as well as factors that, according to the literature, specifically affect conventional and electric bicycle traffic. The study area is urban areas in the Netherlands. In these regions, bicycle use is high and continues to increase. As of 2016, there were around 23 million bicycles in the country, and a quarter of the trips were made using a bicycle (KiM, 2018). The Netherlands is the only country in the world in which people have more than one bicycle per capita (Bicycle Dutch, 2018). Urban areas are defined as postcode areas where the address density is higher than 1000 addresses per square kilometer. This is similar to urbanity levels 3, 4, and 5 of Statistics Netherlands (CBS, 2022c).

The contributions of this study are threefold. First, we construct a conceptual (e-)bike mode choice model, where we use earlier studies to identify 36 possible driving factors as well as possible interactions, and hierarchically structure these into three main categories and six subcategories. Second, we refine the (e-)bike choice model specifically for mode choice for short and medium distances in the Netherlands for an average day, which is applicable for municipalities to estimate their (e-)bicycle traffic share. Here, we used revealed preference data to analyze the correlations among the driving factors and between mode choice and each factor, as well as to show the presence of interaction effects and quadratic components in these factors. Third, we estimate the (refined) mode choice model with the modes car, BTM (Bus, Tram, Metro), train, bike, and e-bike, thus yielding an applicable model and several insights, for example, bicycle and e-bike mode choices are strongly correlated but also explained differently by their own driving factors. Here, we use the same data to compare a multinomial logit and a nested logit, and perform sensitivity and uncertainty analyses.

The remainder of this paper is organized as follows. Section 2 presents the conceptual (e-)bike mode choice model. Section 3 presents the data analysis used to refine the mode choice model for the Dutch situation. Section 4 presents model estimations and analyses. Section 5 discusses the usefulness of the model in case studies and policy applications. Finally, Section 6 provides conclusions and recommendations derived from this study.

2 Literature review

This section reviews the existing literature on the factors that influence the mode choice. The categorization of factors used in this section are spatial characteristics, personal and household characteristics, and travel characteristics (see Witte *et al.* (2013)). The goal of the literature review is to create a conceptual model with factors that are possibly significant and influential for Dutch urban areas, and to estimate bicycle traffic shares for municipalities.

2.1 Spatial characteristics

Several studies have shown a positive relationship between bicycle network characteristics and cycling levels. However, most network aspects are more related to route choice, and therefore, it is expected that spatial characteristics are not the most influential for mode choice. Often, the factors that can influence the mode choice are the length or density of the bicycle network and connectivity. Density was found to be positively correlated and somewhat influential on mode choice in multiple locations in Europe (Santos *et al.*, 2013), Washington, USA (Buehler, 2012), South East Queensland, Australia (Wati and Tranter, 2015) and Hamilton, Canada (Eldeeb *et al.*, 2021). The last study also showed a negative relationship with choosing public transport. This means that a higher density bicycle network leads to a higher preference for bicycles, which is mostly at the expense of the preference for public transport.

A highly used factor in network design is the separation of bicycle lanes. According to a literature review by Buehler and Dill (2015), separate bicycle tracks most often have a positive influence on bicycle choice. This has been proven by research conducted in different locations around the world. These are from Colombia (Orozco-Fontalvo *et al.*, 2018), Cyprus (Kamargianni, 2013) and Trieste, Italy (Scorrano, 2021). The influence on the choice for bike differs from small to large positive magnitudes.

The degree of access to public transport can certainly have a large influence on the choice for public transport, but it does not have a sufficiently significant influence on the choice for the bike and the car (Charreire, 2021) (Ko, 2019). This factor is often defined differently. It mostly depends on the type of research conducted to determine which definition is most valuable. The availability of a car can be defined as the possibility of going to a destination with the car. A literature review shows that car-free city centers often improve physical activity and have higher levels of active mode use (Nieuwenhuijsen, 2016). Another way of defining the availability of cars is the presence and cost of parking spaces and how these factors influence the choice for the bike. It was found that this was negatively correlated with commuting trips. One study (Ko, 2019) analyzed parking at the origin and another paper (Buehler, 2012) analyzed parking at the destination, wherein the destination is more influential than at the origin.

One of the biggest influences on bicycle mode choice is the weather. A factor often associated with weather is the season. This finding is often significant in different studies. For example, compared to the other variables in the papers, an average-sized negative influence on mode choice is found with the winter season for trips in Germany (Muller, 2008), Washington, USA (Buehler, 2012), Copenhagen (Hallberg, 2021) and Sweden (Holmgren, 2020).

2.2 Personal and household characteristics

The four most common factors related to personal traits were gender, age, occupation, and education. A logical relation to mode choice cannot be given or found in the literature for these factors for different transport modes because the results are inconsistent. Often, region and culture play a role in these different outcomes, but even within the same countries, the results can differ (Simsekoglu and Klockner, 2019) (Nello-Deakin and Harms, 2019) (Mattisson *et al.*, 2018) (Barberan *et al.*, 2017). Another important factor for mode choice is having a driver's license. Having a driver's license is expected to increase the choice for the car and sometimes has such a large influence that it also decreases the choice for the bike and the e-bike (Eldeeb *et al.*, 2021) (Kroesen, 2017).

Within households, the availability of transport modes often plays a key role in choosing a trip mode. The availability or number of cars in a household is a widely included factor in analyses because it often leads to the same result: significantly positively correlated with choosing the car. The availability of a car is so influential that even the choice for other transport modes, such as the e-bike and bike, are negatively impacted by this (Kroesen, 2017). However, the influence of this factor has the highest magnitude for the car. Besides a car that can be owned by a household, owning a bicycle is also an option that has been found to possibly influence bicycle mode choice (Hallberg, 2021) (Wati and Tranter, 2015). However, these results differed from each other. One study showed a positive influence (Heinen, 2012), while another did not find any significance in the Netherlands (Ton, 2020). It even shows that bike availability can positively influence e-bike choices (Kroesen, 2017). However, the availability of e-bikes is negatively correlated with bike mode choice (Kroesen, 2017). Moreover, owning a bike has a lower magnitude of influence on the choice for the bike than owning a car has influence on the choice for the car. The number of household members is related to the availability of transportation modes. More people in one household means less availability of cars or bicycles. An increasing number of household members can show a positive influence on choosing the bike or e-bike (Holmgren, 2020) (Ton, 2019) (Kroesen, 2017), although it was not found to be significant for the bike in all papers (Von Behren, 2020). Finally, household income can be a defining factor in choosing a transport mode. However, the results for this factor in the literature are inconclusive. When including income levels as categories instead of continuous variables, one study found that low-income groups often travel with active modes, middle-income groups travel with public transport, and high-income groups take the car more often (Ko, 2019). Nonetheless, some studies have shown no significance (Charreire, 2021) (Rodriquez, 2021). The number of household members and income are often statistically significant predictors of choosing to bike, yet their magnitudes are often small.

Population density is a city trait that has many definitions and different outcomes. The most often found definition is the size of the population density at the origin. Although the majority of papers show a positive correlation between bike and public transport choice and a negative correlation between bike and car choice, the sign of the effects can still differ significantly between papers (Kroesen, 2017) (Hallberg, 2021). Therefore, the magnitude of the influence varies significantly.

Attitude is an often-seen psychological factor. It can be described by multiple variants and is generally positively correlated with mode choice. Variants are attitudes towards efficiency (Simsekoglu, 2019), the environment (Barberan, 2017), health (Heinen, 2011), convenience (Von Behren *et al.*, 2020) and comfort and safety (Piatkowski and Marshall, 2015). Another psychological factor is habits. People are often used to certain ways of travelling, and that habit is difficult to change, which is also often true for habitual cyclists (Heesch *et al.*, 2014). Heinen *et al.* (2011) stated that the longer the distance, the more influential a personal habit is on commuting motives.

2.3 *Travel characteristics*

Modelling mode choice is often focused on using travel distance or travel time as the most important influencing factor. Oftentimes, travel time and travel distance are indeed very influential for mode choice. Most research papers that include discrete choice modelling use travel distance and travel time as a linear effect, but an exponential effect or boundary values have also been used for distance (Heinen *et al.*, 2011). The best way to model these factors for mode choice remains debated. The travel distance is related to the travel time. Ko *et al.* (2019) showed that for travel durations smaller than 30 minutes, bicycles are more sensitive to increases in travel time. The relationship between travel distance and travel time can be presented as linear, where the speed of the modes is assumed to be constant. More accurate would be to show this relationship with non-constant speeds as the speed changes throughout a trip. Rietveld and Daniel (2004) included the bicycle speed relative to the car in the Netherlands. They found a small positive influence for choosing the bike if the bicycle speed compared to the car would increase. They concluded that this is an essential element that can be influenced by municipalities by designing the spatial network in such a way that there are more direct routes and fewer stops for cyclists.

In addition to the characteristics of the trip, there are also possible influences on departure that can influence mode choice. These can be departure times or days. This has not been studied often, and in approximately 40% of these studies, it was significant (Witte *et al.*, 2013). It has been found with data from the Netherlands that a weekday has a positive relation to choosing the bike compared to a weekend day, while the time of day does not matter (Ton *et al.*, 2019). However, the magnitude is not very large.

2.4 Conceptual model

Based on the findings, the significant factors found in the literature are presented in the conceptual model shown in Figure 1. Although influential factors from other countries do not necessarily need to be as influential in the Netherlands, they are included to analyze its possible effect in the Netherlands.

Some possible influential factors were not included in this study. Most psychological factors are too time-consuming to collect during the modelling phase. It is considered to be out of the scope of this research and, therefore, not included further. Health, travel costs, travel group size, safety levels, and comfort levels were influential in the studies but were not included because there were no data directly available. Hilliness is not expected to be very influential because, on average, the Netherlands is very flat. Finally, some variables could not be included because they were not useful for the research goal. The goal is to find a model that can be used to predict modal splits between O-D pairs in urban areas for an average day, and is applicable for municipalities to stimulate bicycle use. The variables that are not included further in the research are thus short-term temporal variables: departure day, departure time, weather, season, and variables for which data is often not generally available to municipalities, such as habit and travel motives. The remaining variables are analyzed and modelled in this study. The lines connecting the two factors are hypothesized to have an interaction effect (to be tested in Section 3), which is based on literature findings (Grudgings, 2021), (Ibeas, 2014), (Heinen, 2012), (Rietveld, 2004).

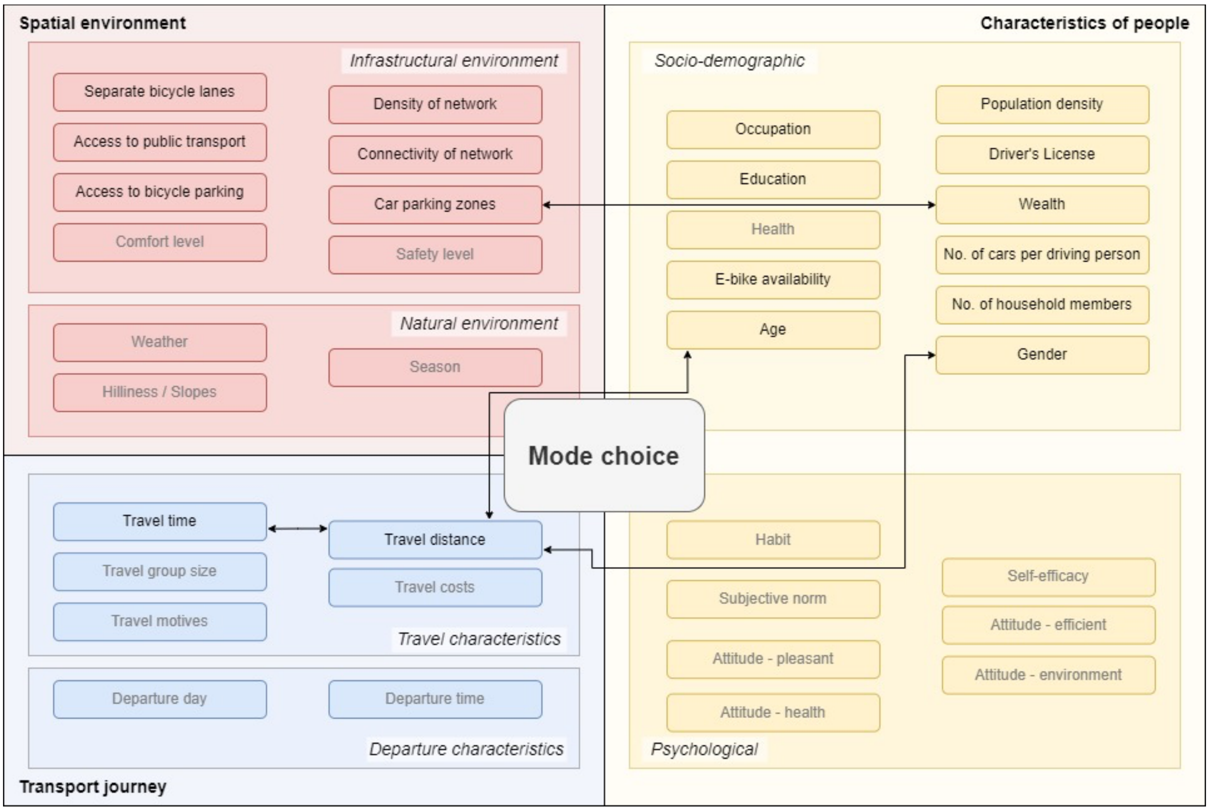

Figure 1. Conceptual model of mode choice in Dutch urban areas

3 Data

This section presents an analysis of the data used in this study. Data is gathered from the Dutch National Travel Survey, ODiN (CBS, 2022a), for 2018 and 2019. ODiN (Onderweg in Nederland) contains information about the trips taken by a respondent. It is important to note that where the trip mode is specified, bikes and e-bikes are distinguished. Furthermore, data is added from OSM (Openstreetmap, 2022) and NDOV (Nationale Databank Openbaar Vervoer, 2022).

The data is filtered by fitting the scope and cleaning the data. For the scope, the data is filtered on the recorded travel distances up to 15 kilometers, the transport modes analyzed in this research, and the urbanity level. The origin and destination zip codes of a trip are recorded in ODiN. A zip code is determined as 'urban' when the address density is larger than 1000 addresses per square kilometers. For cleaning the data, the trips that are removed are zero values in the given zip codes, trips with answers of a respondent to categorical factors such as 'not asked', 'unknown' and 'other' as it is not specific in its definition and thus not useful for modelling, and distances between the zip codes that are illogical compared to the given travel distance by the respondent. The original size of the data for 2018 and 2019 in ODiN was 374329 trips. After the filtering steps mentioned above, the size of the dataset was 160838 trips after filtering for the scope, and 116783 trips after cleaning the data. The largest filtering steps are the selection of distances and modes. ODiN contains 24 different modalities, of which seven are selected (car, bus, tram, metro, train, bike, and e-bike).

3.1 Correlation

Correlation between factors

To create a model that can predict the modal split for municipalities, the correlation between the factors must be limited. Correlation indicates whether two factors possibly explain the same phenomenon; therefore, one of the two factors must be removed from the model.

The highest positive correlations were found between travel distance and travel time. This is a logical result, and therefore, most studies include only one of the two factors. Because of this resemblance, one of the two must be chosen to limit the correlations in the model. The given travel distance and travel time from ODiN are often rounded to whole numbers or incorrectly recorded by the respondents; therefore, these factors are calculated via the OSM network. Although it is generally assumed that travel time better predicts the choice for a mode than travel distance, travel distance is calculated more accurately than travel time with the OSM network when compared to the given travel distance and travel time by respondents in ODiN, and is thus selected.

Furthermore, the street density of the car and the bike and separately the street connectivity, defined as the number of edges divided by the number of nodes of the car and the bike correlate positively. This is probably because most streets can be used for both modes; therefore, these values are often very similar. Address density correlates highly with paid car parking and access to BTM (see Table 3 for definition). This is because high address densities are mostly present in the city centers, which is also the location of most paid car parking zones and more stops for BTM modes. Because of these correlations, it is expected that the address density will explain the same phenomenon as the combination of other correlated factors when used in a model. Therefore, address density was excluded from the model.

Correlation between factor & modes

The correlation for each variable with the choice for bike and e-bike indicates that travel distance is negatively associated with the two modes, so a longer trip is less likely to be made by bike. The choice for the e-bike is mostly correlated with the ownership of the e-bike in the household.

Factors that show counter-intuitive correlations with modes are separate bicycle lanes (percentage of the total length of bicycle lines), street connectivity, and bicycle parking (presence of paid or free parking in the origin/destination zip code). The first two factors are often chosen by public transport modes when the value is increased. This was expected to have a positive influence on the (e-)bike or car modes. Bicycle parking has a positive influence on public transport modes and a negative influence on (e-)bike modes in the model. An explanation for the positive impact on public transport modes is that many of the secured and free bicycle parking spaces are at a train station or a BTM stop. ODiN records a trip mode based on the mode with which it travels the longest. Access and egress will thus be done by bike; however, the main mode of the trip will be the train or BTM.

3.2 Operationalization and description of the variables

Spatial factors

The 4-digit zip code determines origin and destination locations. Most spatial factors are determined at the zip-code level. The frequency of public transport is based on the average frequency per hour of BTM stops in the origin or destination zip codes. Access to public transport was based on the catchment area of the BTM and train stops as a percentage of the total area. The catchment area of the BTM stops is 400m, and that of the train stops is 800m (Rijsman *et al.*, 2019). Street density was calculated as the length of streets on the route divided by the total area. Paid car parking zones were calculated similarly as the area covered by paid car parking zones divided by the total area. To calculate the shortest travel distance for the car and the (e-)bike, the package OSMnx was used in Python, and for public transport, the shortest route from NDOV was used.

The exact locations from which the calculated routes are determined are based on the center of gravity of the addresses in the given origin and destination zip code in ODiN, as this provides the highest probability that the trip has started or ended close to these locations. Table 3 in paragraph 4.1 further gives the source, type, and values of the factors.

Distance boundaries

The data used to create the Cumulative Density Function (CDF) curves are the distances from ODiN. This is typically rounded to the whole kilometer-travel distances. Thus, the curves are not smooth but are more stepwise plotted. A drawback of the CDF curves is that they show the percentage of trips at a given distance, which depends on the total number of trips. Therefore, the interpretation should be performed carefully when comparing the curves with those of other modes.

The CDF curve in Figure 2 (left) shows that the bike is chosen for the shortest routes, whereas the train has the largest share of trips for longer distances. 90% of all e-bike trips and more than 95% of all bike trips were undertaken at distances of 15 km or less. This supports the research by Schneider *et al.* (2020) that the majority of bicycle trips are short- and medium-distance trips. The distance to cover when analyzing modal shifts between the bike and other transport modes is thus expected to be most useful between 0-15 km. Therefore, the data is limited to trips of less than 15 kilometers.

The modal split for the car, train, BTM, bike, and e-bike of the Dutch urban areas in 2018 and 2019 combined is shown in Figure 2 (right) for distances shorter than 15 km. It shows that the largest number of trips is still taken by car, although only approximately 65% of its trips lie between 0-15 km. For the BTM mode, approximately 68% of the trips are still included. For the train, only approximately 9% of the trips are within this selection, which leads to a very low share within the modal split of 1.4%. Train trips are often taken at larger distances as they travel between cities. Trips below 15 kilometer are possible trips for which the origin and destination are close to a train station, for which it is then feasible to take the train. Although the share is low compared to other modalities, the train is an important travel option in these cases.

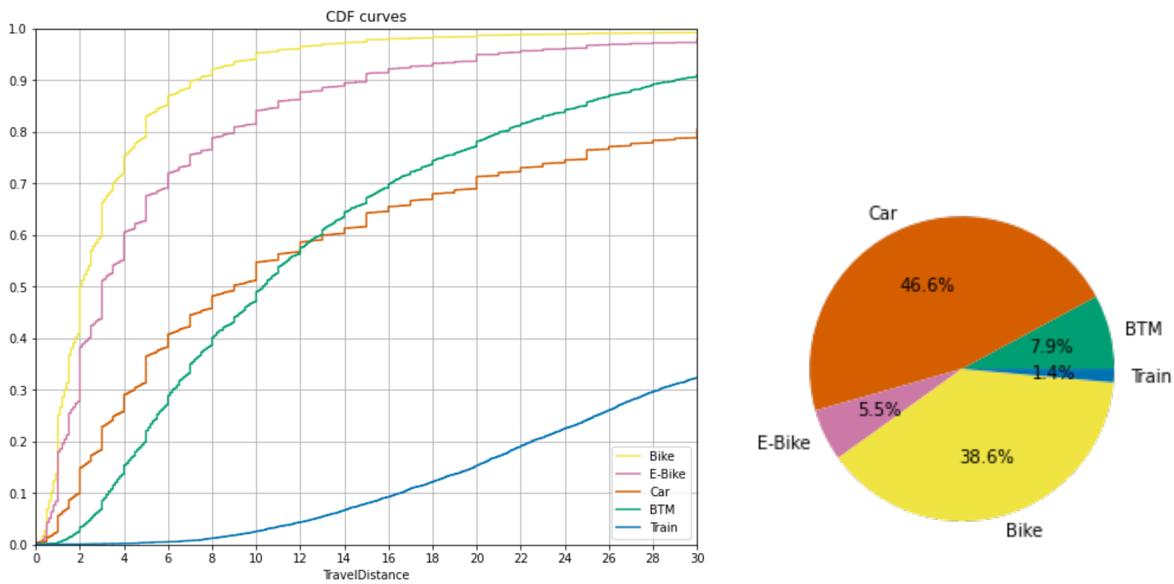

Figure 2. Difference between transport modes in urban areas (left); modal split for distances < 15 km (right)

Representativeness of the sample

The representativeness of the ODiN sample was determined by comparing the shares of the sample characteristics with national shares. The statistics in Appendix A, Table 10, show that the sample is not representative of all people's characteristics. The differences to consider in the modelling are that the ODiN sample has more respondents being middle-aged, higher educated, and living in wealthier households. Because of the large ODiN sample, a model can still provide reliable parameter estimations, even for underrepresented types of people.

3.3 *Influence of interaction effects and quadratic components*

Interaction effects and quadratic components were applied to increase the accuracy of the model further. The interaction effects and quadratic components were assessed using an out-of-sample prediction. The dataset was divided into a test set (50% of the data) and validation set (other 50% of the data). A nested logit model was then re-estimated using the test set and the estimated model was applied to the validation set. This is done for a model with the interaction, main variables, and ASC, and a model solely with the main variables and ASC. The log-likelihood results of the validation set were compared and checked for their significance and influence.

Interaction effects

The interaction effects were found in the literature and are summarized in Table 1. An interaction that can be tested based on literature are gender and travel distance. Heinen *et al.* (2012) found that males have a smaller resistance to cycling longer distances. Other interaction effects that are based on the literature are the age with the travel distance, where the expectation is that the elderly have more resistance to cycling longer distances. In addition, paid car parking zones with the wealth of a household, where it is expected that lower-wealth households will have more resistance to taking the car to a paid parking zone. Finally, paid car parking zones with the residential zip code, where it is expected that people living in a paid parking zone could have an exemption for paid parking. Another variable that will be assessed is travel speed. It was calculated by dividing the calculated travel distance by the calculated travel time.

Not all interaction effects were assessed by including the main effects of both variables. In the case of assessing the impact of paid car parking zones with the residential zip code, only the main effect of the paid car parking zones is included. A likelihood ratio test was performed to check whether the interaction between the two models with the validation set led to a better out-of-sample fit. The results of these interactions are presented in Table 1. Distance and gender were tested for all modes and were significant for bikes and e-bikes. As expected, females had a greater resistance against cycling longer distances. Thus, these results match the findings of Heinen *et al.* (2012). For the interaction between distance and age, it is expected that the elderly have a greater resistance to cycling. It was found to be significant for all modes. Travel speed was only tested for the car, BTM, and train, and these were all significant. Bike and e-bike are not incorporated for travel speed because the travel speed depends on the person riding, and it is thus not measured for these modes. Paid car parking in combination with wealth was tested only for cars, but it was not found to be significant and did not converge. Finally, paid car parking with the residential zip code was tested for all modes and was shown to be significant for all modes. Thus, four interaction effects can be used in the final model: distance and gender, distance and age, travel speed, and paid car parking and residential zip code.

Table 1. Interaction effects results of the validation set (* = p<0.05; ** = p<0.01)

	Main effects LL	Main effects + interaction LL	LR-test	Significance
Distance & Gender	-50603	-50508	0.00**	Bike & E-Bike
Distance & Age	-49731	-49683	0.00**	All
Travel Speed	-50714	-49603	0.00**	Car, BTM & Train
Paid car parking & Wealth	-52142	*Did not converge*	-	None

Paid car parking & Zip code	-52360	-51826	0.00**	All

Quadratic components

Quadratic components were used in the modelling to better fit the data with variables that are possibly not linearly described. The quadratic components analyzed were travel distance, paid car parking zones, and the number of cars per driver's license. The results in Table 2 show a comparison of the estimated values in separate nested logit models for continuous and quadratic relations. The results for the travel distance show that modelling the variable as a quadratic variable increases the out-of-sample fit in terms of the log-likelihood and rho-square bar. It can be seen that it is significant for all but the train. It is expected for the number of cars per driver's license that the other modes are being chosen less when there are more cars available. The results show a better out-of-sample fit when modelling as a quadratic variable compared with a continuous variable. The expectation of the influence of paid car parking zones is that a higher percentage of paid car parking zones in a zip code leads to higher resistance to car use. The results show that adding the quadratic variable leads to a better fit than the continuous variable and that it is significant for all modes.

Table 2. Quadratic components results of the validation set (* = p<0.05; ** = p<0.01)

	Continuous LL	Quadratic LL (LR-test)	Significance
Travel distance	-50714	-50496 (0.00**)	(E-)Bike, Car, BTM
No. of cars per driver's license	-54281	-54060 (0.00**)	(E-)Bike, Car, Train
Paid car parking zones	-52294	-52199 (0.00**)	All

4 Results

Mode choice models based on multinomial logit (MNL) and nested logit (NL) were estimated and compared in this section, and the best-fitting model was selected for further sensitivity and uncertainty analyses, as well as internal and external model validation. Note that in the previous section, we divided the dataset into a test set and a validation set to determine the best model structure. In this section, we use the whole dataset to estimate the model (with the exception of excluding data from the municipality of Haarlem for validation in section 4.4).

4.1 Comparison of Logit models

The output of the model is a modal split. This presents the possibility of analyzing changes in the modal split, and thus bicycle shares, by adjusting the input of variables when applying the model. The input of the model is data of the factors of an urban area. Each parameter is alternative-specific and in reference to the car alternative. An exception to having a reference alternative is when the data of the variable is different for each alternative, for example, travel distance. For each categorical variable, the first category of each categorical variable is the reference category. It is assumed that all Dutch residents have access to a bike and public transportation. The availability of a car and/or e-bike was registered in the ODiN dataset.

Table 3. Overview of variables included in the logit models

	Data source	*Type*	*Values*
Spatial characteristics			
Density of network	OSM, NDOV	Continuous	0...∞
Separate bicycle lanes	OSM	Continuous	0...100%
Paid car parking zones	RDW	Continuous	0...100%
Bicycle parking	VeiligStallen	Categorical	1: Security & Paid, 2: Security & Free
Access to public transport	NDOV	Continuous	0...100%
Frequency of public transport	NDOV	Continuous	0...∞
Characteristics of people			
Gender	ODiN: *Geslacht*	Binary	0: male, 1: female
Age	ODiN: *Leeftijd*	Categorical	1: 6-17, 2: 18-40, 3: 41-66, 4: 67-100
Occupation	ODiN: *MaatsPart*		
Education	ODiN: *Opleiding*	Categorical	1: primary education, 2: vmbo/mavo, 3: havo/vwo, 4: hbo/university
Driver's License	ODiN: *OPRijbewijsAu*	Binary	0: no license, 1: license
No. of household members	ODiN: *HHPers*	Continuous	0…9
Wealth	ODiN: *HHWelvG (combination of equity and income)*	Categorical	1: first 20% group (poorest), 2: second 20% group, 3: third 20% group, 4: fourth 20% group, 5: fifth 20% group (richest)
No. of cars per driving person	ODiN: *HHAuto/HHRijbewijsAu*	Continuous	0...2
Travel characteristics			
Travel distance	OSM, NDOV	Continuous	0...∞

Interaction effects	**Quadratic components**
Distance & gender	Travel distance
Distance & age	Paid car parking zones
Travel speed	No. of cars per driving person
Paid car parking zones & residential zip code	

A *multinomial logit model* is commonly used for this purpose. In this model, there are more than two dependent variables. The probability that mode i is chosen for individual n, given that the utilities equals Equation (1).

$$P_{n,i} = \frac{e^{V_{n,i}}}{\sum_j e^{V_{n,j}}} = \frac{e^{\sum_m \beta_m * x_{i,m} + \epsilon_n}}{\sum_j e^{\sum_m \beta_m * x_{j,m} + \epsilon_n}} \quad (1)$$

The advantage of this model is that it is simple and rapid in calculation. However, this method has certain assumptions and limitations. First, it assumes independence from irrelevant alternatives (IIA). It assumes that the utility associated with the common factors between alternatives does not vary across individuals. Finally, it does not account for possible correlations between choices made by the same individual over time. The results of the multinomial logit analysis are shown in Appendix A Table 11.

A *nested logit model* is used when there is a correlation between the alternatives. The nested logit (NL) calculation of an alternative in a nest is the probability of nest (B) times the probability of the alternative within the nest. The probability that mode i is chosen for individual n is given by Equation (2).

$$P_{n,i} = P[n, i | n, i \in B(n,i)] * P[n, i \in B(n,i)] \quad (2)$$

The expectation is that e-bikes and bikes have some correlation. By testing different nests in NL models and comparing the significance of the outcome, it is shown that the only significant nested structure is indeed the e-bike and bike. Another tested nest was the train and BTM mode. The reason these two modes do not appear to be significant is that the train is not often chosen at distances shorter than 15 kilometers, and the behavior of people choosing the train under these circumstances can be different from what is normally seen. The results of the nested logit model can be found in Appendix A Table 12.

Table 4 presents a comparison of the logit model results. It shows the log-likelihood and rho-square bar, which are measures that indicate the level of model fit and the number of insignificant variables. The log-likelihood for the nested logit seems to show that the model fit is better than that of the MNL. The rho-squared bar does not change significantly between the two logit models. This indicates that both models have a similar model fit. The nested logit model have a greater number of insignificant variables. Table 5 presents the estimation of the modal split for each model using the data fitted with. This shows that the MNL and NL have similar modal split estimations. However, the MNL has a better e-bike and car share prediction, whereas the NL has a better bicycle share prediction. The predictions for the BTM and train shares are the same.

Table 4. Comparison of logit models

	MNL	NL
Log likelihood	-86921.31	- 86657.77
Rho-square-bar	0.468	0.469
No. of insignificant variables	23	29

Table 5. Estimation of modal split of logit models

	Bike	E-Bike	Car	BTM	Train
Actual modal split	33.1%	6.5%	50.7%	8.1%	1.6%
Estimated modal split with MNL	34.1%	7.5%	53.1%	4.7%	0.4%
Estimated modal split with NL	33.1%	8.1%	53.7%	4.7%	0.4%

4.2 Sensitivity

The final model (nested logit) was tested on its sensitivity. The purpose of this analysis is to determine the extent to which beta estimates contribute to changes in the estimated modal split.

A sensitivity analysis was performed to assess the sensitivity of the model outcome as a result of the uncertainty in the estimated beta parameters. For each variable, a random value was drawn 10000 times from the normal distribution of the beta parameter in each mode. The mean and standard deviation of the final model are used. These mean parameter values were then changed, and the probabilities of the modes were calculated based on the averages and standard deviations of the population of the Netherlands (CBS, 2022b) or, when not available, the ODiN dataset. The calculation of the probabilities is based on the occurrence of people owning an e-bike or car. It is assumed that everyone has access to a bike, train, or BTM. The occurrences are based on the averages of the Netherlands, which are 86% for owning a car and 13% for owning an e-bike. Four occurrences can then be formed based on availability. There is an 11% chance of owning all modes, a 75% chance of owning no e-bike, a 2% chance of owning no car, and a 12% chance of owning no e-bike and no car. The final probability was the weighted average of the probabilities and occurrences.

The sensitivity analysis results in Table 6 show the first- and second-order analyses. Values with deviations larger than 1 percentage point for the standard deviation are shown in the first-order analysis, and these are also used in the second-order analysis. Second-order sensitivity effects analyze two variables changing simultaneously, which could lead to larger sensitivities in the model.

The variables do not show large variations or differences from the original mean in the first-order analysis. This results from the fact that these variables are significant, which means that the beta estimates do not have large standard deviations. Therefore, the model is robust and insensitive to any possible changes in parameter estimates. A change in one variable generally leads to very small changes in the modal split. The second-order effect was also analyzed. Here, it can be seen that the change in the beta parameters of the two variables does not amplify the variation when they are adjusted at the same time compared with their variations separately. Thus, second-order sensitivities were not present in the model.

Table 6. Result of sensitivity analysis

	Bike	E-Bike	Car	BTM	Train
Original	36.28%	4.66%	52.47%	6.26%	0.33%
First-order					
Alternative Specific Constants	36.32% (3.08)	4.64% (0.83)	52.28% (3.24)	6.41% (1.46)	0.35% (0.15)
Cars per driver's license	36.28% (1.60)	4.66% (0.43)	52.44% (1.81)	6.29% (0.68)	0.34% (0.09)
Travel distance	36.27% (1.77)	4.65% (0.23)	52.43% (1.46)	6.31% (0.99)	0.34% (0.10)
Second-order					
ASC & Travel distance	36.23% (2.68)	4.66% (0.82)	52.36% (3.05)	6.41% (1.54)	0.34% (0.12)
ASC & Cars per driver's license	36.24% (3.20)	4.65% (0.88)	52.32% (3.35)	6.42% (1.56)	0.37% (0.19)
Cars per driver's license & Travel distance	36.27% (2.36)	4.65% (0.48)	52.38% (2.32)	6.35% (1.20)	0.35% (0.15)

4.3 Uncertainty analysis

An uncertainty analysis was performed to assess the uncertainty of the outcome because of changes in the input values. The same calculation was performed for sensitivity analysis. The only difference is that the input values were changed instead of the beta parameters. The mean and standard deviation for the normal distribution of the input values were based on the population of the Netherlands (CBS, 2022b) or the ODiN dataset.

The uncertainty analysis results in Table 7 show first-order effects with the largest standard deviations (shown in brackets) and their second-order effects. It can be seen that the travel distance is the most influential variable for the probabilities of the modes. None of the other input variables changed the original mean probabilities by more than one percentage point. Other notable results are larger than the average standard deviations of street density and driver's license. For at least one of the modes, the standard deviation was larger than five percentage points. A change in the input of these variables can therefore change the probability of the modes the most.

To analyze second-order effects in the uncertainty analysis, the variables that most influence the outcome are combined to analyze their behavior together. This shows that the largest variations can be found for the combinations with travel distance, although the combined variations are not significantly larger than the variation in travel distance alone. Because the largest changes can be found in the travel distance and mostly for the bike, it is thus important for the travel distance to have accurate input values.

Table 7. Result of uncertainty analysis

	Bike	E-Bike	Car	BTM	Train
Original	36.28%	4.66%	52.47%	6.26%	0.33%
First-order					
Street density	36.56% (6.90)	4.60% (0.52)	52.18% (6.43)	6.31% (1.15)	0.35% (0.16)
Driver's license	36.29% (5.42)	4.59% (0.57)	51.76% (9.52)	7.00% (3.51)	0.36% (0.16)
Travel distance	39.61% (21.78)	4.47% (1.96)	47.48% (20.71)	5.39% (3.89)	3.06% (8.95)
Second-order					

Travel distance & Street density	39.85%	4.44%	46.86%	5.43%	3.42%
	(22.43)	(1.98)	(21.24)	(4.03)	(10.09)
Travel distance & Driver's license	39.78%	4.46%	46.60%	5.91%	3.25%
	(21.96)	(1.95)	(21.67)	(5.32)	(9.43)
Street density & Driver's license	36.47%	4.53%	51.58%	7.04%	0.39%
	(8.73)	(0.73)	(11.24)	(3.74)	(0.26)

4.4 *Validation*

First, an internal validation is performed at the municipal level to check if and how the model can be used by municipalities. The data used in the internal validation were used in the model estimation. Therefore, a location-based external validation was also performed. The location-based validation determines the mode choice for each trip to and/or from Haarlem, because it is a municipality and an urban area in the Netherlands. The data were separated from the ODiN 2018/2019 dataset before modelling. The modal split generated from the estimated probabilities per trip is compared with the original modal split of the dataset to determine how well the model can predict out-of-sample.

The results of the internal validation are presented in Table 8. The estimated modal split is, unlike the input of the sensitivity and uncertainty analysis, calculated with the average per variable from ODiN because the data is used in the estimation of the model parameters. Internal validations for municipalities were performed for two large cities, Rotterdam and Amsterdam, and a relatively smaller city, Dordrecht. It is shown that the accuracy of the model is highest for Amsterdam. The accuracy of Rotterdam was lower, with deviations of almost 10 percent point for the bike and BTM. The validation of Dordrecht has deviations in accuracy of up to approximately 8 percent points for the car. The bike and e-bike were better estimated in Dordrecht than in Rotterdam.

The results of the external validation are presented in Table 9. This shows that, for Haarlem, the model underestimates the choice for the bike. It can be said that the respondents of Haarlem seem to have a larger preference for bikes compared to the average preference of the ODiN dataset.

Table 8. Results of internal validations

Internal validation (Amsterdam)	Bike	E-Bike	Car	BTM	Train
Actual modal split	43.7%	2.2%	21.7%	28.8%	3.6%
Estimated modal split	40.1%	3.8%	21.2%	31.9%	2.4%

Internal validation (Rotterdam)	Bike	E-Bike	Car	BTM	Train
Actual modal split	26.1%	3.2%	37.6%	30.6%	2.5%
Estimated modal split	35.3%	5.5%	37.6%	20.4%	1.2%

Internal validation (Dordrecht)	Bike	E-Bike	Car	BTM	Train
Actual modal split	28.0%	6.7%	60.2%	4.3%	0.7%
Estimated modal split	32.9%	6.5%	52.9%	7.0%	0.7%

Table 9. Results of external validation

External validation (Haarlem)	Bike	E-Bike	Car	BTM	Train
Actual modal split	45.1%	3.9%	42.7%	7.1%	1.2%
Estimated modal split	30.8%	8.6%	51.1%	8.7%	0.8%

5 Discussion

This chapter discusses the application of the model, as well as the limitations and recommendations of the research.

5.1 *Application of the model*

The contribution of this study is that it provides a mode-choice model that has been made more accurate for calculating the bicycle share in Dutch urban areas with trips up to 15 kilometers. Moreover, to the best of our knowledge, it is the first time e-bikes are included in a mode choice model for Dutch urban areas.

The model is practically applicable for municipalities to predict modal shifts towards (e-)bikes in urban areas for changes in their networks or policies. The validation showed that the model could predict fairly well at the municipal level. However, the prediction accuracy differed for each municipality. Therefore, it is advised to validate the model for the area in which it will be used.

Examples of measures from the municipalities of Rotterdam (2022) and Amsterdam (2022) that can be analyzed with the final model include the following:

- Network measures, such as changes in the street network for the (e-)bike and the car, can be modelled by adjusting the factors of street density and travel distance.
- Service measures, such as changes in the accessibility of public transport, can be modelled by adjusting the frequency of BTM modes or increasing the number of train stations and BTM stops, thus increasing the percentage of catchment area covering the area of a zip code.
- Policy measures, such as changes in maximum speed limits for cars, could be modelled by adjusting the travel speed and changes in paid car parking restrictions by adjusting the amount of paid car parking in a zone.
- Behavioral measures, such as the impact of stimulating e-bicycle ownership, can be observed by increasing the availability of e-bikes in the model.

Some measures cannot be modelled as they are not (directly) a factor in the model, such as cost-related parking restrictions, bicycle parking facilities, traffic control systems where (e-)bicycles would receive priority at intersections, and shared mobility.

This review of possible measures shows that not all options can be modelled and analyzed by municipalities. Moreover, specifically modelling these *changes* in networks and policies has not been validated, and thus, it cannot be stated that the model will provide accurate results for the modal *shift* to form expectations.

5.2 Limitations and recommendations of the research

Regarding the data available for this study, a few limitations can be noted, and recommendations can be made on how this could be improved. First, the sample was not an accurate representation of the Netherlands. To further optimize the model, it is recommended that the weight factors in ODiN are studied and used in the modelling phase in other similar studies. Second, the access and egress modes were not included. This leads to unexpected effects on the mode choice for some factors, such as bicycle parking. Third, walking is not included in the model, as the data used in this study, which is at the zip code level, cannot make accurate predictions for the travel distances, while this is expected to have a large effect on choosing to walk. Thus, walking should not be overlooked as a possible substitute for (e-)bikes. Finally, the data may indicate some level of panel structure. ODiN asks respondents to record the trips they have made. However, the average number of trips per respondent in the 2018 and 2019 datasets was only 2.87 trips. Oftentimes, a respondent recording two trips gave their round trip, with the first trip being very similar to the second trip. Nevertheless, it would be valuable for future research to analyze the level of panel structure in the ODiN dataset and how this would affect the estimations of a logit model.

Second, not all data sources were reliable or accurate, which led to the exclusion of factors from the model that could be significant for mode choice. The calculation of the infrastructural factors for public transport modes is less accurate than that for the other modes. It is difficult to calculate the shortest path when the data consists only of the network and stops of public transport. Therefore, it does not include access modes, egress modes, or accurate transfers. A comparison of the calculated and given travel distances shows that the travel distance is often overestimated by the respondents from ODiN, which can cause deviations in the accuracy of the predictions. However, the validation showed that public transport was accurately predicted for the goal of this research. Furthermore, bicycle parking includes biased/missing data, where most bicycle parking facilities are located close to train stations. It is advised that the national government improve and expand their datasets for bicycle parking facilities and that researchers analyze the influence of this factor on bicycle mode choice.

Third, some of the data were ill-defined or incomplete. Street connectivity is calculated as the number of edges divided by the number of nodes. It has been shown that this is not the best representation of street connectivity as it did not show the expected results. It is advised to research other definitions for street connectivity, such as the number of cul-de-sacs or the number of intersections. Travel speed is not used for bikes and e-bikes, as there are no limitations to cycling speeds. However, travel speeds can differ between cycling in the city and cycling in the countryside. Therefore, it is advised to analyze and research the influence of travel speed on the mode choice for these two modes.

Fourth, this study used cross-sectional data. The validation in this study was based on internal and cross-sectional external validation. A limitation of this study is that longitudinal external validation was not performed. The data source, ODiN, was previously called OViN until 2017, which gathered the data differently. After 2019, travel behavior was possibly different because of the covid-19 pandemic. No relevant data is thus available to perform a longitudinal external validation, which would have been useful for municipalities to assess the usefulness of the model for analyzing the modal split in future scenarios. To improve the applicability of the model, it is advised to gather data from before and after bicycle network or policy measures to validate the predicted modal split of the model for these changes. After validation, conclusions for implementations could be made with the model, which can be used by municipalities in their decision-making.

Finally, this study focused on 2018 and 2019. In recent years, travel behavior has been severely influenced by the covid-19 pandemic. Private transport was preferred over public transport due to hygienic considerations, and working from home led to less congestion on the road (Scorrano, 2021). The degree to which after the pandemic preferences may have changed is not yet known. It is advised to repeat this analysis as data on travel behavior after 2022 becomes available.

6 Conclusion

The objective of this research was to find factors to develop a mode choice model for bikes, e-bikes, cars, BTM, and trains for trips of up to 15 kilometers in urban areas in the Netherlands. The factors used to specify the model were derived from the literature. Data was obtained from the National Travel Survey ODiN and enriched with other national sources (NDOV, CBS, RDW, and CROW). Interaction effects and quadratic components were tested to further improve the model. The final model is based on a nested logit model, as shown in Appendix A Table 12. Internal validation showed that the accuracy of the prediction of the modal split was very close to the actual modal split. However, the accuracy of the prediction of modal splits differs by municipality. Location-based external validation for the city of Haarlem showed that the model underestimated the choice for the bike.

The factor that is highly influential on both bikes and e-bikes is the travel distance. Other influencing factors are owning a driver's license and street density. The least influential factors of mode choice were personal and household characteristics. These statements demonstrate how infrastructure and spatial networks can greatly affect mode choice. Additionally, the choice for the (e-)bike is often dependent on factors related to the choice for the car. Discouraging the choice for the car, for example, by increasing the travel distance or introducing paid car parking zones in popular destinations, will thus lead to a larger choice for the (e-)bike.

The choice for the bike mainly differs from that for the e-bike in the availability of the e-bike. While most Dutch people have access to more than one bike, not everyone owns an e-bike. Moreover, the choice for the bike can be more heavily influenced by factors of other modes than the choice for the e-bike. Examples include car ownership, the frequency of BTM stops, and owning a driver's license. In addition, person and household characteristics have different influences on the choice for the bike compared to the e-bike. Lastly, street density has a greater positive influence on the choice for the bike than on the choice for the e-bike. These differences demonstrate the need to model the e-bike independently of the bike in mode choice models.

This study shows that a nested logit approach provides the most accurate modal split prediction for Dutch urban areas over short distances from the tested models. Using this model, the differences between the choice between conventional bicycles and e-bikes can be assessed, and changes in (e-)bicycle policies and networks can be analyzed by altering the input accordingly.

Contributor Statement

Conceptualization: Chantal Huurman, Adam Pel, Winnie Daamen, Kees Maat
Data Curation: Chantal Huurman
Formal analysis: Chantal Huurman
Investigation: Chantal Huurman
Methodology: Chantal Huurman, Adam Pel, Winnie Daamen, Kees Maat
Project administration: Chantal Huurman
Software: Chantal Huurman
Validation: Chantal Huurman
Visualisation: Chantal Huurman
Writing – Original Draft: Chantal Huurman
Writing – Review & Editing: Chantal Huurman, Adam Pel, Winnie Daamen, Kees Maat

Acknowledgement

This study is partly based on a master's thesis written at the Delft University of Technology in collaboration with Witteveen+Bos (available at: http://resolver.tudelft.nl/uuid:89003a8b-3917-4ea6-8398-8d9723faa092). In this regard, the authors would like to thank Sander Veenstra from Witteveen+Bos for his practical point of view on the master's thesis and guidance in processing the data using Python and QGIS.

Conflict of Interest (COI)

There is no conflict of interest.

References

Barberan, A., Silva, J. D. A. E. and Monzon, A. (2017). Factors influencing bicycle use: a binary choice model with panel data. *Transportation Research Procedia,* 27, 253.

von Behren, S., Schubert, R., and Chlond, B. (2020). International comparison of psychological factors and their influence on travel behavior in hybrid cities. *Research in Transportation Business and Management,* 36.

Bicycle Dutch (2018). Dutch cycling figures.

Buehler, R., and Dill, J. (2015). Bikeway Networks: A review of effects on cycling. *Transport Reviews,* 36, 9.

Buehler, R. (2012). Determinants of bicycle commuting in the Washington, DC region: The role of bicycle parking, cyclist showers, and free car parking at work. *Transportation Research Part D: Transport and Environment*, 17, 525.

CBS (2022a). Onderweg in Nederland (ODiN) 2018-2020. *Centraal Bureau Voor De Statistiek* https://www.cbs.nl/nl-nl/longread/rapportages/2022/onderweg-in-nederland--odin---2018-2020

CBS (2022b). StatLine. *Centraal Bureau Voor De Statistiek*. https://opendata.cbs.nl/statline/#/CBS/nl/

CBS (2022c). Stedelijkheid (van een gebied). *Centraal Bureau Voor De Statistiek*. https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/stedelijkheid--van-een-gebied--

Charreire, H., Roda, C., Feuillet, T., Piombini, A., Bardos, H., Rutter, H., Compernolle, S., Mackenbach, J. D., Lakerveld, J., Oppert, J. M., and et al. (2021). Walking, cycling, and public transport for commuting and noncommuting travels across 5 European urban regions: Modal choice correlates and motivations. *Journal of Transport Geography,* 96, 103196.

Eldeeb, G., Mohamed, M., and Paez, A. (2021). Built for active travel? Investigating the contextual effects of the built environment on transportation mode choice. *Journal of Transport Geography,* 96, 103158.

Gemeente Rotterdam (2022). Vervoer. https://www.rotterdam.nl/vervoer

Gemeente Amsterdam (2022). Volg het beleid: Verkeer en vervoer. https://www.amsterdam.nl/bestuur-organisatie/volg-beleid/verkeer-vervoer/

Grudgings, N., Hughes, S., and Hagen-Zanker, A. (2021). The comparison and interaction of age and gender effects on cycling mode-share: An analysis of commuting in England and Wales. *Journal of Transport and Health*, 20, 101004.

Hallberg, M., Rasmussen, T. K., and Rich, J. (2021). Modelling the impact of cycle superhighways and electric bicycles. *Transportation Research Part A: Policy and Practice*, 149, 397.

Heesch, K. C., Giles-Corti, B., and Turrell, G. (2014). Cycling for transport and recreation: Associations with socioeconomic position, environmental perceptions, and psychological disposition. *Preventive Medicine***,** 63, 29.

Heinen, E., Maat, K., and Van Wee, B. (2012). The effect of work-related factors on the bicycle commute mode choice in the Netherlands. *Transportation,* 40, 23.

Heinen, E., Maat, K., and Van Wee, B. (2011). The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances. *Transportation Research Part D: Transport and Environment*, 16, 102.

Holmgren, J., and Ivehammar, P. (2020). Mode choice in home-to-work travel in mid-size towns: The competitiveness of public transport when bicycling and walking are viable options. *Transportation Research Procedia,* 48, 1635.

Ibeas, A., Dell'Olio, L., Bordagaray M. E., and De Dios Ortúzar, J. (2014). Modelling parking choices considering user heterogeneity. *Transportation Research Part A-policy and Practice*, 70, 41–49.

Kamargianni, M., and Polydoropoulou, A. (2013). Hybrid choice model to investigate effects of teenagers' attitudes toward walking and cycling on mode choice behavior. *Transportation Research Record,* 151.

KiM (2018). Cycling Facts 2018.

Ko, J., Lee, S., and Byun, M. (2019). Exploring factors associated with commute mode choice: An application of city-level general social survey data. *Transport Policy***,** 75, 36.

Kroesen, M. (2017). To what extent do e-bikes substitute travel by other modes? Evidence from the Netherlands. *Transportation Research Part D: Transport and Environment*, 53, 377.

Mattisson, K., Idris, A. O., Cromley, E., Hakansson, C., Ostergren, P. O., and Jakobsson, K. (2018). Modelling the association between health indicators and commute mode choice: a cross-sectional study in southern Sweden. *Journal of Transport & Health,* 11, 110.

Muller, S., Tscharaktschiew, S., and Haase, K. (2008). Travel-to-school mode choice modelling and patterns of school choice in urban areas. *Journal of Transport Geography,* 16, 342.

NDOV loket (2022). https://ndovloket.nl/

Nello-Deakin, S., and Harms, L. (2019). Assessing the relationship between neighbourhood characteristics and cycling: Findings from Amsterdam. *Transportation Research Procedia,* 41, 17.

Nieuwenhuijsen, M. J., and Khreis, H. (2016). Car free cities: Pathway to healthy urban living. *Environment International,* 94, 251.

Orozco-Fontalvo, M., Arevalo-Tamara, A., Guerrero-Barbosa, T., and Gutierrez-Torres, M. (2018). Bicycle choice modeling: A study of university trips in a small Colombian city. *Journal of Transport & Health,* 9, 264.

OpenStreetMap (2022). https://www.openstreetmap.org/

Piatkowski, D. P., and Marshall, W. E. (2015). Not all prospective bicyclists are created equal: The role of attitudes, socio-demographics, and the built environment in bicycle commuting. *Travel Behaviour and Society,* 2, 166.

Rietveld, P., and Daniel, V. (2004). Determinants of bicycle use: do municipal policies matter? *Transportation Research Part A: Policy and Practice,* 38, 531.

Rijsman, L., Van Oort, N., Ton, D., Hoogendoorn, S., Molin, E., and Teijl, T. (2019). Walking and bicycle catchment areas of tram stops: factors and insights.

Rodriguez-Valencia, A., Rosas-Satizabal, D., Unda, R., and Handy, S. (2021). The decision to start commuting by bicycle in Bogota, Colombia: Motivations and influences. *Travel Behaviour and Society,* 24, 57.

Santos, G., Maoh, H., Potoglou, D., and Brunn, T. V. (2013). Factors influencing modal split of commuting journeys in medium-size European cities. *Journal of Transport Geography,* 30, 127.

Schneider, F., Jensen, A. F., Daamen, W., and Hoogendoorn, S. (2020). Bicycle accessibility: What can we learn from best-practice examples*? Journal of Transport Geography*.

Scorrano M., and Danielis, R. (2021). Active mobility in an Italian city: Mode choice determinants and attitudes before and during the Covid-19 emergency. *Research in Transportation Economics,* 86.

Simsekoglu, O., and Klockner, C. A. (2019). The role of psychological and socio-demographical factors for electric bike use in Norway. *International Journal of Sustainable Transportation,* 13, 315.

Ton, D., Bekhor, S., Cats, O., Duives, D. C., Hoogendoorn-Lanser, S., and Hoogendoorn, S. P. (2020). The experienced mode choice set and its determinants: Commuting trips in the Netherlands. *Transportation Research Part A: Policy and Practice,* 132, 744.

Ton, D., Duives, D. C., Cats, O., Hoogendoorn-Lanser, S., and Hoogendoorn, S. P. (2019). Cycling or walking? Determinants of mode choice in the Netherlands. *Transportation Research Part A: Policy and Practice,* 123, 7.

Wati, K., and Tranter, P. J. (2015). Spatial and socio-demographic determinants of South East Queensland students' school cycling. *Journal of Transport Geography***,** 47, 23.

Witte, A. D., Hollevoet, J., Dobruszkes, F., Hubert, M., and Macharis, C. (2013). Linking modal choice to motility: A comprehensive review. *Transportation Research Part A: Policy and Practice*, 49, 329.

Appendix A: Supplementary tables

Table 10. Representativeness of ODiN data sample (CBS, 2022)

		National share	Filtered ODiN share
Gender	*men*	49.64%	47.85%
	women	50.36%	52.15%
Age	*average*	41.9	45.60
	0-20	22.05%	1.83%
	20-40	24.85%	40.43%
	40-65	34.05%	42.91%
	65+	19.00%	14.83%
Occupation	*part-time*	11.16%	18.94%
	full-time	34.81%	45.11%
	own housekeeping	15.30%	3.38%
	student	12.99%	11.62%
	unemployed	1.38%	4.21%
	unable to work	4.35%	
	retired	20.01%	16.74%
Education	*primary education*	9.96%	2.57%
	vmbo/mavo	20.75%	13.50%
	havo/vwo	37.23%	35.07%
	hbo/university	30.57%	48.86%
	not known	1.49%	0.00%
Driver's License	*no*	35.45%	13.37%
	yes	64.55%	86.63%
Cars per driver's license	*average*	0.79	0.68
Household members	*average*	2.15	2.64
Wealth	*first 20% group*	20.00%	11.35%
	second 20% group	20.00%	15.56%
	third 20% group	20.00%	19.60%
	fourth 20% group	20.00%	24.73%
	fifth 20% group	20.00%	28.77%
E-bike ownership	*no*	86.90%	77.22%
	yes	13.10%	22.78%

Table 11. Multinomial logit model

MNL - All variables: Summary

	BIKE	*E-BIKE*	*CAR*	*BTM*	*TRAIN*
ASC	1.060**	1.890**	*ref*	-4.600**	-3.220**
Access to public transport (Origin)	-	-	-	1.320**	0.800**
Access to public transport (Destination)	-	-	-	1.400**	0.949**
Age: 0-17	*ref*	*ref*	*ref*	*ref*	*ref*
Age: 18-40	-0.837**	-0.754**	*ref*	-0.691**	-1.510**
Age: 41-66	-0.776**	-0.468*	*ref*	-0.903**	-1.620**
Age: 67-100	-0.953**	-0.444*	*ref*	-0.067	-3.230**
Bicycle parking: Security & Paid (Origin)	*ref*	*ref*	*ref*	*ref*	*ref*
Bicycle parking: Security & Free (Origin)	0.113**	0.076	*ref*	0.200**	0.178*
Bicycle parking: Security & Paid (Destination)	*ref*	*ref*	*ref*	*ref*	*ref*
Bicycle parking: Security & Free (Destination)	0.056*	0.028	*ref*	0.247**	0.241**
Car ownership	0.796**	-0.013	*ref*	0.544**	1.410**
Quadratic: car ownership	-0.788**	-0.288**	*ref*	-0.618**	-1.510**
Travel distance	-0.307**	-0.295**	-0.142**	0.175**	-0.135**
Quadratic: travel distance	0.010**	0.009**	0.002	-0.007**	0.000
Interaction: distance & age 0-17	*ref*	*ref*	-	*ref*	*ref*
Interaction: distance & age 18-40	-0.068**	-	-	-	0.047**
Interaction: distance & age 41-66	-0.065**	-	-	-	0.023
Interaction: distance & age 67-100	-0.175**	-0.080**	-	-0.087**	0.023
Interaction: distance & gender	-0.068**	-0.077**	-	-	-
Travel speed	-	-	*0.012***	-0.514*	10.600**
Car parking (Origin)	0.500**	0.949**	*ref*	0.776**	2.030**
Quadratic: car parking (Origin)	0.291*	-0.276	*ref*	0.008	-1.590**
Interaction: car parking (Origin) & residential postcode	0.105*	0.002	*ref*	0.769**	0.386**
Car parking (Destination)	0.594**	1.000**	*ref*	0.508*	2.350**
Quadratic: car parking (Destination)	0.072	-0.466	*ref*	0.132	-2.090**
Interaction: car parking (Destination) & residential postcode	0.154**	0.082	*ref*	0.827**	0.583**
Education: Primary education	*ref*	*ref*	*ref*	*ref*	*ref*
Education: vmbo/mavo	0.014	-0.237**	*ref*	0.553**	0.139
Education: havo/vwo	0.139*	-0.414**	*ref*	0.427**	0.222
Education: hbo/university	0.484**	-0.298**	*ref*	0.241**	0.343
Frequency of BTM stops (Origin)	0.135**	0.084*	*ref*	0.431**	0.424**
Frequency of BTM stops (Destination)	0.198**	0.094*	*ref*	0.410**	0.320**
Gender	0.101**	0.693**	*ref*	0.119**	-0.247**
Driver's License	-0.900**	-0.615**	*ref*	-1.620**	-1.510**
Household Members	0.049**	-0.083**	*ref*	0.078**	-0.030
Occupation: Part-time job	*ref*	*ref*	*ref*	*ref*	*ref*
Occupation: Full-time job	-0.229**	-0.269**	*ref*	-0.143**	-0.276**
Occupation: Own housekeeping	-0.160**	0.049	*ref*	0.120	-0.897**
Occupation: Student	0.788**	-0.795**	*ref*	1.060**	0.914**
Occupation: Unemployed / Unable to work	-0.462**	-0.153*	*ref*	0.008	-1.320**
Separate bicycle lanes	-3.030**	-1.320**	-	-	-
Street Density	1.590**	2.300**	1.590**	0.591**	-7.730**
Wealth: First 20% group	*ref*	*ref*	*ref*	*ref*	*ref*
Wealth: Second 20% group	0.199**	0.214**	*ref*	0.194**	0.259**
Wealth: Third 20% group	0.335**	0.311**	*ref*	0.478**	0.306**
Wealth: Fourth 20% group	0.407**	0.441**	*ref*	0.344**	0.663**
Wealth: Fifth 20% group	0.588**	0.337**	*ref*	0.251**	0.674**
Log likelihood: -86921.31 Rho-square-bar: 0.468					

ref = reference alternative / category ** = p < 0.05* *** = p < 0.01*

Table 12. Nested logit model

NL - All variables: Summary

	BIKE	*E-BIKE*	*CAR*	*BTM*	*TRAIN*
ASC	1.190**	1.930**	*ref*	-4.560**	-3.150**
Access to public transport (Origin)	-	-	-	1.320**	0.802**
Access to public transport (Destination)	-	-	-	1.400**	0.950**
Age: 0-17	*ref*	*ref*	*ref*	*ref*	*ref*
Age: 18-40	-0.831**	-0.853**	*ref*	-0.692**	-1.510**
Age: 41-66	-0.737**	-0.611**	*ref*	-0.894**	-1.620**
Age: 67-100	-0.853**	-0.663**	*ref*	-0.031	-3.230**
Bicycle parking: Security & Paid (Origin)	*ref*	*ref*	*ref*	*ref*	*ref*
Bicycle parking: Security & Free (Origin)	0.110**	0.099*	*ref*	0.196**	0.176
Bicycle parking: Security & Paid (Destination)	*ref*	*ref*	*ref*	*ref*	*ref*
Bicycle parking: Security & Free (Destination)	0.054	0.053	*ref*	0.244**	0.238*
Car ownership	0.720**	0.390**	*ref*	0.458**	1.320**
Quadratic: car ownership	-0.750**	-0.499**	*ref*	-0.578**	-1.460**
Travel distance	-0.307**	-0.318**	-0.143**	0.177**	-0.135**
Quadratic: travel distance	0.010**	0.009**	0.002	-0.008**	0.000
Interaction: distance & age 0-17	*ref*	*ref*	-	*ref*	*ref*
Interaction: distance & age 18-40	-0.064**	-	-	-	0.047**
Interaction: distance & age 41-66	-0.061**	-	-	-	0.023
Interaction: distance & age 67-100	-0.161**	-0.072**	-	-0.090**	0.023
Interaction: distance & gender	-0.069**	-0.057**	-	-	-
Travel speed	-	-	*0.013***	-0.508	10.600**
Car parking (Origin)	0.497**	0.741**	*ref*	0.783**	2.040**
Quadratic: car parking (Origin)	0.267*	-0.081	*ref*	-0.005	-1.610**
Interaction: car parking (Origin) & residential postcode	0.111*	0.072	*ref*	0.766**	0.383*
Car parking (Destination)	0.593**	0.779**	*ref*	0.513*	2.350**
Quadratic: car parking (Destination)	0.046	-0.243	*ref*	0.121	-2.100**
Interaction: car parking (Destination) & residential postcode	0.160**	0.134	*ref*	0.824**	0.582**
Education: Primary education	*ref*	*ref*	*ref*	*ref*	*ref*
Education: vmbo/mavo	0.037	-0.150	*ref*	0.550**	0.138
Education: havo/vwo	0.103	-0.273**	*ref*	0.409**	0.205
Education: hbo/university	0.417**	-0.108	*ref*	0.213	0.317
Frequency of BTM stops (Origin)	0.128**	0.101**	*ref*	0.428**	0.421**
Frequency of BTM stops (Destination)	0.188**	0.135**	*ref*	0.407**	0.317**
Gender	0.125**	0.499**	*ref*	0.123**	-0.243**
Driver's License	-0.868**	-0.679**	*ref*	-1.600**	-1.500**
Household Members	0.043**	-0.053**	*ref*	0.075**	-0.033
Occupation: Part-time job	*ref*	*ref*	*ref*	*ref*	*ref*
Occupation: Full-time job	-0.244**	-0.268**	*ref*	-0.145*	-0.277*
Occupation: Own housekeeping	-0.145*	-0.020	*ref*	0.124	-0.894*
Occupation: Student	0.745**	-0.239*	*ref*	1.040**	0.902**
Occupation: Unemployed / Unable to work	-0.440**	-0.225**	*ref*	0.024	-1.300**
Separate bicycle lanes	-2.990**	-1.850**	-	-	-
Street Density	1.660**	2.560**	1.660**	0.599*	-7.690**
Wealth: First 20% group	*ref*	*ref*	*ref*	*ref*	*ref*
Wealth: Second 20% group	0.210**	0.229**	*ref*	0.198*	0.266
Wealth: Third 20% group	0.362**	0.324**	*ref*	0.486**	0.315*
Wealth: Fourth 20% group	0.442**	0.445**	*ref*	0.356**	0.675**
Wealth: Fifth 20% group	0.608**	0.426**	*ref*	0.259**	0.683**
Nest parameter: Bike & E-Bike	1.810**		-	-	-
Log likelihood: -86657.77 Rho-square-bar: 0.469					

ref = reference alternative / category $* = p < 0.05$ $** = p < 0.01$