J. van Ewijk

Stimulating cyclists to choose a safer route

A study with a focus on infrastructure

July 2020

Master Thesis

Construction Management and Engineering

Eindhoven University of Technology

Graduation Committee

Chairman	:	prof.dr.ir. B. (Bauke) de Vries
1st supervisor	:	dr.ing. P.J.H.J. (Peter) van der Waerden
2nd supervisor	:	ir. R.P. (Robert) van Dongen

Colophon

Graduation thesis – Master Construction Management & Engineering

Stimulating cyclists to choose a safer route – A study with a focus on infrastructure

Author

Research conducted by	:	Jacob van Ewijk
Student number	:	1277383
In cooperation with	:	Eindhoven University of Technology
Mail	:	j.v.ewijk@student.tue.nl

Graduation Committee

Chairman	:	prof.dr.ir. B. (Bauke) de Vries
1st supervisor	:	dr.ing. P.J.H.J. (Peter) van der Waerden
2nd supervisor	:	ir. R.P. (Robert) van Dongen

University

Eindhoven University of Technology
Department of the Built Environment
Den Dolech 2, 5612AZ Eindhoven

Location and date

Tienhoven, 13-07-2020

Acknowledgements

After months of hard work I am proud to present my master thesis which is the final part of my master's degree in Construction Management & Engineering. Before I present the content of this report, I would like to thank a number of people who have contributed to making this study possible.

Great appreciation goes to my main supervisor dr.ing. Peter van der Waerden. I am very thankful for his very personal help and professional guidance throughout the project. Apart from the fact that his guidance was very helpful, I also experienced it as very pleasant and enjoyed every meeting. Secondly, I would like to thank my second supervisor ir. Robert van Dongen. His comments have given me valuable insights and have had a positive effect on my understanding of the subject matter. Last, I want to thank my family and friends for their support and suggestions.

Jacob van Ewijk

Tienhoven, 2020

Table of contents

Summary

Over the last couple of decades, cycling has become increasingly popular in the Netherlands and is still gaining popularity. This makes sense, because cycling has many advantages. Cycling has health benefits, environmental benefits, economic benefits and it is often also very practical. Unfortunately, cycling also has disadvantages. The biggest disadvantage of cycling is that it is relatively dangerous compared to other means of transport. Many more Dutch cyclists get seriously injured or even killed than one would expect based on the share cycling has in Dutch mobility.

Since cycling mainly has advantages, the Dutch government has been investing heavily in bicycle-friendly infrastructure since the 1970s. These investments have contributed greatly to making cycling more attractive, but not so much to making cycling safer. This is a problem, because it is definitely important that cycling becomes safer and new investments in bicycle-friendly infrastructure should contribute more to this.

Due to the fact that in recent years a lot of research has been done into the relationship between infrastructure and the safety of cyclists, engineers know exactly how to design infrastructure that is safe for cyclists. It is therefore unlikely that making infrastructure even safer for cyclists is the most efficient solution to make cycling safer in general. Getting cyclists to make more use of infrastructure that is safe for cyclists seems to be a much more efficient way to make cycling safer.

Ensuring that cyclists make more frequent use of infrastructure that is safe for cyclists is not easy. The main reason for this is that there is relatively little of this infrastructure in the Netherlands yet. Cyclists wanting to use new, comfortable and above all safe bicycle routes often have to deviate far from the fastest route and unfortunately this is something they do not like to do. This means that smart solutions have to be found to motivate cyclists to choose a safe route instead of a fast one. In order to provide these solutions it is important to understand which safety related infrastructural attributes motivate cyclists to deviate from the fastest route and to understand how these attributes do this.

This study is about gaining insight in the role that various safety-related infrastructural attributes play in the route choice behaviour of cyclists. The ultimate goal of this study is to provide Dutch transportation planners with information that can help them better understand how route choice behaviour is related to bicycle safety, so that they can ensure that more cyclists are going to use infrastructure that is safe for cyclists.

In order to determine the effect of safety-related infrastructural attributes on the route choice behaviour of cyclists, this study uses the 'shortest route' technique. This is a technique in which driven routes between certain origins and destinations are compared with shortest routes between the same origins and destinations. The differences between the driven routes and the shortest routes form the final dataset that will be used to estimate both a multi-

variable linear regression and a Tobit regression. The difference in distance between the driven routes and the shortest routes is the dependent variable and all safety-related infrastructural attributes are the independent variables.

The dataset that is created for this study consists of roughly two parts. The first part is the road network, including all required information about the infrastructure. This information mainly comes from government agencies such as the Dutch Central Bureau of Statistics, the Dutch Ministry of Infrastructure and Water management (Rijkswaterstaat) and the Municipality of Eindhoven. However, data from the Fietstelweek and from OpenStreetMap is also used. The second part of the dataset is the route data. The route data used for this study comes from the B-Riders programme. From this programme, which aims to stimulate cycling, 145 bicycle routes have been selected that have been driven in the Eindhoven region between January and May 2014. All data files that are used to build the dataset are brought together using GIS software.

The results of this study are based on the estimation of a Tobit model with only the significant variables, because this variant fitted the dataset best. These results offer three important insights. First of all, this study shows that there are multiple safety-related infrastructural attributes that play a role in the extent to which cyclists deviate from the shortest route. These are: traffic intensity, road sides, entrances & exits, trees, and cyclists intensity. A higher percentage of softer road sides, a higher number of entrances & exits and a higher amount of trees along the route cause that cyclists are willing to deviate more than average from the shortest route. A higher traffic intensity, a higher cycling intensity and a higher age cause that cyclists are willing to deviate less than average from the shortest route. The main conclusion that can be drawn regarding the importance of these attributes relative to each other is that traffic intensity has a larger impact on the route choice behaviour of cyclists than road sides do. Secondly, in contrast to what literature suggested, the results of this study show that when infrastructure related attributes positively affect route choice behaviour they do not necessarily positively affect cycling safety as well. This became clear, because this study showed that encountering more entrances and exits (crossings) and higher cyclist intensities are positively related with route choice, whereas recent literature clearly showed that higher levels of these attributes are negatively related with cycling safety. Third and last, this study shows that age also plays a significant role in the extent to which cyclists deviate from the shortest route. By showing that older people are less willing to deviate from the shortest route than younger people, this study underlines the heterogeneity of the population. The effect of the studied attributes on route choice behaviour differs per individual and this is important to consider when designing bicycle infrastructure.

All in all, this study shows that it is possible for transportation planners to improve the cycling safety of a route while at the same time making it more attractive for cyclists. Lowering traffic intensity and increasing the amount of soft road sides seem to be the easiest and most realistic design options to make infrastructure both more attractive for cyclists as well as safer. Planting more trees is an option to attract more cyclists to a route, but does not necessarily makes cycling safer. Furthermore, transportation planners need to keep in mind that the population is heterogeneous, which causes that not everyone's route choice behaviour is equally affected by the safety improving measures that are taken by transportation planners. This makes that it is important to consider the composition of the population when designing safe and attractive cycling infrastructure. Lastly, this study has highlighted multiple factors that should be included in travel models in order to make them reflect cycling behaviour more accurately. Given the fact that there is no literature available about the effect of road sides on route choice behaviour, chances are high that this attribute is not yet considered by at least some travel models. This makes the roadside attribute particularly interesting for transportation planners.

Dutch summary

Fietsen is de afgelopen jaren steeds populairder geworden in Nederland en de populariteit neemt nog steeds toe. Dit is niet verwonderlijk, want fietsen heeft veel voordelen. Fietsen is goed voor de gezondheid, het milieu, de economie en het is vaak ook heel praktisch. Helaas heeft fietsen ook nadelen. Het grootste nadeel van fietsen is dat het relatief gevaarlijk is ten opzichte van andere vervoermiddelen. In het verkeer raken veel meer fietsers ernstig gewond of komen zelfs om het leven dan verwacht zou mogen worden op basis van het aandeel dat fietsen heeft in de Nederlandse mobiliteit.

Vanwege het feit dat fietsen overwegend voordelen heeft, wordt er door de Nederlandse overheid al sinds de jaren 70 veel geïnvesteerd in fietsvriendelijke infrastructuur. Deze investeringen hebben sterk bijgedragen aan het aantrekkelijker maken van fietsen, maar niet zo sterk aan het veiliger maken van fietsen. Dit is een probleem, want het is wel degelijk belangrijk dat fietsen veiliger wordt en nieuwe investeringen in fietsvriendelijke infrastructuur moeten hier ook meer aan gaan bijdragen.

Gezien het feit dat er de afgelopen jaren veel onderzoek is gedaan naar de relatie tussen infrastructuur en de veiligheid van fietsers weten ingenieurs exact hoe zij infrastructuur moeten ontwerpen die veilig is voor fietsers. Het is dan ook onwaarschijnlijk dat het nog veiliger maken van deze infrastructuur de meest efficiënte oplossing is om fietsen in zijn algemeenheid veiliger te maken. Er voor zorgen dat fietsers meer gebruik gaan maken van infrastructuur die veilig is voor fietsers lijkt een veel efficiëntere manier om fietsen veiliger te maken.

Er voor zorgen dat fietsers vaker gebruik gaan maken van infrastructuur die veilig is voor hen is niet eenvoudig. De belangrijkste reden hiervoor is dat er nog relatief weinig van deze infrastructuur is in Nederland. Fietsers die gebruik willen maken van nieuwe, comfortabele en vooral ook veilige fietsroutes moeten hiervoor vaak ver afwijken van de snelste route en laat dit nu net iets zijn dat fietsers niet graag doen. Dit betekent dat er slimme oplossingen moeten worden gevonden om fietsers te motiveren om voor een veilige route te kiezen in plaats van een snelle route. Om deze oplossingen te kunnen bieden is het belangrijk om te begrijpen welke aan veiligheid gerelateerde infrastructurele attributen fietsers motiveren om af te wijken van de snelste route en om te begrijpen hoe deze attributen dit doen.

De functie van deze studie is het verkrijgen van inzicht in de rol die diverse aan veiligheid gerelateerde infrastructurele attributen spelen in het routekeuzegedrag van fietsers. Het uiteindelijke doel van deze studie is om Nederlandse transport planners informatie te verstrekken die hen kan helpen om beter te begrijpen hoe routekeuzegedrag gerelateerd is aan fietsveiligheid, zodat zij er voor kunnen zorgen dat meer fietsers gebruik gaan maken van infrastructuur die voor hen veilig is.

Om te achterhalen wat het effect van infrastructuur op het routekeuzegedrag van fietsers is maakt deze studie gebruikt van de 'kortste route' techniek. Dit is een techniek waarbij gereden routes tussen bepaalde oorsprongen en bestemmingen worden vergeleken met kortste routes tussen diezelfde oorsprongen en bestemmingen. De verschillen tussen de gereden routes en de kortste routes vormen de uiteindelijke dataset die zal worden gebruikt om zowel een multivariabele lineaire regressie als een Tobit regressie te schatten. Hierbij is het verschil in afstand tussen de gereden routes en kortste routes de afhankelijke variabele en zijn alle aan veiligheid gerelateerde infrastructurele attributen de onafhankelijke variabelen.

Het databestand dat gemaakt is voor dit onderzoek bestaat uit ruwweg twee onderdelen. Het eerste onderdeel is het routenetwerk met de bijbehorende informatie over de infrastructuur. Deze informatie komt hoofdzakelijk van overheidsinstanties zoals het Centraal Bureau voor de Statistiek, het Ministerie van Infrastructuur en Waterstaat (Rijkswaterstaat) en de Gemeente Eindhoven. Echter, er is ook data van de Fietstelweek en van OpenStreetMap gebruikt. Het tweede onderdeel van het databestand is de routedata. De routedata die voor dit onderzoek is gebruikt is afkomstig van het B-Riders programma. Uit dit programma, dat als doel heeft om fietsen te stimuleren, zijn 145 fietsroutes geselecteerd die tussen januari en mei 2014 zijn gereden in de regio Eindhoven. De verschillende databestanden zijn bij elkaar gebracht met behulp van GIS software.

De resultaten van deze studie komen voort uit de schatting van een Tobit model met alleen de significante variabelen, omdat deze het beste paste bij de gebruikte dataset. Deze resultaten bieden drie belangrijke inzichten. Allereerst toont deze studie aan dat er meerdere infrastructurele en aan veiligheid gerelateerde attributen zijn die een rol spelen in de mate waarin fietsers afwijken van de kortste route. Dit zijn: verkeersintensiteit, bermen, in- en uitritten, bomen en fietsersintensiteit. Een hoger percentage zachtere bermen, een groter aantal in- en uitritten en een groter aantal bomen langs de route zorgen ervoor dat fietsers meer dan gemiddeld van de kortste route willen afwijken. Een hogere verkeersintensiteit, een hogere fietsintensiteit en een hogere leeftijd zorgen ervoor dat fietsers minder dan gemiddeld van de kortste route willen afwijken. De belangrijkste conclusie die getrokken kan worden met betrekking tot het belang van deze attributen ten opzichte van elkaar is dat verkeersintensiteit een grotere impact op het routekeuzegedrag van fietsers heeft dan bermen. Ten tweede, de resultaten van deze studie tonen aan dat wanneer infrastructuur gerelateerde attributen een positieve invloed hebben op het routekeuzegedrag, zij niet noodzakelijk ook een positieve invloed hebben op de fietsveiligheid. Dit werd duidelijk, omdat uit deze studie bleek dat het tegenkomen van meer in- en uitritten (kruisingen) en hogere fietsersintensiteiten positief gerelateerd zijn aan de keuze voor een bepaalde route, terwijl recente literatuur duidelijk aantoonde dat hogere niveaus van deze attributen negatief gerelateerd zijn aan fietsveiligheid. Ten derde toont dit onderzoek aan dat leeftijd ook een belangrijke rol speelt in de mate waarin fietsers afwijken van de kortste route. Door te laten zien dat ouderen minder geneigd zijn om van de kortste route af te wijken dan jongeren, onderstreept deze studie de heterogeniteit van de bevolking. Het effect van de attributen op het routekeuzegedrag van fietsers verschilt per individu en dit is belangrijk om te mee te nemen bij het ontwerpen van fietsinfrastructuur.

Al met al laat deze studie zien dat het voor transportplanners mogelijk is om de fietsveiligheid van een route te verbeteren en deze route tegelijkertijd aantrekkelijker te maken voor fietsers. Het verlagen van de verkeersintensiteit en het vergroten van het aantal zachte bermen lijken de gemakkelijkste en meest realistische ontwerpopties die de infrastructuur zowel aantrekkelijker als veiliger maken voor fietsers. Meer bomen planten is een manier om een route aantrekkelijker te maken, maar maakt fietsen niet per se veiliger. Bovendien moeten transportplanners er rekening mee houden dat de bevolking heterogeen is, waardoor het routekeuzegedrag van fietsers niet altijd op dezelfde manier en in dezelfde mate wordt beïnvloedt door de veiligheidsbevorderende maatregelen die door transportplanners worden genomen. Dit maakt het belangrijk om bij het ontwerpen van nieuwe fietsveilige infrastructuur rekening te houden met de samenstelling van de bevolking. Tot slot, deze studie heeft meerdere factoren belicht die in routekeuze modellen zouden moeten worden opgenomen om deze fietsgedrag nauwkeuriger te laten weerspiegelen. Gezien het feit dat er nog geen literatuur beschikbaar die het effect van bermen op het routekeuzegedrag van fietsers bespreekt is de kans groot dat dit attribuut nog niet wordt meegenomen in op zijn minst sommige routekeuze modellen. Dit maakt het attribuut bermen bijzonder interessant voor transportplanners.

Abstract

Over the last couple of decades, cycling has become increasingly popular in the Netherlands and is still gaining popularity. Due to the multitude of benefits that cycling has, such as health and environmental benefits, the increasing popularity of cycling is generally considered to be very positive. However, cycling also has one major drawback, which is that the bicycle is a relatively unsafe means of transport. Despite the fact that the Dutch government invested a lot of money in bicycle friendly infrastructure over the last couple of years, cycling safety improved only slightly. In order to increase the positive effect of bicycle friendly infrastructure on cycling safety, it is important that cyclists are going to use this infrastructure more often. For this purpose, better understanding of how safety exactly affects the route choice behaviour of cyclist is required. This study investigates how cyclists can be stimulated to choose a safer route by examining the role that safety-related infrastructural attributes play in the extent to which cyclists deviate from the shortest route. For this effort, the GPS data of 145 cycling trips from the Eindhoven region is used to estimate a multivariable linear regression and a Tobit regression. On average, cyclists deviate 340 meters from the shortest possible route. A higher percentage of softer road sides, a higher number of entrances & exits and a higher amount of trees along the route cause that cyclists are willing to deviate more than average from the shortest route. A higher traffic intensity, a higher cycling intensity and a higher age cause that cyclists are willing to deviate less than average from the shortest route. All in all, this study shows that it is possible for transportation planners to improve the cycling safety of a route while at the same time making it more attractive for cyclists. However, transportation planners need to keep in mind that not everyone's route choice behaviour is equally affected by the safety improving measures that they take.

List of tables

List of figures

1. Introduction

1.1. Problem definition

Over the last couple of decades, cycling has become increasingly popular in the Netherlands. Where Dutch people cycled 14 billion kilometres in 2005, they cycled 15.5 billion kilometres in 2017 (CBS, 2017 in KiM, 2018). The fact that cycling is getting increasingly more popular makes sense, because cycling has many benefits. Cycling is healthy, practical, good for the environment and good for the economy (ANWB, 2020).

Unfortunately, the bicycle is also a relatively unsafe means of transport. Something that is mainly caused by the fact that cyclist are very vulnerability (Reurings et al., 2012). Cyclists nowadays account for approximately one third of the total number of road casualties and over 50% of all serious traffic injuries in the Netherlands, while cycling only accounts for about a quarter of Dutch daily mobility (CBS, 2018; Weijermars et al., 2016; KiM, 2018).

The fact that the safety of cyclists has developed a lot less favourably than that of other road users suggests that cycling safety is relatively hard to improve. Where the total number of road fatalities decreased by approximately 50% over the last 20 years, the number of cycling fatalities remained approximately the same (CBS, 2018). Given the fact that the number of bicycle kilometres travelled has increased by approximately 10% over the past 15 years and that the amount cyclists who got killed in traffic did not rise, it can be concluded that cycling has become safer. However, the improvement is very slight. Especially when compared to the improvement of overall traffic safety.

Due to the fact that cycling has so many advantages, the Dutch government started to actively stimulate the use of the bicycle since the 1970s, which led to increasing investments in bicycle infrastructure (van Goeverden & Godefrooij, 2010). Some recent examples of large investments in the Dutch cycling infrastructure are a large bicycle bridge in Utrecht and various so-called fast cycle routes that have been built throughout the country (Gemeente Utrecht, 2020; Fietsersbond, 2020). These investments in cycling infrastructure have clearly made cycling more attractive by improving cycling comfort and cycling convenience. However, since cycling safety has improved only very little the last couple of decades, their effect on cycling safety appears to be relatively limited. This is a problem, because there are still many cyclists who get seriously hurt or even killed in traffic every year.

In recent years, many studies and reports have focussed on the relationship between infrastructure and cycling safety (e.g. Schepers, 2008; Wijlhuizen & Aarts, 2014). As a result, transportation planners know exactly how to design infrastructure that is safe for cyclists and generally do so. Increasing cycling safety by making new cycling infrastructure even safer does therefore not seem to be a realistic option. It seems that getting cyclists to use new infrastructure that is safe for cyclists more often has a lot more potential to significantly increase cycling safety in general.

Getting cyclists to use newly built infrastructure that is safe for cyclist more often is relatively difficult. The main reason for this is that newly developed infrastructure that is safe for cyclists still forms a relatively small percentage of the total network in the Netherlands and that, in order to use it, cyclists generally have to deviate significantly from their fastest route. Something that cyclists do not like to do as they tend to minimise travel time (Stinson & Bhat, 2003; Yang & Mesbah 2013). This means that smart solutions have to be found that get cyclists to move away from the fastest route and choose a safer route. In order to provide these solutions, it is important to understand which safety-related infrastructural attributes motivate cyclists to deviate from the fastest route and to understand how these attributes do this. Only when this information is available, it is possible to get cyclists to use safer routes.

Despite the fact that a lot of research has been done into cycling safety and that there is a lot of information available on this topic, there is almost no information about how cycling safety relates to the route choice behaviour of cyclists. Various studies (e.g. Broach et al., 2012; Huisman & Hengeveld, 2014) show that cycling safety affects the route choice behaviour of cyclists, but no one yet studied how cycling safety exactly affects the route choice behaviour of cyclists.

Getting more insight into the role that safety-related infrastructural attributes play in the route choice behaviour of cyclists will lead to better understanding of why cycling facilities are used in the way they are used and will ultimately lead to better understanding of how cyclists can be tempted to choose a safer route. This is important because, in the coming years again a lot of money will be invested in cycling facilities in an attempt to realise the Dutch cycling ambitions of making cycling more attractive (Rijksoverheid, 2018). It would be very helpful if these investments contributed more to increasing cycling safety than the previous investments did.

1.2. Research question & objective

Current literature emphasizes the relevance of studying the influence of safety on the route choice behaviour of cyclists (Broach et al., 2012). Broach et al. (2012), who investigated which types of bicycle facilities are preferred by cyclists, found that the value cyclists gave to some facility types could not be explained by the detailed facility variables that they were able to measure. They concluded that something more subtle as perceived safety might explain the results of their study and suggest that this should be studied in future research. The aim of the current study is to contribute to filling this research gap.

Over the past couple of decades, various methods have been used in route choice behaviour studies. All with their own pros and cons, which will be explained in the literature review. One of these methods is to compare the characteristics of actually driven routes between certain origins and destinations with shortest routes between the same origins and destinations (Winters et al., 2010; Larsen & El-Geneidy, 2011). A sound technique since there is evidence that travel time is the most important route choice attribute (Stinson & Bhat, 2003) and that distance is a good proxy for travel time (Rietveld et al., 1999).

As already mentioned in the previous paragraph, this study is about getting insight into the role that safety-related infrastructural attributes play in the route choice behaviour of cyclists. For this endeavour, following the 'shortest path technique' described above, the main research question for this study is formulated as follows:

> "What role do safety-related infrastructural attributes play in the extent to which cyclists deviate from the shortest route?"

In order to answer the main research question, various sub-questions are defined. These are:

1. What is route choice behaviour?
2. What is cycling safety?
3. Which infrastructure related attributes affect cycling safety and the route choice behaviour of cyclists and how do they do this?
4. How can the effect of safety-related infrastructural attributes on the route choice behaviour of cyclists be measured?

The main objective of this study is to provide Dutch transportation planners with information that will help them to design infrastructure that is not only safe but also so attractive that a lot of cyclists are going to use it. The way in which this will be done is by providing insight into the way in which various safety-related infrastructural attributes are related to route choice behaviour of cyclists. The information will be presented in a way that is accessible for Dutch transportation planners.

1.3. Research design

Over the last couple of decades, a lot of research has been done into cycling. In order to make good use of the existing information and to make sure that this study will be a valuable addition to existing literature, this study begins with an extensive literature review. In this literature review, the purpose of which is to elaborate on the research domain and to illustrate the current opinions in the scientific literature, an answer will be given on the first three sub-question of this study. By answering these questions, the literature review will provide a lot of relevant information about cycling safety and the route choice behaviour of cyclists.

The methodology of the current study is determined based on the information presented in the literature review and can be seen as the answer to the fourth and last sub-question of this study. A decisive factor in determining the methodology is the choice to base this study on revealed preference data instead of stated preference data. Something that is not uncommon in route choice behaviour (e.g. Dill & Gliebe, 2008; Broach et al., 2012; Hood et al., 2011). The main reason for using revealed preference data is that this study aims to shed light on how certain infrastructural attributes affect the route choice behaviour of cyclists. For such a situation, it is more accurate to look at which routes cyclists have actually chosen instead of asking cyclists which routes they should have chosen given certain circumstances (stated preference). Choosing for revealed preference data instead of stated preference data takes away the risk of respondents saying that they would make a certain choice, but in reality make a different choice.

Following the choice for revealed preference data, this study will be based on GPS data of routes that people actually cycled to get from an origin to a destination and compare the attributes of these routes with the attributes of alternative routes between those same origins and destinations. This operationalisation makes it possible to investigate how the infrastructural attributes that were identified in the literature review affect the route choice behaviour of cyclists and ultimately find out which of these attributes have the potential to make cyclists choose safer routes.

The results of this study will be generated by using two different regression based model analyses. First a standard linear regression is specified. However, due to the fact that the data that will be used for this study is censored (this will be explained in detail in chapter 3), a more suitable Tobit regression is specified as well (Tobin, 1958).

Finally, after the results have been generated, the final conclusions will be drawn and the main research question will be answered. These conclusions will form a design advice for transportation planners that they can use to make infrastructure both safe and attractive for cyclists and thereby improve general cycling safety.

1.4. Reading guide

This remainder of this study consists of four chapters and this paragraph briefly explains what these chapters are about. Chapter two presents the results of the literature review of this study. This chapter addresses route choice behaviour, cycling safety and how infrastructure, traffic conditions and infrastructure surroundings are related to the route choice behaviour of cyclists and cycling safety. Chapter three presents the details of the methodology. This chapter explains the conceptualisation of the research problem and elaborates on various theories, methods and techniques that are relevant for answering the research questions and ultimately achieving the research objective. Chapter four presents the results of this study and contains the descriptive and statistical analysis of the dataset. Chapter 5 presents the conclusions and recommendations of this study. This chapter answers the research question, discusses the conclusions of this study, provides recommendations for practice and provides recommendations for further research.

2. Literature review

2.1. Introduction

This chapter presents the literature review of this study and consists of 7 paragraphs. The purpose of this literature review is to elaborate on the research domain and to illustrate the current opinions in scientific literature. Paragraph 2.2. zooms in on route choice behaviour. This paragraph explains the principle of route choice behaviour, elaborates on various approaches in route choice behaviour literature and highlights various important aspects of route choice behaviour, such as its determinants. Paragraph 2.3. zooms in on cycling safety. This paragraph explains the principle of cycling safety, elaborates on various approaches in cycling safety literature and addresses the most important determinants of cycling safety. Paragraph 2.4. explains how infrastructure is related to the route choice behaviour of cyclists and cycling safety. More specifically, this paragraph describes how various infrastructural attributes affect the route choice behaviour of cyclists, objective safety (safety in numbers) and subjective safety (safety as perceived). Paragraph 2.5. and 2.6. follow the same structure as paragraph 2.4., but address the traffic situation and infrastructure surroundings instead of infrastructure respectively. Finally, in paragraph 2.7., the conclusions of the literature review are presented.

Since this literature review directly dives into cycling safety, more general information about cycling is provided in appendix A for anyone who would like a more comprehensive introduction to the subject matter. This appendix discusses cycling in the Netherlands and provides information about the benefits of cycling, drawbacks of cycling and the Dutch cycling ambitions.

2.2. Route choice behaviour

This paragraph zooms in on route choice behaviour and focuses specifically on cyclists. In doing so, this paragraph provides a clear picture of what route choice behaviour exactly is and how the route choice behaviour of cyclists has been studied in the past.

2.2.1. The principle of route choice behaviour

Route choice behaviour is the overarching concept that stands for the decision-making process in which people consciously or unconsciously choose a route based on their knowledge of or opinion towards certain characteristics of the route network. There are several theories regarding route choice behaviour. However, the choice theory that is mostly used in route choice modelling is the utility theory. The Saylor Academy (2020) explains on their website that the utility theory is originally developed in economics and aims to explain the observed general behaviour of individuals. A key element of this theory is that it assumes that individuals always try to choose the alternative that will benefit them most (Louviere et al, 2000). A principal that is referred to as utility maximisation. Within route choice modelling research, utility maximisation means that travellers choose the route of which they think that it suits their preferences best, given its attributes. More detailed information about choice behaviour and the utility theory is presented in appendix B.

2.2.2. Approaches in route choice behaviour literature

Over the past couples of decades, many studies on route choice behaviour have been conducted. Although in most cases the aim of these studies (which is mostly to figure out how certain attributes relate to route choice decisions) is very similar, they are all slightly different when they are looked at in more detail.

The first important point on which studies differ is the target area. Many route choice behaviour studies were carried out in typical 'cycling countries' like the Netherlands (e.g. Rietveld et al., 1999; Claasen & Rienstra, 2017), but also in North America a lot of research has been done on this topic (e.g. Winters et al. 2010; Buehler & Pucher, 2012).

The second important point of difference is the type of data that is used. Where most route choice behaviour studies used stated preference data (e.g. Stinson & Bhat, 2003; Larsen & El-Geneidy, 2011), there were also quite a few that used revealed preference data (e.g. Dill & Gliebe, 2008; Broach et al., 2012). Why stated preference is the most commonly used data type in route choice behaviour studies has been clearly explained by Broach et al. (2012). They explain that data collection is often easier, it is not necessary to obtain detailed network data, there is no need to generate realistic choice sets, model specification and estimation is generally easier and it is possible to study non-existent options. However, stated preference in route choice behaviour studies also has some drawbacks, the most important of which is that it is very difficult for respondents to empathise with the hypothetical situation they are presented with and give answers that match their actual behaviour. This problem plays a much smaller role in revealed preference route choice behaviour studies, as they use actual routes and real world network data. The most important shortcomings of these studies is that they generally consider limited choice sets (Broach et al., 2012).

The third important point on which studies differ is the research technique that is used. Within route choice studies different research techniques can be distinguished. A technique that is often used in stated preference studies is discrete choice analysis (e.g. Stinson & Bhat, 2003; Sener et al. 2009). Studies that use this technique generally present respondents with route options from which they can choose. The options are designed in such a way that respondents have to trade off certain features, such as parking levels and travel time. Finally, a discrete choice model, such as the Multinomial Logit Model (MNL), is used to estimate the coefficients connected to the examined attributes. A Technique that is used in revealed preference studies is to compare driven routes with other logical alternatives, such as the shortest route (e.g. Winter et al., 2010; Larsen & El-Geneidy, 2011). Studies that use this technique often ask respondents to recall their routes or are based on GPS data. Smart algorithms are then used to generate alternative routes. Finally, the routes are analysed by means of regression analysis.

The last important point on which studies differ is the specific focus of the research. Some studies focus on trip characteristics (such as commuting or leisure), others on traveller characteristics (such as age and gender) and others on infrastructure characteristics (such as quality of the surface). There are also studies that focus on a combination of these attributes or that incorporate attributes that cannot be placed in the earlier mentioned groups such as the weather.

Even though most route choice behaviour studies are different (different target area, different type of data, different research technique, different focus) the results show many commonalities. This will become clear in paragraph 2.3, 2.4 and 2.5.

2.2.3. Determinants of route choice behaviour

Based on current literature, it becomes clear that route choice behaviour has three main determinants. These are convenience, comfort and safety. Convenience generally refers to the ease of cycling, comfort is mostly about the state of physical and mental ease experienced during cycling and safety is mainly about the risks of getting injured while cycling.

Convenience is probably the most important determinant of route choice behaviour as it is mainly determined by travel time. Despite the fact that there are no studies that directly mention convenience as an important determinant of route choice behaviour, there are many studies that have showed the great importance of travel time and thus indirectly the great importance of convenience. For example, Stinson & Bhat (2003), who used a stated preference survey to investigate commuter bicyclist route choice behaviour, found that travel time is the most important route choice attribute for bicycle commuters. These findings are in line with earlier findings by Bovy & Bradley (1985), who conducted a comparable study. A study conducted by Yang & Mesbah (2013), also shows that travel time is the most important route choice attribute for regular cyclists. They studied the route choice behaviour of cyclists in general based on both stated and revealed preference data.

Based on current research, comfort seems to be an important route choice determinant as well. Despite the fact that only a study by Stinson & Bhat (2003) directly mentions that comfort plays a significant role in the route choice behaviour of cyclists, there are many studies that do so indirectly. An example of a study that indirectly mentions the relation between comfort and route choice behaviour is the one conducted by Bovy & Bradley (1985). Their study shows that various comfort related attributes, such as the quality of the pavement, have a significant impact on route choice behaviour. A more recent study by Sener et al. (2009) shows similar results for parking, continuous bicycle facilities, lower roadway speed limits and the number of traffic lights, stop signs and cross streets. These are also aspects that can be related to comfort. Interestingly, research by Stinson & Bhat (2003) suggest that the role that comfort plays within route choice behaviour is age dependent. The results of their study suggests that older people seem to be much more sensitive to comfort aspects than younger people.

Safety is the third important determinant that affects the route choice behaviour of cyclists. As already explained in the introduction, the bicycle is a relatively unsafe means of transport. People are generally aware of this relative unsafety, which in some cases affects their route choice decisions. This has been demonstrated in a large variety of studies that all look at route choice behaviour from a different perspective and focus on different parts of the world. One of the more recent studies that has showed that safety affects the route choice behaviour of cyclists is conducted in the Netherlands by Claasen & Rienstra (2017). They conducted a written survey among cyclists in Utrecht (Netherlands) in order to assess the influence of the various aspects, such as safety, on the route choice behaviour of cyclists and found safety to be very important. This finding is in line with earlier statements done in a Belgian study by Huisman & Hengeveld (2014). They wrote a report on bicycle counts for the province of Antwerpen and stated that traffic and transport factors, of which safety is one, determine to a large extent the route choice decisions of cyclists. Evidence for the relationship between safety and route choice behaviour was also found in Ireland and Denmark. In these countries, the effect of safety on route choice behaviour was shown in some studies that specifically focused on subjective (perceived) safety. In Ireland, an interesting study on the safety perception of cyclists in Dublin conducted by Lawson et al. (2013), showed clearly that cyclists prefer routes that they perceive as safe and that cyclists are willing to adapt their route choice accordingly. In the same year, in Denmark, a study on bicyclists' experiences in Copenhagen was carried out by Snizek et al. (2013). They also found that subjective safety has a significant influence on the route choice behaviour of cyclists.

Despite the fact that safety, comfort and convenience are three different determinants of route choice behaviour, there appears to be a close link between them. The attributes (e.g. infrastructural attributes) that determine safety, comfort and convenience often do not affect one determinant, but to multiple determinants. A good example of an attribute that shows the overlap between the determinants is quality of pavement as it can influence both comfort and safety.

2.2.4. Travel time in route choice behaviour literature

As already mentioned in the previous section, travel time is generally considered to be the most important attribute in route choice behaviour and its relationship with route choice behaviour of cyclists has been studied many times. This is relatively easy in stated preference studies, because respondents can just be asked to value travel time against other attributes. In revealed preference studies it is a lot harder to incorporate travel time, because this information is generally not available and has to be collected, for example, by recruiting respondents and equipping them with small computer devices (e.g. Dill and Gliebe, 2008).

An easier way to incorporate travel time in revealed preference studies is to assume a linear relationship between travel time and distance and use distance as proxy for travel time. This assumption is supported by Rietveld et al. (1999), who show high correlations between travel time and distance as well as a linear relationship between travel time and distance for short distance/low travel time routes. Despite the fact that their study focused on commuting by car, it is highly likely that their findings also apply to cyclists as they move through traffic in a similar way (i.e. have to stop or slow every now and then, maintain different speeds, etc.). In fact, their findings might even apply more strongly to cyclists as they generally travel shorter distances and maintain a more constant speed.

A fundamental assumption in many route choice behaviour studies is that cyclists generally try to minimize travel time and thus distance. However, this does not mean that cyclists always choose the shortest route. In fact, cyclists only rarely choose the shortest route and this is proven in a few studies that compared chosen routes with shortest routes. For example, Dill & Gliebe (2008) found that for all trips under 16km, cyclist, on average, divert 384 meters from the shortest route. They explain that this represents approximately 1.5 minutes of travel. Similar results were found by Winters et al. (2010), who noticed that cyclists, on average, divert 360 meters from the shortest possible route. These findings suggest that chosen routes have certain characteristics that make it worthwhile to cycle a little further and a little longer. A suggestion that is in line with later findings by Segadilha & Sanches (2014) and van Overdijk (2016). Segadilha & Sanches (2014) found that attributes such as traffic intensity and quality of pavement outweigh trip length and van Overdijk (2016) found that the type of bicycle facility, pavement quality and presence of slopes outweigh travel time.

Despite the fact that the directness of the trip generally seems to be outweighed by other attributes such as pavement quality, this is not necessarily case for everybody. Just as with comfort aspects, the value that people add to the directness of the trip seems to be influenced by age. This has been demonstrated by Bernhoft & Carstensen (2008), who suggest that older people generally add higher value to minimizing the distance they cycle than younger people.

2.3. Cycling safety

This paragraph zooms in on cycling safety and provides a clear picture of what cycling safety exactly is and how cycling safety has been studied in the past.

2.3.1. The principle of cycling safety

Cycling safety is a term that mostly relates to the objective safety level of bicycle use in traffic and can be used to describe the safety risk for cyclists in certain areas or on certain types of infrastructure (ITF, 2018). Objective safety is 'real' safety which can, for example, be measured in terms of cycling related accidents per million inhabitants (Heinen et al., 2010) or by counting the amount of cycling fatalities per billion kilometres cycled (ITF, 2018). A high level of cycling safety generally means that there is a low chance of getting involved in any type of accident while cycling and a low level of cycling safety means that this chance is high. However, it is important to note that indications of cycling safety are often too optimistic as minor/non-fatal accidents often go unreported (Vanparijs et al., 2015).

Despite the fact that the term cycling safety mostly relates to objective safety, subjective safety is also part of cycling safety (Heinen et al., 2010). Subjective safety relates to the way in which individuals perceive safety, and is "mostly measured in terms of the stated safety experience of users or other respondents" (Heinen et al., 2010, Pg. 63). Although being different types of safety, objective and subjective safety do not necessarily have to differ from each other, they can also correspond with each other (Heinen et al., 2010). In fact, research conducted by Sørensen & Mosslemi (2009) & Manton et al. (2013) shows that there often appears to be a close relationship between objective and subjective safety.

So is it objective or subjective safety that influences the behaviour of cyclists? An answer to this question can be found in a study of Parkin (2007), who states that: “While actual, or objective risk, is relatively high for cycling compared with other modes, the perceived risk, that is the risk that is assumed to exist by existing and would-be mode users, is the important criterion in terms of behavioural response” (Pg. 2). This is confirmed by Heinen et al. (2010), who state that “cyclists’ preferences are based on subjective notions of safety” (Heinen et al., 2010, Pg. 65). So, where objective safety is decisive for how safe cycling really is, it is subjective safety that affects the behaviour of cyclists. This means that when investigating the route choice behaviour of cyclists, subjective safety is the determining factor. However, since subjective safety is hard to measure and it seems like there often is a relationship between objective and subjective safety, it can be very useful to investigate objective safety as well.

2.3.2. Approaches in cycling safety literature

In recent years, a lot of research has been done into cycling safety. The literature on this topic is very diverse, but can be roughly divided into two groups: literature that focuses on objective safety and literature that focuses on subjective safety.

The literature that focuses on objective safety generally uses revealed preference data, such as accident data. A good example of such literature is the SWOV report on Monitoring Bicycle Safety written by Wijlhuizen & Aarts (2014). Wijlhuizen & Aarts (2014) used accident data to identify indicators for cycling safety. Another good example is a study about the effects of roundabouts on cycling safety conducted by Daniels et al. (2008). This study used various types of data, such as traffic intensity data and traffic accident data, in order to determine the impact of roundabouts on the safety of cyclists.

In addition to reports and studies, various methods have been developed to assess the objective safety of bicycle infrastructure. The best known and most extensive model is CycleRAP. A Dutch method to assess the safety of bicycle infrastructure inspired by the European branch of iRAP: EuroRAP (Wijlhuizen et al., 2016). CycleRAP is a method to proactively map the safety of cycling infrastructure, without having to use accident data. Instead, experts that are connected to CycleRAP map the safety of cycling infrastructure based on the characteristics of the infrastructure. CycleRAP has been developed for all bicycle paths, bicycle lanes and roads used by cyclists, both inside and outside the built-up area (ANWB, 2018b). According to the ANWB (2018a), CycleRAP is both a method and a certification system to systematically estimate how the shape and layout of a road protect road users. Safety is determined by the risk of an accident and the severity of the outcome. This is expressed in a "Road Protection Score" (RPS). The RPS is a traffic safety indicator for road design, in which the protection is expressed on a five-point scale.

The literature that focuses on subjective safety is mostly based on stated preference surveys. Most of these studies use surveys in which questions are asked about hypothetical situations that are presented in images. For example, a study about the comfort and safety perception of cyclists conducted by Jain et al. (2010). However, there are also stated preference studies that use videos in order to evaluate cycling safety. For example, a study about the risk perception of cyclists conducted by Lehtonen et al. (2016). They presented their respondents with various videos of traffic situations about which they asked questions.

When it comes to assessing subjective cycling safety, various methods have been developed in recent decades. The most relevant one is the Bicycle Compatibility Index (BCI). According to the FHWA (1998), the BCI is a practical instrument that can be used by traffic engineers and other practitioners to determine the level of bicycle compatibility of infrastructure by predicting the perceptions of cyclists for that specific infrastructure. The Bicycle Compatibility Index was developed for all kinds of infrastructure and includes all variables (e.g. lane width) that cyclists typically use to assess the “bicycle friendliness” of infrastructure. The BCI model can be used for the operational evaluation of existing infrastructure, redesign of existing infrastructure, design of new infrastructure and to assess long-range bicycle transportation plans. The BCI was developed based on perspectives of cyclists that were obtained by having them assess and rate infrastructure shown on video with respect to how comfortable they would be cycling there. Subsequently, the reliability of the results obtained from the cyclists was validated in a pilot study (FHWA, 1998).

2.3.3. Determinants of cycling safety

According to recent literature, there are many things that affect cycling safety. However, the most important determinants seem to be the infrastructure itself, the traffic situation, the direct surroundings of the infrastructure and personal safety measures for cyclists (e.g. helmets). Since this study focusses on attributes that, to a greater or lesser extent, can be related to infrastructure it is decided to focus on the first three determinants.

When it comes to the infrastructure itself, current literature shows a very clear relationship between infrastructure and cycling safety. According to Schepers & Klein Wolt (2012), over 50% of the single-bicycle crashes are related to infrastructure. Also when it comes to crossing accidents there is a relationship with infrastructure. According to Wijlhuizen & Aarts (2014), the amount of intersections and roundabouts cyclists come across is of great influence on their objective safety. Regarding subjective safety, the relationship between safety and infrastructure is clear as well. Manton et al. (2013) found that the presence of roundabouts, width of the road lane, presence of a car parking lane, number of junctions passed through, width of the cycling lane and maximum gradient of the road all have a significant impact on the safety perception of cyclists.

Recent literature shows that the traffic situation is also an important determinant of cycling safety. Two aspects of the traffic situation that have often been shown to affect objective and subjective cycling safety are the amount of traffic and the speed of other traffic. For example, by Schepers et al. (2017), who state that an important measure to increase objective cycling safety is to separate traffic. They explain that shifting away motor vehicles from where cycling levels are high helps to increase objective safety, because it reduces the amount and speed of other traffic. When it comes to subjective safety, the amount of traffic and speed of other traffic are important as well. Manton et al. (2013) found that the speed of traffic and the number of cars passing by are highly important in influencing the subjective safety of cyclists.

A last determinant of which recent literature has shown that it affects cycling safety is the surroundings of the bicycle path/road. An aspect of the infrastructure surroundings that has been related to both objective and subjective cycling safety is the presence of trees. Research by Schoon & Blokpoel (2000) and VeiligheidNL (2016) shows that the presence of trees can be related to objective cycling safety and studies by Krabbenborg et al. (2015) and van der Waerden (2018) show a relationship between the presence of trees and subjective cycling safety. Another aspect of the infrastructure surroundings that several studies have shown to affect both objective and subjective cycling safety is land use. For example, Cho et al. (2009) found that the extent to which land use is mixed is related to both the objective and subjective safety of cyclists.

The following three paragraphs each zoom in on one of the three determinants discussed above and explain how these determinants are related to the route choice behaviour of cyclists and cycling safety.

2.4. Infrastructure, route choice behaviour and cycling safety

As already mentioned in the previous paragraphs, there appears to be a strong relationship between infrastructure, the route choice behaviour of cyclists and cycling safety. However, infrastructure is not always one and the same thing and can take various forms. Infrastructure consists of various attributes (e.g. road surface and roadside) that all have their own specific properties (e.g. road surface is asphalt and roadside is grass). These attributes all have their own relation with the route choice behaviour of cyclists, objective cycling safety and subjective cycling safety and this paragraph discusses those relations in more detail. Although for some attributes the information was rather limited, this paragraph is still able to provide a thorough overview.

The attributes that are included in this section are selected on the basis of an extensive body of existing literature. Attributes that, according to existing literature, affect objective cycling safety are incorporated in this paragraph, regardless of whether existing literature also provides evidence for a relationship between these attributes and subjective cycling safety or the route choice behaviour of cyclists. For this reason, for each attribute, first its relationship with objective cycling safety is discussed, then its relationship with subjective cycling safety (if possible) and then the relationship with the route choice behaviour of cyclists (if possible). The attributes are now being discussed one by one.

1. **Quality of the pavement**

<u>Safety – Objective</u>

According to Hendriks (2018), who conducted a study on the causes of cycling accidents, low pavement quality is the most important infrastructural cause of cycling accidents. He explains that accidents are mostly caused by tree roots, holes and uneven pavement. Schepers (2008), who specifically investigated single-bicycle crashes, found that quality of pavement also plays a major role in single-bicycle accidents. He explains that a lack of roughness of the surface seems to be the most important cause of single-bicycle crashes, but that bumps and pits are important causes as well.

Safety – Subjective
A recent study by Lawson et al. (2013) shows that quality of the pavement has a large influence on the subjective safety of cyclists. In their study, 81% of the respondents indicated that a low pavement quality has a major negative impact on their safety perception. These results are similar to the findings of studies elsewhere (Doherty et al., 2000).

Route choice behaviour
Research by Stinson & Bhat (2003) suggests that quality of pavement is a very important determinant in the route choice behaviour of cyclists. In their study, that specifically focussed on commuter cyclists, they found that cyclists prefer even surfaces over uneven surfaces. Comparable result were found by Segadilha & Sanches (2014), who conducted a similar study. However, they not only found that quality of pavement plays a major role in the route choice behaviour of commuter cyclists, but that type of pavement does so as well.

2. Cycling facility
Studies about cycling safety distinguish between various types of cycling facilities, which can be roughly divided into two groups. The first group contains cycling facilities that separate cyclists from other traffic and includes one-way and two-way bicycle paths. The second group contains cycling facilities that do not separate cyclists from other traffic and includes bicycle lanes and mixed traffic situations. These different types of bicycle facilities all have different effects on cycling safety.

Safety – Objective
Research by Thomas & DeRobertis (2013) shows that cycling facilities that separate cyclists from other traffic significantly reduce the risk of bicycle-motor vehicle crashes compared to cycling facilities that do not separate cyclists from other traffic. Their study also showed that one-way bicycle paths are generally safer than two-way bicycle paths. However, two-way bicycle paths are still significantly safer than roads without facilities (Lusk et al. 2011).

Safety – Subjective
In recent years a number of studies has investigated the effect of the type of cycling facility on the subjective safety of cyclists. According to Manton et al. (2013) and Parkin et al. (2007), separated cycling infrastructure increases subjective cycling safety significantly. However, Lawson et al. (2013) found that this is only the case for inexperienced cyclists and that regular, confident and experienced cyclists prefer to cycle on the road. When it comes to the effect of on-road cycle lanes on the subjective safety of cyclists, Parkin et al. (2007) found that these contribute only very little to the moderation of subjective risk compared to situations with no facility. Though, research by Dollisson et al. (2013) shows that most cyclists do feel safer on a designated on-road bicycle lane.

Route choice behaviour
According to Stinson & Bhat (2003), the type of cycling facility is an important determinant in the route choice behaviour of cyclists. Their study shows that cyclists prefer bicycle lanes and bicycle paths over routes without designated cycling facilities. Broach et al. (2012) go into a

bit more detail and explain that cyclists prefer separated cycle paths the most. In addition, they found that bicycle lanes are only preferred when quiet neighbourhood roads are no option. Larsen & El-Geneidy (2011) also found that the type of cycling facility affects the route choice behaviour of cyclists and explain that greater separation from vehicle traffic generally positively affects trip distance. However, their results also show that this effect only seems to apply to recreational cyclists and not to regular and frequent cyclists.

3. Parking

Safety – Objective

Accidents between cyclists and parked or parking cars can be roughly divided into three categories: Accidents in which cyclists crash into a parked car, accidents between cyclists and cars that are parking and accidents in which cyclists crash into a car door that is opened in their path (Schepers, 2008).

When it comes to accidents in which cyclists crash into an opening car door, Johnson et al. (2013) explain that these are more frequent every year. Furthermore, they explain that this type of accident, also known as being 'doored', can lead to serious injuries with fatal outcomes and that it seems that infrastructure plays a significant role in these kind of accidents. Measures like physically separating cyclists from car traffic or making sure that there is enough space for cyclists to safely avoid opening car doors are likely to reduce 'dooring' accidents (Schepers, 2008; Johnson et al., 2013). Regarding accidents in which cyclists crash into a parked car, Schepers (2008) explains that the absence of a parking lane or parking spaces contributes to collisions between cyclists and parked vehicles, because then the cars are not parked in a straight line which causes that there is no straight line for cyclists. Accidents between cyclists and cars that are parking appear to be relatively rare (Isaksson-Hellman, 2012). No infrastructural cause for this type of accident has been identified in recent literature.

Safety – Subjective

According to Manton et al. (2013), parking also effects subjective cycling safety. They explain that parking along roads negatively impacts the subjective safety of cyclists. This is in line with earlier findings from Jain et al. (2010) and Parkin et al. (2007), who noticed that when the amount of cars that are parked along the road is lower, the level of subjective safety is higher.

Route choice behaviour

Existing research suggests that parking plays a significant role in the route choice behaviour of cyclists. For example, Stinson & Bhat (2003) found that cyclists generally tend to avoid routes on which parking is permitted. They suggest that this might be caused by the negative safety effects of parking and explain that parked cars can limit sight at intersections and cyclists who cycle past parked cars risk getting doored. Similar results were found by Sener et al. (2009), who go into a bit more detail and also explain that cyclists prefer routes with angled parking over routes with parallel parking.

Furthermore, Stinson & Bhat (2003) found that parking does not affect everyone's route choice behaviour equally. For people who live in urban/sub-urban areas and younger people parking is less of a concern. Sener et al. (2009) also found different effects for males and females.

4. Street lights

Safety – Objective

The effect of street lighting on objective cycling safety has hardly been investigated. However, one study did. According to Kim et al. (2007), street lights appear to have a substantial positive effect on the safety of cyclists at night.

Route choice behaviour

According to Segadilha & Sanches (2014), the presence of street lights plays a major role in the route choice behaviour of cyclists. They explain that cyclists generally prefer routes that are illuminated.

5. Obstacles

Safety – Objective

Obstacles on the road play a significant role in accidents with cyclists (Schepers, 2008; Hendriks, 2018). These objects are mostly poles, but also obstacles that indicate a narrowing of the road are the cause of a significant number of accidents. According to Schepers (2008), most of these accidents happen, because cyclists did not see the obstacles or saw them too late. He explains that possible reasons for this are that:

- The colours of the obstacle do not sufficiently contrast with the background or are insufficiently illuminated in darkness;
- Introductory markings are missing;
- There is not enough space between or next to the obstacles.

According to Hendrik (2018), another possible reason for cyclists crashing into obstacles is that other cyclists obstruct the view of the obstacle.

6. Curbs and edges

Safety – Objective

Curbs and other edges are an important cause of cycling accidents (Schepers, 2008; Hendriks, 2018). A study by Hendriks (2018) shows that curbs and other edges play a role in about 14% of the accidents that happen with cyclists in the Netherlands. Furthermore, Hendriks (2018) explains that important reasons why these accidents happen are that curbs and edges are often poorly visible and that cyclists misjudge the height of curbs and edges. When it comes to single bicycle crashes, curbs and edges play a role in about 13% of the accidents in the Netherlands (Schepers, 2008).

7. Roadside

<u>Safety – Objective</u>

One of the few studies that describes the effect of road sides on cycling safety is a study from Schepers (2008) on single-bicycle crashes in the Netherlands. In his study, Schepers (2008) found that approximately 7% of all single-bicycle accidents in the Netherlands are accidents in which people end up on the roadside. He explains that the main reasons why people that end up in the roadside get an accident are height difference between the road and the roadside, the absence of specific type of roadside that allows people to cycle on and obstacles in the roadside.

8. Road width

<u>Safety – Objective</u>

According to Wijlhuizen & Aarts (2014), the width of the road is an important infrastructural factor affecting both single-bicycle crashes and crashes with other road users. They explain that road width influences the chance that cyclists end up in the roadside, collide with an obstacle or collide with another cyclist. Schepers (2008) explains that when the effective width of the path or lane is reduced, cyclists cycle closer to the curb which increases the chance of a collision with the curb. According to Hendriks (2018), reduction of the effective width of the cycle path or lane also makes that dodging and overtaking is more dangerous.

<u>Safety – Subjective</u>

Recent literature shows that road width also affects subjective cycling safety. Manton et al. (2013), who conducted a study on perceptions of cycling safety in Ireland, found that when the width of the road increases, the subjective safety of cyclists generally does as well.

<u>Route choice behaviour</u>

Research by Sener et al. (2009), who investigated a few very specific road situations, shows that road width has only a minor impact on the route choice behaviour of cyclists. Interestingly, Segadilha & Sanches (2014) found the effect of road width on the route choice behaviour of cyclists to be rather large. A possible explanation for the discrepancy between these results could be that Sener et al. (2009) only looked at very specific situations, whereas Segadilha & Sanches (2014) just asked cyclists about their general preferences in a survey. Unfortunately, both studies were not able to draw conclusions about the direction of the effect.

9. Intersections

There are many different types of intersections in the Netherlands. They differ from each other on the basis of their design (amount of lanes, separation of traffic, etc.) and how they regulate traffic (priority, traffic lights, etc.). This attribute does not relate to one type of intersection in particular, but relates to multiple aspects of intersections.

<u>Safety – Objective</u>

Over the past 10 years, various studies and reports have investigated or incorporated the effect of intersections on the safety of cyclists. Despite the fact that most of these studies have a different focus, they all draw the same conclusion: Intersections play a major role in cycling safety.

One thing that has been studied quite a lot is how intersections that regulate traffic in a different way relate to cycling safety. Kroeze et al. (2010) found that priority intersections are the most dangerous for cyclists, followed by signalised intersections and intersections without priority. Crossing side roads seems to have a large influence on cycling safety as well. Due to the fact that cyclists pass side road very often, on a network level, crossing side roads appears to be even more dangerous than crossing priority intersections (Kroeze et al., 2010). These findings contrast with findings from Schepers & Voorham (2010). Their study suggests that it is not the way in which intersections regulate traffic, but the design and the location in the network that are important.

Speed reducing measures seem to have a positive effect on crossing accidents. Intersections with a speed inhibitor for traffic coming from a side road have fewer cyclist crashes than intersections without a speed inhibitor (Schepers & Voorham, 2010; Reurings et al., 2012). Raising intersections on a plateau also seems to positively affect cycling safety (Schepers & Voorham, 2010; Schepers et al., 2011). However, at intersections with cycle paths only, placing the intersections on a plateau seems to result in more accidents with cyclists (Schepers & Voorham (2010).

When it comes to the width of the road that has to be crossed, research by Schepers & Voorham (2010) suggests that cycling safety is not related to the width of the road that has to be crossed.

Safety – Subjective

Recent literature shows that intersections affect subjective cycling safety. For example, a study conducted by Ng et al. (2017) shows that priority intersections negatively influence the subjective safety of cyclists. However, intersections where cyclists have to give way seem to impact subjective safety less than intersections where cyclists should get priority.

When it comes to intersections with signals, the relationship with cycling safety is also evident. A study by Jain et al. (2010) shows that intersections with signals have a large negative impact on the subjective safety of cyclists.

Route choice behaviour

According to Stinson & Bhat (2003) intersections play a significant role in the route choice behaviour of cyclists. The results of their study show that the fewer large intersections a route contains, the more likely it is that this route is chosen. Cyclist in urban and suburban areas and younger cyclists seem to be less adverse to major intersections.

When it comes to the effect of stop signs and traffic lights on the route choice behaviour of cyclists, Stinson & Bhat (2003) found that these only play a relatively small role. This is an interesting finding, because they also found that travel time does play a large role. This could indicate that people do not see avoiding stop signs and traffic lights as a way to reduce travel time. Also Dill & Gliebe (2008) and Segadilha & Sanches (2014) found that the effect of stop signs and traffic lights on the route choice behaviour of cyclists is relatively small compared to other attributes.

10. Roundabouts

Safety – Objective

Roundabouts generally have a positive effect on safety (Daniels et al., 2008). However, the specific way in which roundabouts are designed negatively impacts cycling safety (Daniels and Wets, 2005). According to Daniels et al. (2008), a lot more accidents with cyclists happen on roundabouts than can be expected based on the way in which cyclists occur in overall traffic. However, the negative effects of roundabouts on cycling safety seem to be much larger for roundabouts that are located inside the built-up area as compared to the ones located outside the built-up area (Daniels et al., 2008).

Safety – Subjective

There are not a lot of studies that address the effect of roundabouts on subjective cycling safety. However, a relatively recent study conducted by Jain et al. (2010) suggests that the presence of roundabouts negatively impacts the safety perception of a significant number of cyclists. Whether or not roundabouts have cycling facilities does not seem to make a difference for the effect they have on the subjective safety of cyclists (Parkin et al., 2007).

Route choice behaviour

The specific effect of roundabouts on the route choice behaviour of cyclists has not been studied a lot. However, according to Segadilha & Sanches (2014), roundabouts play a significant role in the route choice behaviour of cyclists. This makes sense because roundabouts are also some sort of intersection and intersections play a significant role in the route choice behaviour of cyclists (Stinson & Bhat, 2003).

Unfortunately, there are no studies that have elaborated on what the exact effect of roundabouts on the route choice behaviour of cyclists is yet. However, since roundabouts are some sort of intersection and a lower amount of intersections increases the chance that a route is chosen, it is likely that a lower amount of roundabouts will also increase the chance that a route is chosen.

11. Road markings

Safety – Objective

Studies that investigated the effect of road markings on objective cycling safety can be roughly divided into two groups: studies that focus on line and signal markings and studies that focus on colouring entire road surfaces.

When it comes to the objective safety benefits of line and signal marking for cyclists, the results of recent studies are positive. According to Schepers & Brinker (2011) and Fabriek et al. (2012), line and signal markings help increase the safety of cyclists by keeping them on track and by helping them to avoid obstacles.

The results of studies on the safety benefits of colouring entire road surfaces are mixed. When it comes to coloured cycling crossings, research by Thomas & DeRobertis (2013) suggests that they only appear to improve cycling safety if the colour is only applied to one side of a four-sided intersection. They explain that this is likely caused by the fact that the coloured parts lose the property of being clearly distinguishable if applied to multiple sides of an intersection. With regard to coloured bike lanes, Kroeze et al. (2010) found that there are significantly more accidents with cyclists at intersections where the main road has a red bicycle lane. They furthermore found that also priority roads with separate bicycle paths that have good markings (blocks, not worn out) and a red colour at side roads are significantly more unsafe for cyclists. This finding is in line with Reurings et al. (2012), who found that the number of accidents with crossing cyclists is smaller at side roads where no colour or marking has been applied.

<u>Safety – Subjective</u>
According to Fabriek et al. (2012), road markings generally increase the subjective safety of vulnerable cyclists (visually impaired and older cyclists). The results of their study show that cyclists' feelings of safety are worse in situations where the visibility of obstacles and the road's course is low. Fabriek et al. (2012) state that visibility, and thus the feeling of safety, can be enhanced by applying high contrast road markings.

<u>Route choice behaviour</u>
Winters et al. (2010) found that the presence of road markings is a factor that affects the route choice behaviour of cyclists. Their study shows that cyclists are willing to deviate from the shortest route in order to cycle on a road with more markings.

An overview of the information presented in this paragraph will be given in paragraph 2.7.

2.5. The traffic situation, route choice behaviour and cycling safety

This paragraph describes how the traffic situation affects cycling safety and the route choice behaviour of cyclists. Similar to infrastructure, the traffic situation consists of various attributes that all have their own relation with objective cycling safety, subjective cycling safety and the route choice behaviour of cyclists. These are now being discussed one by one.

1. Traffic intensity

This factor relates to the intensity levels of motorised traffic.

<u>Safety – Objective</u>
According to recent literature, the amount of traffic on a particular section of infrastructure has a large influence on the objective safety of cyclists on that same section of infrastructure. According to Wijlhuizen & Aarts (2014), the amount of traffic affects the amount of cycling accidents and casualties directly. Schepers et al. (2017) go into a bit more detail and explain

that the amount of traffic affects cycling safety, because the number of motor vehicles encountered by cyclists greatly determines their exposure to risk. They state that "volumes are important because it has been found, at different levels such as intersections, road sections and jurisdictions, that the amount of bicycle and motor vehicle traffic affect the likelihood of bicycle-motor vehicle crashes" (Jacobsen, 2003 and Elvik, 2009 in Schepers et al., 2017, Pg. 266).

Safety – Subjective

That the amount of traffic affects subjective cycling safety has been shown in various studies. An example of such a study is that of Manton et al. (2013). Their study showed that when more cars pass a cyclist, perceived cycling safety is lower. A finding that is in line with earlier research by Chirstmas et al. (2010), who found that when the traffic volume increases, the amount of stress that cyclists experience increases as well. Also Stone and Gosling (2008) found that the perceived safety of cyclists decreases as traffic volumes increase.

Route choice behaviour

Current research shows that there is definitely a link between traffic intensity and the route choice behaviour of cyclists. According to Stinson and Bhat (2003), cyclists prefer more quiet residential streets over arterial roads. They found that cyclists on average are willing to endure 10% extra travel time if they can cycle on residential roads instead of arterial roads and suggest that this might be caused by the fact that arterial roads are generally busier and therefore more unsafe. Also Sener et al. (2009) found evidence for the link between traffic intensity and the route choice behaviour of cyclists. Their findings suggest that traffic intensity is one of the most important attributes in the route choice behaviour of cyclists and that cyclists generally try to avoid high traffic intensities. Similar results were found by Winters et al. (2010), who noticed that chosen routes have more traffic calming features than shortest routes. This indicates that cyclists prefer routes with lower traffic intensities and that they are willing to travel a bit further for quieter roads.

2. Modal split

This attribute relates to the mixture of transport modes.

Safety – Objective

The modal split (in particular the ratio between cyclists and motor vehicles) affects objective cycling safety in a similar way as the traffic intensity does. Schepers et al. (2017) explain that the modal split relates to the amount of motor vehicles that cyclists come across and thereby affects objective cycling safety. They suggest that a larger proportion of motor vehicles leads to higher risks for cyclists.

Another reason why the modal split affects the objective safety of cyclists is that more cyclists on the road causes that motorists behave more safely towards cyclists (Jacobsen, 2003). "A motorist is less likely to collide with a person walking and bicycling when there are more people walking or bicycling" (Jacobsen, 2003, Pg. 208). This means that, more cyclist means safer cycling.

Route choice behaviour

The effect of the modal split on the route choice behaviour of cyclists has not been studied specifically. However, there are various indications that suggest that the modal split does play a role. One important indication is that cyclists tend to choose routes with segregated cycling facilities (explained in paragraph 2.3.). This means that cyclists prefer routes with lower amounts of motorised traffic and relatively high amounts of cyclists. However, there is also evidence that cyclists dislike busy cycle paths (Krabbenborg et al., 2015). Another indication that the modal split is an important attribute is provided by Segadilha & Sanches (2014), who interviewed cyclists in Brazil and found that the amount of trucks and busses encountered are the most important attributes in the route choice behaviour of cyclists. Despite the fact that their study does not shed any light on how these attributes exactly affect the route choice behaviour of cyclists, their study does seem to indicate that the effect of motor vehicles on the route choice behaviour of cyclists is also determined by the type of motor vehicles.

3. Speed differences

This attribute relates to the size of the speed differences between cyclists and other road users.

Safety – Objective

Wijlhuizen & Aarts (2014) explain that speed differences between cyclists and motor vehicles play an important role in the severity of cycling accidents. Cyclists are very vulnerable and large speed differences between cyclists and motor vehicles can therefore easily lead to serious injuries. Another indication that lower speed differences lead to more safety for cyclists is the fact that speed-reducing measures and separating traffic flows have a positive effect on cycling safety (Schepers et al., 2017).

Safety – Subjective

According to Manton et al. (2013) and Vandebona & Kiyota (2001), the level of speed differences also affects subjective cycling safety. Both studies seem to suggest that higher speed differences generally result in lower levels of subjective cycling safety. However, a more recent study about cycling safety conducted in Zeeland (Netherlands) by Lankhuijzen, et al. (2016) showed something really different. Their study showed that car traffic that passes cyclists very fast in most cases does not negatively affect the safety perception of cyclists. A possible explanation for these contrasting results are differences in cycling experience between the sample populations.

Route choice behaviour

Research by Sener et al. (2009) shows that the level of speed differences has a significant effect on the route choice behaviour of cyclists. They found that cyclists generally prefer roads with lower speed limits, but that this effect is stronger for inexperienced cyclists than for experienced cyclist. Sener et al. (2009) suggest that a possible cause for this difference is that experienced cyclists are more comfortable riding with vehicles travelling at higher speeds and see the health benefits from riding at higher speeds. However, even experienced cyclists avoid roads with a high speed limits as these are substantially more dangerous than roads with low or moderate speed limits. A later study by Segadilha & Sanches (2014) confirms that traffic speed is an important attribute in the route choice behaviour of cyclists.

An overview of the information presented in this paragraph will be given in paragraph 2.7.

2.6. Infrastructure surroundings, route choice behaviour and cycling safety

This paragraph describes how the infrastructure surroundings affects cycling safety and the route choice behaviour of cyclists. The infrastructure surroundings consist of various attributes that all have their own relation with objective cycling safety, subjective cycling safety and route choice behaviour. These are now being discussed one by one.

1. Land use

<u>Safety – Objective</u>

The existence of a relationship between land use and the objective safety of cyclists has been suggested in literature from different countries. For example in the Netherlands. A Dutch report on land use planning and safer transportation network planning, states that land use planning can have an important influence on general traffic safety (Hummel, 2001). Thirteen years later, a statement with a similar meaning is done by Schepers et al. (2014), who stated that "adapting land use and infrastructure is a means for governments to improve cycling safety and increase bicycle use" (Schepers et al., 2014, Pg. 331). In Denmark, the relationship between land use and objective cycling safety has been proven as well. In a study about the effect of land use and network effects on the frequency and severity of bicycle-motor vehicle crashes in the Copenhagen region, it was found that land use is associated with the number of bicycle-motor vehicle crashes (Kaplan & Prato, 2015). Cho et al. (2009) found evidence for the relationship between land use and objective cycling safety in the US. In their study about the role of the built environment in explaining relationships between perceived and actual pedestrian and bicyclist safety, they found that mixed land use is positively related with the objective safety of cyclists.

<u>Safety – Subjective</u>

According to Cho et al. (2009), land use influences subjective cycling safety. They explain that mixed land use decreases the perception of crash risk and thereby has a positive effect on subjective cycling safety. However, Götschi et al. (2018), who evaluated cycling infrastructure, did not find a relationship between land use and subjective cycling safety.

<u>Route choice behaviour</u>

The effect of land use on the route choice behaviour of cyclists is only studied by Winters et al. (2010), who did not find any proof that this attribute affects the route choice behaviour of cyclists.

2. Urban density

Safety – Objective

The existence of a relationship between urban density and objective cycling safety has only recently been demonstrated by Cho et al. (2009), who found that a higher urban density is positively related with bicycle-motor vehicle crashes.

Safety – Subjective

According to Cho et al. (2009), urban density affects subjective cycling safety. The results of their study, which is based on a North American survey, suggest that a higher urban density negatively influences subjective cycling safety.

Route choice behaviour

One of the few studies that investigated the effect of urban density on the route choice behaviour of cyclists is one conducted by Winters et al. (2010). Winters et al. (2010) did not find any proof that urban density affects the route choice behaviour of cyclists.

3. Trees

Safety – Objective

The existence of a relationship between trees and cycling safety has been proven in multiple accident data based studies, but appears to be very weak. Schoon & Blokpoel (2000) conducted a study on the frequency and causes of single bicycle crashes in the Netherlands. They found that of all 1617 single-bicycle crashes they examined, 201 (12%) were collisions with an object. Of this 201 collisions, 12 collisions (6%) were collisions with a tree.

Almost two decades later, in 2016, another study on cycling accidents was conducted in the Netherlands by VeiligheidNL (2016). This study did not only look at single-bicycle crashes, but looked at all types of cycling crashes. VeiligheidNL (2016) found that off all cycling accidents less than 1% is a collision with a tree or bush.

Safety – Subjective

Recent literature shows that the presence of trees can have both positive and negative effects on the subjective safety of cyclists. According to van der Waerden (2018), trees and bushes right next to the road or cycle path have a negative effect on subjective cycling safety. However, the presence of trees in general seems to contribute positively to subjective cycling safety (Krabbenborg et al., 2015).

Route choice behaviour

Research by Segadilha & Sanches (2014) shows that the presence of trees has a moderate effect on the route choice behaviour of cyclists. Krabbenborg et al. (2015) went into a bit more detail and found that the presence of trees along the route positively relates to that route being chosen. Interestingly, Winters et al. (2010), who studied 'greenness' in general, did not find a relation between greenness and the route choice behaviour of cyclists.

2.7. Conclusion

Over the past couple of decades a lot of research has been done into both cycling safety and the route choice behaviour of cyclists. The studies devoted to this subject are very diverse, cover almost all aspects of these topics and offer great insights. However, on some points they fall short. One of these points is the relationship between the route choice behaviour of cyclists and cycling safety. As already mentioned in the introduction, many studies suggest that safety affects the route choice behaviour of cyclists, but none of these studies explains how safety exactly affects the route choice behaviour of cyclists. This literature review provides a solid basis for filling in this research gap by bringing information on cycling safety together with information about the route choice behaviour of cyclists. In doing so, the focus was on infrastructure related attributes.

The literature review provided answers to the first three sub-question of this study. These answers are presented below.

1. *What is route choice behaviour?*

Route choice behaviour is the overarching concept that stands for the decision-making process in which people consciously or unconsciously make route choices based on their knowledge or opinion towards certain characteristics of the route network. The theory that is mostly used in studies on route choice behaviour is the utility theory. A theory that assumes that individuals always try to choose the alternative that will benefit them most.

At the highest level of abstraction, route choice behaviour is determined by trip characteristics, traveller characteristics and road characteristics. When it specifically comes to the road characteristics, the main determinants are convenience, comfort and safety. Overall, it seems that travel time is the most important attribute of route choice behaviour.

2. *What is cycling safety?*

Cycling safety is a term that mostly relates to the objective safety level of bicycle use in traffic and can be used to describe the safety risk for cyclists. Objective safety is factual safety and can, for example, be measured in terms of cycling related accidents per million inhabitants or by counting the amount of cycling fatalities per billion kilometres cycled. A higher level of cycling safety generally means that the chances of getting an accident are lower. However, subjective safety can also be part of cycling safety. Subjective safety relates to the way in which individuals perceive safety, and is mostly measured based on stated experience. Despite being different types of safety, there often appears to be a close relationship between objective and subjective safety. Where objective safety is decisive for how safe cycling really is, it is subjective safety that affects the behaviour of (potential) cyclists.

There are many things that have an influence on cycling safety. However, the main determinants seem to be the infrastructure itself, the traffic situation, the direct surroundings of the infrastructure and personal safety measures for cyclists (e.g. helmets).

3. *Which infrastructure related attributes affect cycling safety and the route choice behaviour of cyclists and how do they do this?*

The answers to these two questions are presented in the two tables below. The first table (Table 1) shows whether an attribute is related to cycling safety, the route choice behaviour of cyclists or both. In doing so, a distinction is made between objective safety and subjective safety. The second table (Table 2) presents how the infrastructure related attributes affect cycling safety and the route choice behaviour of cyclists by explaining under what conditions these attributes have a positive effect on cycling safety and the route choice behaviour of cyclists (i.e. the attractiveness of a route). The two most important conclusions that can be drawn based on these tables are that most of the attributes that affect cycling safety also affect the route choice behaviour of cyclists and that when they do the direction of the effect is often similar.

Table 1: Attribute relations with cycling safety and route choice behaviour

Attribute Group	Attribute		Objective safety	Subjective Safety	Route choice behaviour
Infrastructure	1	Quality of pavement	+	+	+
	2	Cycling facility	+	+	+
	3	Parking	+	+	+
	4	Street lights	+	-	+
	5	Obstacles	+	-	-
	6	Curbs and edges	+	-	-
	7	Roadside	+	-	-
	8	Road width	+	+	+
	9	Intersections	+	+	+
	10	Roundabouts	+	+	+
	11	Road markings	+	+	+
Traffic conditions	12	Traffic intensity	+	+	+
	13	Modal split	+	-	+
	14	Speed differences	+	+	+
Infrastructure surroundings	15	Land use	+	+	-
	16	Urban density	+	+	-
	17	Trees	+	+	+

Note: + means that there is evidence for a relationship, - means no evidence is found

Table 2: Attribute effects on cycling safety and route choice behaviour

Attribute group	Attribute		Positive effect on cycling safety when	Positive effect on route choice behaviour when (i.e. the attractiveness of a route)
Infrastructure	1	Quality of the pavement	The surface is smoother and less damaged	The surface is smoother and less damaged
	2	Cycling facility	The level of traffic segregation is higher	The level of traffic segregation is higher
	3	Parking	The amount of parking is lower	The amount of parking is lower
	4	Street lights	The road is better lit	The road is better lit
	5	Obstacles	There are less obstacles	Effect unknown
	6	Curbs and edges	There are less curbs and edges	Effect unknown
	7	Roadside	The roadside is softer and with fewer objects	Effect unknown
	8	Road width	The road is wider	The road is wider
	9	Intersections	The amount of intersections is lower	The amount of intersections is lower
	10	Roundabouts	The amount of roundabouts is lower	The amount of roundabouts is lower
	11	Road markings	The amount of road markings is higher	The amount of road markings is higher
Traffic conditions	12	Traffic intensity	The traffic intensity is lower	The traffic intensity is lower
	13	Modal split	The share of cyclists is higher	The share of cyclists is higher
	14	Speed differences	The speed differences between road users are lower	The speed differences between road users are lower
Infrastructure surroundings	15	Land use	When land use is more diverse	Effect unknown
	16	Urban density	When urban density is lower	Effect unknown
	17	Trees	Effect unclear	More trees along the route

3. Methodology

3.1. Introduction

This chapter explains the conceptualisation of the research problem and elaborates on various theories, methods and techniques that are relevant for answering the research question and ultimately achieving the research objective. Paragraph 3.2. explains how the research problem is conceptualised. This is done on both a high and a low level of abstraction. Paragraph 3.3. elaborates on the specification and operationalisation of the research problem. A key element of this study. Paragraph 3.4. discusses statistical modelling. First some information is presented about statistical modelling in general as well as how it is used within route choice behaviour studies. Subsequently, two statistical models that are interesting for this study are presented after which it is decided which of these models fits this study best. Paragraph 3.5. zooms in on the data collection. This paragraph explains how the road network, route and attribute data are collected. Paragraph 3.6. presents the operationalisation of the included attributes. This paragraph elaborates on the general operationalisation approach, describes in detail how the attributes are operationalised and provides information about the excluded attributes. Lastly, in paragraph 3.7., the process of creating the final dataset is described in five steps. This paragraph explains how the raw data files are prepared, how the prepared data files are combined into one large dataset, how the large dataset is cleaned of irrelevant data, how the large dataset is enriched which missing data and how the large dataset is used to generate the final dataset that will be used for further analysis.

3.2. Conceptualisation of the research problem

For the purpose of clarity, the research problem is first conceptualised on a high level of abstraction and then on a low level of abstraction.

3.2.1. High level of abstraction

The literature review provides three important insights that contribute to placing the research problem in the bigger picture and thereby help to understand the context of the research problem. The first insight is that trip characteristics, traveller characteristics and road characteristics affect the route choice behaviour of cyclists at the highest level of abstraction. The second insight is that road characteristics consist of convenience related road characteristics, comfort related road characteristics and safety related road characteristics. The third and last insight is that road characteristics that relate to safety consist of infrastructure attributes, traffic situation attributes and infrastructure surroundings attributes. A visual representation of these findings and thereby a conceptualisation of the research problem on a high level of abstraction is presented in figure 1.

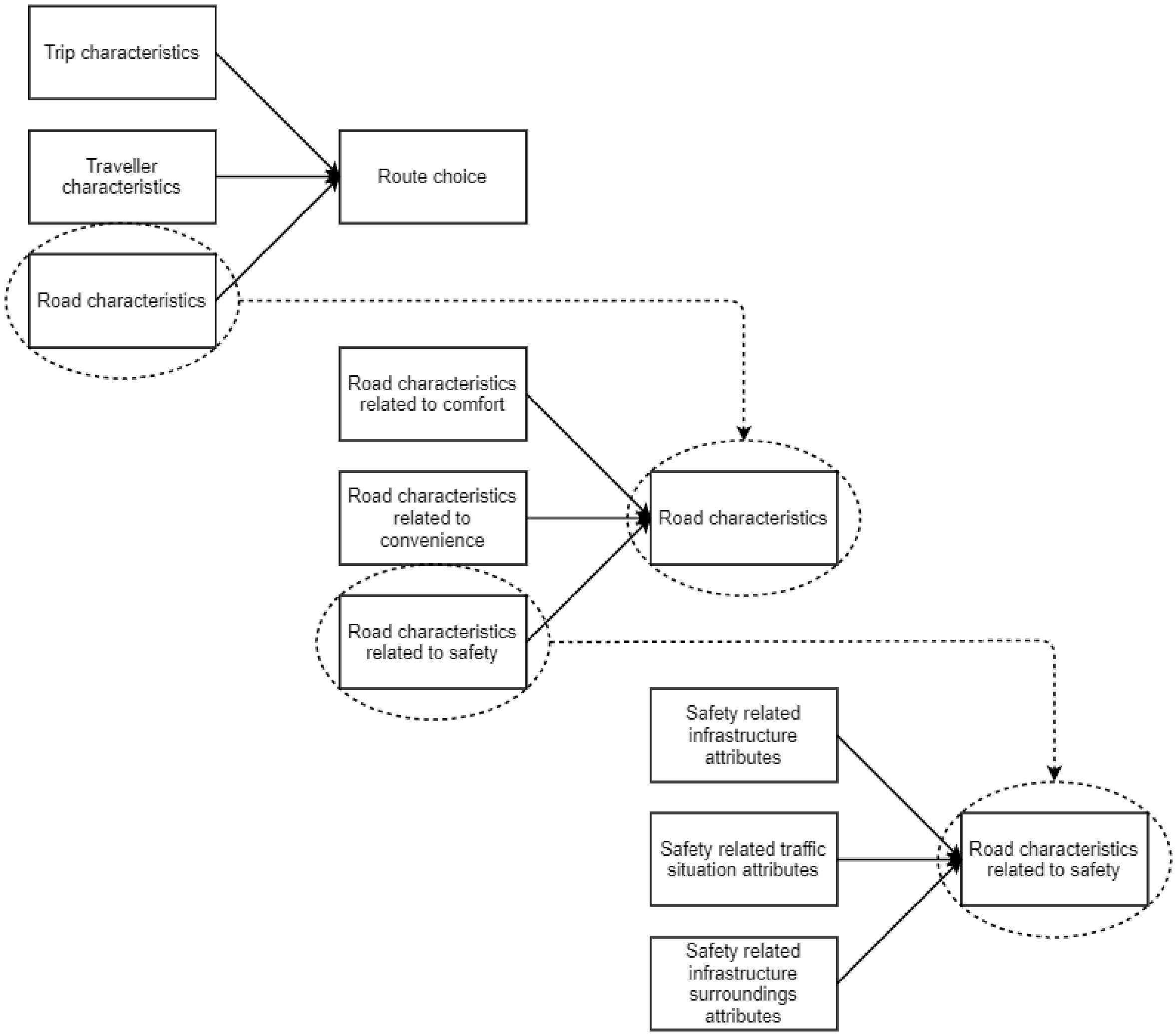

Figure 1: Conceptualisation of the research problem - High level of abstraction

Another important insight that contributes to understanding the research problem on a high level of abstraction is that travel time is often found to be the most important determinant in the route choice behaviour of cyclists. This finding suggest that other attributes that play a role in the route choice behaviour of cyclists, such as road characteristics related to safety, are most strongly dominated by travel time. This makes travel time the most relevant attribute to compare the safety-related infrastructural attributes with.

3.2.2. Low level of abstraction

When it comes to understanding the research problem on a lower level of abstraction, the most important insight gained from the literature review is that there are at least seventeen infrastructure related attributes that might affect cycling safety (See conclusions literature review, paragraph 2.7.). This insight determines the conceptualisation of the research problem on a lower level of abstraction and forms the basis for the continuation of this study. A conceptualisation of the research problem on a low level of abstraction is presented in figure 2.

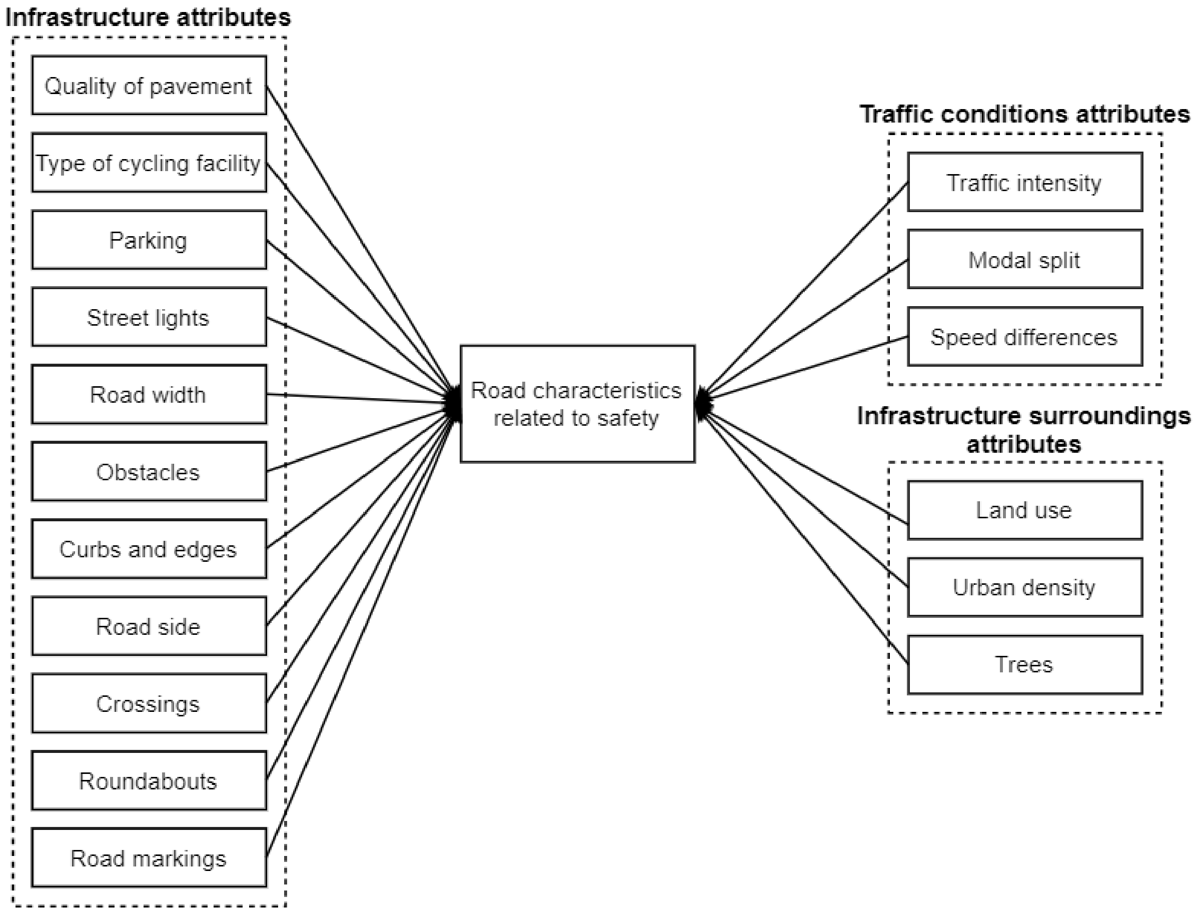

Figure 2: Conceptualisation of the research problem - Low level of abstraction

3.3. Specifying and operationalising the research problem

In order to solve the research problem, it is first specified and operationalised. This means that the information that is needed to solve the research problem is precisely defined after which a technique is chosen to obtain this information.

3.3.1. Specifying the research problem

At this point in time, we know that the route choice decisions of cyclists are affected by many different attributes. The literature review revealed that travel time is one of the most important attributes, if not the most important attribute in influencing these decisions. However, the literature review also showed that safety seems to have a significant influence on the route choice decisions of cyclists. In fact, seventeen infrastructural attributes have been identified as being relevant factors that might affect the safety and route choice behaviour of cyclists.

Despite the extensive literature review, it is not always clear whether the attributes that are identified positively or negatively affect the route choice behaviour of cyclists (i.e. the choice for a particular route). In addition, current literature does not shed much light on the extent to which these attributes affect the route choice behaviour of cyclists. In order to answer the research question and ultimately reach the research objective, there is a need to clarify the direction and size of the effects that these attributes have on the route choice behaviour of cyclists.

3.3.2. Operationalising the research problem

The operationalisation of the research problem is based on various assumptions that follow from findings in the literature review.

As already mentioned, travel time is generally considered to be the most important attribute in route choice decision making. Therefore it is assumed that travel time is the most important attribute that affects the route choice behaviour of cyclists. In addition, it is assumed that people always want to minimize their travel time and that when they do not, some attributes of the route they choose outweigh travel time (this assumption is in line with the utility theory). Lastly, as a first attempt to operationalise travel time, a linear relationship between travel time and distance is assumed.

Based on the theoretical framework that is created by the three assumptions mentioned above, it is decided to answer the research question by comparing the safety-related infrastructural attributes of actually driven routes with those of corresponding shortest routes. This operationalisation is visualised in figure 3, using an example route between origin A and destination B.

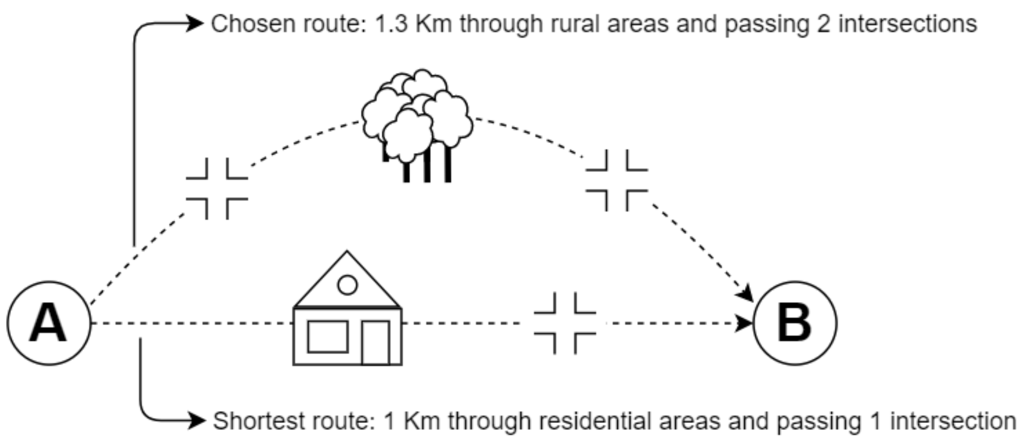

Figure 3: Operationalisation of the research problem

As part of this operationalisation, the strategy chosen to answer the research question is to conduct a regression analysis and use the difference in distance between the chosen route and the shortest route as the dependent variable and all safety-related infrastructural attributes as the independent variables (predictors). This makes it possible to determine the size and direction of the effects that the individual safety-related infrastructural attributes have on difference in distance and thus allows us to get insight into the role that these attributes play in the route choice behaviour of cyclists.

The operationalisation of the research problem is translated to practice as follows:

1. First, route sets will be created. Each route set contains a route that someone actually cycled in order to get from A to B as well as the shortest route between that same A and B.
2. Second, for all routes the values of all relevant attributes will be calculated. These are difference in distance and all safety-related infrastructural attributes that are identified in the literature review.
3. Third, all attribute values of the shortest route will be subtracted from the attribute values of the chosen route to identify the differences between the two considered routes. This will be done for each individual route set and the results of these subtractions together will form the final dataset.
4. Fourth and last, regression analysis is conducted on the final dataset with difference in distance as the dependent variable and all safety-related infrastructural attributes as the independent variables.

3.4. Statistical modelling

In order to analyse the dataset correctly, it is important to use the right statistical model.

3.4.1. Statistical modelling in general

A Statistical model is a non-deterministic mathematical model that consists of multiple statistical assumptions regarding the generation of data (Cox, 2006). The aim of a statistical model is to approach reality as closely as possible in order to, for example, make predictions. However, statistical models almost never completely reflect reality. “Any model is an approximation to reality. A theory is an abstract set of ideas that links together concepts. A model is a formal representation of a theory” (Bollen, 1989, Pg. 71).

3.4.2. Statistical models that are based on a continuous dependent variable

There are many different statistical models used in studies on route choice behaviour. Examples of such models are linear regression models and multinomial logistic regression models. Which model should be used depends on the type of data that is available and the research question that needs to be answered (Field, 2009).

The models that are commonly used in route choice behaviour studies can be roughly divided into two groups. The first group contains all models that are connected to a discrete nature of the dependent variable. Something that is also known as discrete choice modelling. The second group contains all models that are connected to a continuous nature of the dependent variable. As already explained in paragraph 3.3. there is one continuous dependent variable that will be investigated in this study which makes the regression based models the most suitable for the analyses. Therefore, this paragraph only zooms in on relevant models that fall into this category.

Two models that are, based on their characteristics, highly interesting for this study are the standard linear regression and the Tobit regression.

(Multivariable) Linear regression

A linear regression is a relatively basic type of predictive analysis that aims to either examine whether a set of independent (predictor) variables can predict a dependent variable well or to find out which independent variables are significant predictors of the dependent variable and in what way they impact the dependent variable (Field, 2009).

<u>Functioning of the model</u>

The most basic form of linear regression is the univariate linear regression. As the name already suggests, this type of linear regression studies the effect of only one independent variable (predictor) on the dependent variable. The univariate linear regression describes the dependent variable with a straight line and can be defined as shown in equation 1. In this equation, y represents the score of the dependent variable, a represents a constant (intercept), b represents the regression coefficient (slope), x represents the score of the independent variable and ε represents an error term that describes the difference between the observed value and the predicted value (Young, 2018).

$$y = a + b * x + \varepsilon \qquad \text{Eq. 1}$$

When there is one dependent variable and multiple independent variables, a multivariable linear regression is required. This variant of the linear regression describes the dependent variable as a linear function of multiple independent variables and can be defined as shown in equation 2. In this equation, y represents the dependent variable, b_n represent the partial regression coefficients, x_n represent the independent variables and ε represents the error term (Young, 2018).

$$y = a + b_1 * x_1 + b_2 * x_2 + \cdots b_n * x_n + \varepsilon \qquad \text{Eq. 2}$$

When it comes to the independent variables that should be included in the multivariable linear regression, there are roughly two options. One option is to include all potentially relevant independent variables. This option might seem the best way of doing it, because it allows for a better fit, but it has two major drawbacks. The first drawback is the likely lack of observations. A rule of thumb is that the number of observations should be at least 20 times greater than the number of included independent variables. The second drawback is the risk of over adjustment. When many irrelevant independent variables are included, some will have an effect purely by chance. This means that the better model fit that is the result of adding more independent variables is mainly caused by random effects that negatively impact the applicability of the model outside the used dataset. The other, better, option is to include only the independent variables that explain a large portion of the variance. Choosing this option makes the regression model more robust and explain the dependent variable a lot better (Schneider et al., 2010).

Something that one should always be aware of, regardless of which of the above mentioned options is chosen, is multicollinearity. Something that occurs when multiple independent variables are correlated. When the degree of correlation is high enough, it can cause problems with the model fit and interpretation of the results (Allen, 1997). Although there is no threshold value for when a correlation is too high, various studies and websites seem to indicate that one should be aware of multicollinearity when correlations are higher than 0.3 (on a scale from 0 to 1) and that correlations are likely to be problematic when they get above 0.5.

Interpretation of the model output

The usefulness of the linear regression totally depends on the correct interpretation of the regression output. The relevant output of the model should be interpreted as follows:

- The **R-squared (coefficient of determination)** reflects the explained variance and thus how well the model performs (goodness of fit). Since it is easy to make the R-squared artificially high by adding more independent variables. It is better practice to use the **adjusted R-squared (corrected coefficient of determination)**, which reflects the explained variance after correcting for the number of explanatory variables (Schneider et al., 2010).
- The **F-Test (F-Ratio)** indicates whether a linear regression model with various independent variables fits the data better than an intercept only model. The regression model fits the data better than the intercept only model when the P-Value for the F-Test is lower than the significance level (Field, 2009). The **P-Value (short for probability value)** is used to find out whether a relationship also exits in the larger population. This is done by comparing it to some level of acceptance (significance level). More specifically, the P-Value represents the probability that the null hypothesis is rejected when it is actually true. When P-values are lower, the evidence against the null hypothesis is greater. Usually the significance level is 0.05. This means that there is a 5% chance that no relationship exists when there is actually a relationship. However, stricter significance levels of 0.01 and less strict significance levels of 0.1 are also not uncommon (Hensher et al., 2015).
- The **Unstandardized B** represents the regression coefficient. For the dependent variable, this coefficient represents the constant. For the independent variables, this coefficient represents the change in the dependent variable per unit of change in the independent variable. Note that it is important to consider the units of measurement from proper interpretation (Schneider et al., 2010). The **Standardized B** represents the same thing as the unstandardized B. However, the standardized B values are all measured in standard deviation units which makes them directly comparable (Field, 2009).

Drawbacks of the model

As mentioned before, the type of data that is used and the question that needs to be answered determine which model is suitable. Given the way in which the research problem is operationalised, the linear regression at first seems a sensible approach. However, when you look a bit further, it potentially has one major drawback. This drawback stems from the fact that the linear regression is based on the Ordinary Least Squared (OLS) estimation. A technique that optimises the regression estimates by minimising the sum of squared errors

(Young, 2018). As with most statistical analyses, OLS has underlying assumptions. When these assumptions are satisfied, the model produces the best estimates possible. However, when one or more of these assumptions are not satisfied, the best estimates are not guaranteed anymore (Young, 2018). Unfortunately, as a result of the operationalisation of the research problem, not all of these assumptions are met.

The most important violation of the OLS assumptions is the fact that the dependent variable used in this study is always larger than 0 (When data has this property it is also referred to as censored data). The reason for this is that the chosen route simply cannot be shorter than the shortest route. As a consequence of this lower bound, the linearity assumption is violated which might result in a structural bias of the predicted values (Young, 2018).

Another violation of the OLS assumptions is caused by the distribution of the residuals. The modelling errors are not normally distributed, but limited from below as well. Something that will likely cause that the modelling errors do not all have the same variance. This means that the data is likely to be heteroskedastic, while one of the assumptions of OLS is that the data is homoscedastic (Young, 2018).

Because of the fact that the dependent variable has a lower bound, it might seem more logical to use a non-linear regression instead of a linear regression. However, the fact that the data is heteroskedastic is likely to cause that a non-linear regression is not suitable as well (Lim et al., 2012). In addition, using a non-linear regression would mean that the interpretation of the estimates would be much less intuitive.

The best modelling option for this study might be found in regression models that are able to work with censored data more appropriately. A good example of such a model is a Tobit regression.

Tobit regression

A Tobit model is a type of regression model that is able to take a dependent variable with a constrained range into account. The model is first proposed by Tobin in 1958, who developed the model in an effort to estimate relationships for limited dependent variables (Tobin, 1958). Just as with the linear regression, the aim of the model is to examine whether a set of independent variables can predict a dependent variable well or to find out which independent variables are significant predictors of the dependent variable and in what way they impact the dependent variable. However, contrary to a regular linear regression model, a Tobit model is specifically designed to work with censored data.

The Tobit model knows many variations that can be made by changing where and when the censoring occurs. According to Amemiya (1984), variations of the Tobit model can be divided into 5 categories based on the classification of the dependent variable. For this study, only the most basic version of the Tobit model will be considered.

Functioning of the model

The core concept of the basic Tobit model (and all Tobit models) is that it makes uses of an unobserved latent variable in order to simulate a continuing linearity between the dependent variable and independent variables after the point of censoring. By using the data points generated by the latent variable instead of the data points that are located on the line of censoring, the model is able produce better estimates (Young, 2018).

According to Young (2018), the basic Tobit model can be written as follows:

$$y_t^* = b_t' \, x_0 + \varepsilon_t, \; t = 1, 2, \ldots, n. \qquad \text{Eq. 3}$$

$$y_t = \begin{cases} y_t^* \; if \; y_t^* > c, \\ c \; if \; y_t^* \leq c. \end{cases} \qquad \text{Eq. 4}$$

In equation 3, which is very similar to equation 2, y_t^* represents the dependent variable, b_t' represents the regression coefficient (slope), x_0 represents the score of the independent variable and ε represents the error term. Equation 4 shows that if y_t is larger than *c*, y_t takes on that value. If y_t is smaller than *c* or equal to *c*, y_t takes on the value of *c*.

Interpretation of the model output

Just as with the linear regression, the usefulness of the Tobit regression totally depends on the correct interpretation of the regression output. The relevant output of the model should be interpreted as follows:

- The **ANOVA based fit measure** and the **DECOMP based fit measure** are both transformations of the **Log Likelihood function** and reflect the goodness of fit of the model. Due to the fact that a Tobit regression is not based on the OLS principle, calculating the R^2 (as with the linear regression) is not possible. For this reason, the ANOVA based fit and the DECOMP based fit measure are designed to mimic the R^2 roughly and can be used instead. The main difference between the two measures is that the ANOVA based fit measure only works with the variance of the predicted conditional mean and the variance of the dependent variable, whereas the DECOMP based measure also takes the residual variation into account (Greene, 1986a).
- The **coefficient** is the equivalent of the Unstandardized Coefficient B that we know from the regular linear regression. Similarly, the **Prob. Z** is the equivalent of the P-Value (Greene, 1986b).

3.4.4. Statistical model chosen for this study

For this study, it is chosen to work with both a linear regression as well as a Tobit regression. The reason for this is that although a Tobit model theoretically fits the dataset better, it is not certain whether it will actually produce better estimates than a regular linear regression. As explained by Young (2018), not meeting the OLS assumptions does not necessarily negatively impact the estimates of the linear regression which suggests that it possibly estimates the dependent variable just as good as a Tobit regression. This is a good opportunity to see whether a Tobit regression produces better results for our dataset than a regular multivariable linear regression would.

One consequence of choosing to estimate both a linear regression and a Tobit model is that their goodness of fit has to be compared at some point. However, due to the fact that the linear regression and the Tobit regression use different measures to reflect the goodness of fit, it is difficult to compare these models on this point by using the standard model output only. This difficulty can be overcome by using alternative methods to determine the model fit that can be applied to both models in the same way. One of these alternative methods is to calculate the **Root Mean Square Error (RMSE).** This is the sample standard deviation of the errors and can be interpreted as "the average absolute size of deviations of individuals from the sample regression line" (Young, 2018, Pg. 12). Another alternative method to compare the goodness of fit of a linear regression and a Tobit regression is by calculating the **Mean Prediction Error**. This measure indicates how much the predicted values on average deviate from the observed values and is expressed as a percentage of the average of observed values.

3.5. Data collection

The data is collected based on the methodological decisions. As a result, this study requires data on three topics: the road network, cycle routes and infrastructural attributes. This paragraph presents the most relevant information about the data collection in three paragraphs that each elaborate on one of the three aforementioned data topics.

3.5.1. Road network

The road network that is used for this study is the national road database (In Dutch: Nationaal wegenbestand) developed by Rijkswaterstaat (Rijkswaterstaat, 2020). This network is the most detailed, accurate free available representation of the Dutch road network. The network includes almost all roads that exist and displays them accurately with lines that connect seamlessly. The roads consist of many individual line segments (links) that all have their unique ID number and some other relevant information such as their length and the name of the street they represent. Series of contiguous line segments can be used to represent cycling routes.

3.5.2. Cycle routes

The cycle route data that is used for this study comes from the B-Riders project (B-Riders, 2020). A project that was part of the 'Beter benutten' program set up by the Dutch Ministry of Infrastructure & Environment and the Province of Noord-Brabant for the purpose of improving the accessibility of cities in Brabant. The project started in 2013 and ended in 2018. The B-Riders projects was specifically set up for companies and aimed to stimulate bicycle use among employees by rewarding them for cycling kilometres. The more an employee cycled, the greater the reward. A second purpose of the project was to collect data for analysis and supporting policy decisions. Therefore, the project did not just collect route data, but also sociodemographic data such as age and gender.

Employees who participated in the B-Riders project had to download a smartphone app that was able to track their routes by using GPS. In addition, the employees had to meet some requirements. They had to be at least 18 years of age, their commuting distance had to be at least 4 km and they must have used the car for at least half of their commuting trips over the last three months.

The data from the B-riders project contains detailed information about where and when the participants travelled. However, it is not directly useable for this study, because the data is nothing more than a bunch of GPS traces. Fortunately, the Urban Planning group of the Eindhoven University of Technology developed software that is able to transform the GPS traces into activity-travel diaries (Feng & Timmermans, 2018). Their so-called 'Trace Annotator' software is able to recognise the transportation mode with which a trip is made by combing GPS data with accelerometer data. The software predicts that a bicycle is used as means of transport with 97% accuracy.

The data file that is used for this study is generated by the Trace Annotator software based on B-Riders data of the months January till May and contains the travel data of cyclists only. The data file is an Excel file that contains detailed information about where and when the participants travelled. The Excel sheets contains the GPS coordinates of the starting point, intermediate points and end point of all their routes and indicates for each GPS measurement, down to the second, when it was taken.

3.5.3. Independent attributes

The data that is collected about the independent attributes comes from multiple sources. In search for the right data, special attention has been paid to the use of the most reliable sources only. The result of which is that most infrastructural data comes from governmental institution such as the municipality of Eindhoven, Rijkswaterstaat and the Dutch Central Bureau of Statistics. However, not all infrastructural data could be provided by these highly reliable governmental institutions. Therefore, a limited amount of infrastructural data comes from other sources, such as the Bicycle Week (Fietstelweek) and OpenStreetMap. Despite the fact that these sources are no governmental institutions, their data is considered to be sufficiently reliable for the purpose of this study. Furthermore, the B-Riders project is used to provide data that relates to the moment the trips were made and by who they were made. A brief description about the sources for the attribute data as well as information about which data source is used for which attribute is given in table 3. As one might notice, the attributes about which data is collected do not exactly match the attributes that were identified in the literature review. This will be explained in the next paragraph.

Table 3: Data sources of the independent attributes

Source	Description	Attributes
Gemeente Eindhoven (Gemeente Eindhoven, 2020)	Municipality of Eindhoven. They have a data portal through which all kinds of information about the municipality of Eindhoven is available.	Entrances & Exits, Trees, Speed Bumps, Pavement quality, Facility type, Parking, Road sides, Speed differences
CBS (CBS, 2020).	The Dutch Central Bureau of Statistics is a governmental organisation that collects data about the Dutch society.	Land use, Urban density
Rijkswaterstaat (Rijkswaterstaat, 2020)	Rijkswaterstaat is the executive arm of the Ministry of Infrastructure and Environment and is responsible for most of the Dutch infrastructure.	Speed differences,
Fietstelweek (Fietstelweek, 2020)	Large Dutch cycling study that has been carried out by volunteers in 2016 and 2017 with the aim of collecting data on cycling behaviour.	Modal split, Average cycling speed
OpenStreetMap (OpenStreetMap, 2020)	OpenStreetMap is a project that aims to collect freely available and editable geographic data in order to create maps and other services.	Traffic intensity
B-Riders project (B-Riders, 2020)	A project set up by the ministry of infrastructure & environment and the Province of Noord-Brabant for the purpose of improving the accessibility of cities in Brabant.	Gender, Age, Daylight, Peak, Weekend, Month

3.6. Attribute operationalisation

The infrastructural attributes that are identified in the literature review are operationalised based on the data that is collected. This means that for all infrastructural attributes is decided how they will be measured based on the available data.

3.6.1. General approach

The literature review made clear what the ideal operationalisation for each of the identified attributes would be. However, based on the available data, this ideal operationalisation is not always possible. In some cases the data is not complete enough and in other cases the data is not available at all. Since generating the missing data is too time consuming, it is decided to work around the missing data. In doing so, the strategy is to keep as many attributes on board as possible. Since there are only two attributes that can be operationalised in the theoretically ideal way, this requires a lot of creativity and inventiveness.

Despite good efforts, it was not possible to keep all attributes on board. Sometimes it was simply not possible to work around the missing data. Of the 18 attributes that were identified in the literature review, 12 could be operationalised and are therefore included in the further course of this study. Two attributes (‘Speed bumps’ and ‘Average cycling speed’) that were not directly identified in the literature review but seem to have a relation with cycling safety are also operationalised and included as well as six attributes that relate to the moment of the trip and by who the trip is made. This brings the total of included attributes to 20.

3.6.2. Included attributes

The attributes that are included in this study can be roughly divided into two categories: Continuous attributes and categorical attributes.

The continuous attributes are used in situations where it does not make sense to distinguish levels. These attributes count or measure things on a continuous scale. For example, the number of trees along a route. Given the fact that all routes have a different length, the cumulative value that is expressed by this attribute is, if necessary, converted to a value per 100 or 1000 metres of route.

The categorical attributes are used when it does make sense to distinguish between different levels. However, it is decided to bring all multilevel attributes down to only two levels. In this process, the levels that are most similar to each other are merged into one level. An example of a categorical attribute is 'Parking'. For this attribute, the levels "parking" and "no parking" are created. In order to be able to compare routes based on these two-level attributes, the values of the levels is expressed as a percentage of the route (e.g. 15% of the route is "parking" and 85% is "no parking").

In total, this study includes 6 continuous and 14 categorical attributes. An overview of the how the attributes that are included in this study are operationalised is presented in Table 4.

Table 4: Attribute operationalisation overview

	Nr.	Attribute	Level	Measured as	Measured in
	1	Distance	Continuous	Total length of the route	Km
Continuous attributes	2	Entrances and exits	Continuous	All entrances and exits that are designed for vehicles	Total number for the whole route
	3	Modal split	Continuous	Average cyclist intensity	Average number per 1000 metres of route
	4	Trees	Continuous	All trees within 10 metres measured from the road centre line	Average number per 100 metres of route
	5	Speed bumps	Continuous	Only speed bumps. No raised crossings	Total number for the whole route
	6	Average cycling speed	Continuous	Average cycling speed for the total route	Kilometres per hour
Categorical attributes	7	Pavement quality	Low quality	Elements and unpaved	Percentage of total route
			High quality	Asphalt and concrete	
	8	Type of facility	Cyclists not segregated	Traffic lanes and bicycle lanes	Percentage of total route
			Cyclists segregated	Bicycle paths	
	9	Parking	Parking	Perpendicular or parallel parking	Percentage of total route
			No parking	No parking	
	10	Road side	Hard road side	Concrete, stone and asphalt	Percentage of total route
			Soft road side	Grass and other vegetation.	
	11	Traffic intensity	Quiet roads	Residential and other typically quiet roads	Percentage of total route
			Busy roads	Primary, secondary and tertiary roads	
	12	Speed differences	Slow traffic	Speed limit till 30 km/h	Percentage of total route
			Fast traffic	Speed limit above 30 km/h	
	13	Land use	Non-Built-up	Recreational areas, parks, forest, etc.	Percentage of total route
			Built-up	Residential, industrial, etc.	
	14	Urban density	Low urban	More than 1500 addresses per km^2	Percentage of total route
			High urban	Less than 1500 addresses per km^2	
	15	Gender	Man	Man	Absolute number
			Woman	Woman	
	16	Age	Younger	< 48 years of age	Absolute number
			Older	> 47 years of age	
	17	Daylight	Daylight	Trip starts after sunrise/before sunset	Absolute number
			No daylight	Trips starts before sunrise/after sunset	
	18	Peak	During peak	Trip starts between 7.00-9.00 AM or 4.30-6.30 PM	Absolute number
			Off-peak	Trip starts outside 7.00-9.00 AM and 4.30-6.30 PM	
	19	Weekend	Weekday	Monday till Friday	Absolute number
			Weekend	Saturday and Sunday	
	20	Month	Cold month	January, February, March.	Absolute number
			Warmer month	April, May	

3.6.3. Excluded attributes

Multiple attributes are excluded from this study for various reasons. The most important reason why attributes are excluded is that there is simply not enough data available to operationalise them in way that makes sense. This applies to the attributes 'Obstacles', 'Colour of the road surface', and 'Road markings'. There has certainly been an attempt to work around the lack of data but this did not pay off. As mentioned before, generating this data is not an option due to the short time frame of this study. Another reason why some attributes are excluded is that they are not relevant for the target area. This applies to the attributes 'Street lights' and 'Roundabouts'. The specific reason to exclude 'Street lights' is that 99% of all streets and cycle paths in the study area has street lights. For this reason there is no point in distinguishing between roads with and roads without street lights. The attribute 'Roundabouts' is excluded because there are almost no roundabouts in study area. Of all places where roads intersect less than 1% is a roundabout. This is simply not enough to be meaningful in this study.

3.7. Creating the final dataset

The literature review made clear what data is required for this study and the choices made in the methodology chapter determine what the dataset ideally looks like. Unfortunately, such a dataset is not directly available, but has to be compiled by combining a large variety of data files of various shapes and sizes. A comprehensive process of which the outcome is of great importance for the quality of this study. This paragraph explains how the final dataset that will be used for further analysis is created by successively discussing the five steps that make up this process.

The software that are used in the process of building the dataset are QGIS (https://www.qgis.org/) and TransCAD (https://www.caliper.com/). QGIS is an open source Geographical Information System and TransCAD is a Geographical Information System that is specifically designed for transportation professionals. Detailed information about this software can be found on the websites of the developers.

3.7.1. Data preparation

The first step in the process of creating the final dataset is data preparation. This is the process of preparing all individual data files in such a way that they can be used for building the overall dataset that is required for this study. As already explained in the paragraph about data collection, the dataset is built with data about the road network, data about cycle routes and data about infrastructural attributes. Since these data files are all set up differently and contain information on different topics, they all require a different preparation.

Road network

The road network that is used for this study is more or less ready to use and does not require a lot of adjustments. However, there are a few important things that need to be taken care of. First of all, the data file is reduced in such a way that it only contains data of the target area Eindhoven. What the result of this action looks like is shown in figure 4. Secondly, excess attribute data is removed.

Figure 4: The road network of the Eindhoven region (Rijkswaterstaat, 2020)

Route data

The B-Riders route data (adapted by the Trace Annotator tool) that is used for this study is an Excel file that consists of several thousands of routes that are all represented by series of GPS coordinates. These routes are not all usable for this study which means that a selection needs to be made. Selecting usable routes is a process that consists of two main actions: Selecting the routes that are for the most part located in Eindhoven and selecting routes that are unique and reliable.

The action that is conducted first is to select the routes that are for the most part located in Eindhoven. This selection process is done visually. First, TransCAD is used to convert the Excel GPS coordinates to strings of GPS points and project these GPS strings on the road network. Then, routes that are for a large part or even fully located outside the study area are removed from the dataset. Figure 5 shows all February routes in the Eindhoven area. It is clearly visible that some routes need to be removed from the dataset.

Figure 5: February routes in the Eindhoven area (Rijkswaterstaat, 2020; B-Riders, 2020)

The second step is to select routes that are unique and reliable. This is done based on the following selection criteria:

- Routes must follow infrastructure on which cycling is allowed;
- Routes cannot have a large detour;
- Routes driven by the same respondent need to differ considerably from each other.

What the result of this selection process looks like for one person is shown in figure 6.

Figure 6: All (2) routes of one person (Rijkswaterstaat, 2020; B-Riders, 2020)

Attribute data

Preparing the large amount of attribute data files and getting them ready for use involves many actions. However, the most important actions are that the coordinate systems of all attribute data files are lined up with the coordinate system of the road network and that data that does not relate to the target area or is obviously irrelevant for other reasons is removed.

3.7.2. Data aggregation

When the individual data files are roughly filtered of irrelevant data and are tuned to each other, they are brought together. This process is also referred to as data aggregation and forms the second step in the process of creating the final dataset. The data aggregation consists of roughly two actions that are carried out sequentially. First, attribute data is added to the road network. Then, cycle route data is added to the road network.

Adding attribute data to the road network

The attribute data is spread over seven different types of geographical data files. Some of these data files contain information in points (e.g. trees), others in lines (e.g. speed differences) or polygons (e.g. land use). Transferring the data that is in these data files to the road network requires the use of many different data processing techniques, such as creating buffers, calculate intersections and even run tailor made python scripts. All these actions are carried out in QGIS.

Despite the fact that adding attribute data to the road network is done with utmost care, it is not possible to give all road sections the most optimal attribute values. The differences between some of the data files that contain the attribute data and the road network are too large to get a 100% accurate data transfer by using GIS software only. However, it was possible to get over 90% accuracy. After transferring all attribute data to the road network, visual and manual checks are conducted to correct suboptimal attribute values and get closer to 100% accuracy. These checks are conducted by comparing the attribute values that are now connected to the road network with the attribute values of the original data files they came from. In cases where there was still uncertainty about the correctness of certain attribute values, Google Maps (Google Maps, 2020) is used to verify these values.

Adding cycle route data to the road network

As explained earlier, the cycle route data is basically nothing more than strings of GPS points with some information about the respondent and time information that makes it possible to determine when and in which direction the route is cycled. In order to add the route data to the road network, the strings of GPS points are translated into actual routes that follow the links of the road network. This is done by comparing the strings of GPS points with the road network in TransCAD. First, the locations of the GPS points in relation to the road network are used to determine which links are used for a particular trip and to draw the driven route. Second, the start- and endpoints of these routes are used to determine and draw the corresponding shortest routes. The result of these actions is to have pairs of driven and shortest routes that are selections of interconnected road sections in the road network. This is the basis that is required for determining the attribute values of a whole route and ultimately to compare driven routes to shortest routes. An example of how a set of GPS points is translated into a driven and shortest route is presented in figure 7.

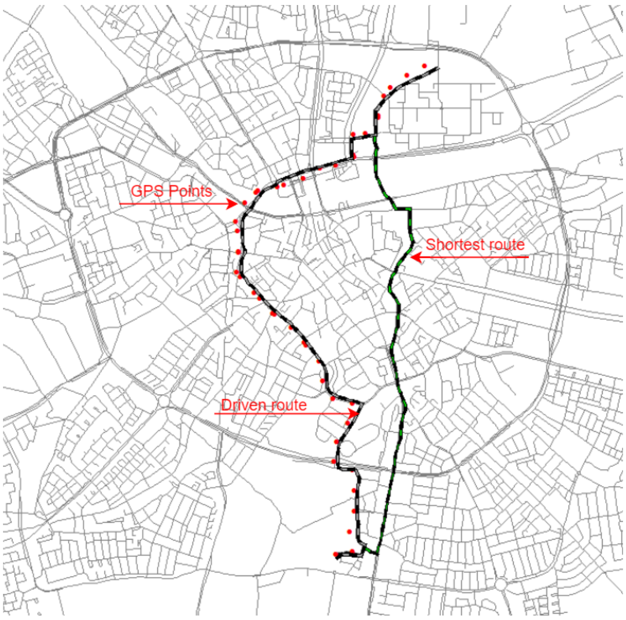

Figure 7: Translating GPS points into routes (Rijkswaterstaat, 2020; B-Riders, 2020)

When it comes to the accuracy of this process, one has to realise that GPS is not 100% accurate. There is always a slight deviation from reality which, in extreme cases, can reach several tens of metres. For this study, this does not cause any big problems. The distance between two roads is generally so large that there is no doubt about which road was chosen.

3.7.3. Data cleaning

When all data files are brought together, the big dataset is cleaned of the last irrelevant data. This is the third step in the process of creating the final dataset. The first action in the data cleaning process is the removal of irrelevant links. The original road network consists of almost 12000 links. However, only 3924 links are used to draw the routes. The roughly 8000 links that are not used are removed from the network to make it better workable. The second action is the removal of strange attribute values. Fortunately, there is only one attribute that has such values. This is the attribute that represents average cycling speed. According to the data, there are some links where people cycle extremely slow (<5 km/h) or extremely fast (>30 km/h). Because of the fact that these speeds are highly unusual for cyclists and are based on only one measurement, the chance that these values are incorrect is very high. Therefore, it is decided to remove all values that are below 5 km/h or above 30 km/h.

3.7.4. Data enrichment

When creating the dataset for this study, the goal was to get a complete (100% filled in) dataset purely by combining several smaller data files. This way, it will be easy to scale up or reproduce the study for a different area. However, this turned out to be impossible and now that all irrelevant data has been removed from the dataset, it is easy to see which data is still missing. Considering the fact that having this data would be of great added value for this study and that the amount of missing data is relatively small, it is decided to fill the empty fields. This process of enriching the dataset with missing data is the fourth step in the process of creating the final dataset and consists of two main actions that are carried out sequentially.

Adding links and nodes to the road network

The first action is adding links and nodes to the road network that are missing, because they are no official road or intersection for car traffic. Since cyclist do make use of these links and nodes, not having them in the road network means that some of the actual driven routes can only be replicated with a large detour. This is a problem, because this would make many of these routes too inaccurate to be used for this study. Therefore, approximately 60 links and 20 nodes are added to the road network. An example of a missing link is presented in figure 8 and an example of a missing node is presented in figure 9.

Figure 8: Example of a missing link (Google Maps, 2020)

Figure 9: Example of a missing node (Google Maps, 2020)

Filling empty fields
The second action is to fill the empty fields. There are multiple reasons why there are empty fields in the data file. However, the most important reason is that various data files are connected to the road system based on geographical similarities. Despite the fact that this is the best option, this does not work for all the links. Sometimes the geographical similarities between the data files are very vague, which causes that the GIS software cannot make the right connection. Filling in the empty fields is done by manually assessing all links that have missing values. Most of the times it is possible to assess a group of links at once, but some links need to be assessed individually.

Assessing the links with missing attribute values is initially done by comparing these links with the data files that are the source of the missing attribute values. When the missing attribute value is found, it is added to the road network. Unfortunately, not all missing values can be found by using this technique. When this is not possible, the missing attribute values are determined by using Google Maps and Google Street View (Google Maps, 2020).

There is one attribute of which the missing data cannot be added by using the techniques described above. This is the average cycling speed attribute. This data is for some links simply not available and this data cannot be derived from Google Maps or Google Street View. The strategy that is chosen to fill these empty fields is by calculating average cycling speed per pavement type. Research shows that cyclists cycle on average 17.51 Km/H on smooth surfaces and 17.26 Km/H on rough surfaces. For this reason, it is decided to give all links that have an empty field for average cycling speed and a rough surface an average cycling speed value of 17.26 KM/H and all links that have an empty field for average cycling speed and a smooth surface an average cycling speed value of 17.51 Km/H.

3.7.5. Dataset generation
When the empty fields are filled, the final dataset that will be used for further analysis is generated. This last step in the process of creating the final dataset consists of two main actions. The first action is to calculate the attribute values of the routes as a whole. This is done based on the length and other attributes values of the individual links out of which the various routes consist.

The second action is calculating the attribute differences between the driven routes and shortest routes. This is done by subtracting the attribute values of the shortest routes from their corresponding driven routes. The result of this last action and thereby the whole data set generation process is a dataset that contains information about the differences between driven routes and corresponding shortest routes. The dataset contains information about the difference in length between these routes, but also about the attribute differences between these routes.

4. Results

4.1. Introduction

The results of this study consist of roughly two parts that are both presented in an individual paragraph. Paragraph 4.2. contains the results of the descriptive analysis, the purpose of which is to describe several basic features of the dataset that is used in this study. This paragraph first presents the descriptive statistics of the respondents, then those of the trips (journey) and lastly those of the routes (the infrastructure used for the trip). Paragraph 4.3. contains the results of the statistical analysis. In order to present the results as clearly and comprehensible as possible, first regression models that include all attributes are discussed. Most of the time, these full models do not give the best result and need to be improved. This might be because their estimation process is negatively influenced by high correlations between independent variables. The full models often provide a rough idea of how the regression performs and provide insights into how a regression model can be improved. These full models can be considered as the starting point and foundation of the regression analyses. Second, regression models that only include the significant attributes are discussed. These are the adapted models that, after various model improvements, produce the most meaningful and accurate output possible and thus, more or less, mark the end point of the regression analyses. Last, the best performing adapted regression model(s) will be discussed in detail. The results of this model are determinative for answering the research question and reaching the research objective.

4.2. Descriptive analysis

As already mentioned in the previous chapter, the B-Riders data that was available for this study covered the months January till May and contained mainly travel data of cyclists. The B-Riders data contained a few thousand trips made by a few hundred respondents. However, these trips were not all usable. After carefully selecting only the unique and reliable trips, 145 unique trips that all have their start- and endpoint in and around the city of Eindhoven and are made by 48 unique respondents remained. These routes have been incorporated in the final dataset and are investigated in this study.

4.2.1. Descriptive statistics of the respondents

This part presents the descriptive statistics of the 48 unique respondents and has, besides providing insight into the sample data, another important purpose. This purpose is to determine the extent to which the sample is representative for the Dutch population. Something that is done by comparing the sample population with the Dutch population. Because of the fact that the final dataset contains data on ‘gender’ and ‘age’, these two attributes are used for the comparison. The results of these comparison are presented in table 4 and 5.

Table 5: Gender distribution sample population in relation to overall population

Gender	Sample population		Overall population in 2019 (CBS, 2020)	
Male	24	50.0%	8581086	49.7%
Female	24	50.0%	8701077	50.3%
Total	48	100%	17282163	100%

Table 6: Age distribution sample population in relation to overall population

Age	Sample population		Overall population in 2019 (CBS, 2020)	
< 35	3	6.3%	7070532	40.9%
35-44	11	22.9%	2056681	11.9%
45-54	20	41.7%	2512575	14.5%
> 55	14	29.2%	5642375	32.6%
Total	48	100%	17282163	100%

Table 5 shows that when it comes to gender, the sample population is representative for the Dutch population. However, table 6 shows that when it comes to age distribution, the sample population is not representative at all. The reason for this discrepancy is likely to be the nature of the B-Riders program in combination with the requirements one has meet to join the program. For example, take the simple fact that participants have to work in order to participate. This means that they are generally at least in their twenties. Knowing that people younger than 25 make up for almost 30% of the total population (CBS, 2020), this explains for a large part the underrepresentation of the <35 category. In addition, the overall population includes all types of people, while the B-riders project probably attracts the more sporty types.

Based on the descriptive statistics presented above, it can be concluded that the sample population shows similarities with the Dutch population based on the attributes age and gender. Despite the fact that some age categories in the sample population are underrepresented compared to the Dutch population, they are all present. Therefore, based on the sample population, the dataset is considered to be useful for further analysis.

4.2.2. Descriptive statistics of the trips

Because of the fact that the B-Riders data that is used for this study only included 48 unique respondents it is decided to make the dataset larger by incorporating multiple trips per respondent when possible. The result of which is that now respondents are responsible for one to six trips. This brings the total number of trips that are included in the dataset to 145.
This part presents the descriptive statistics of the trips in six tables. The first two tables (table 7 and 8) are similar to the ones showed in the previous part. However, this time it is not the sample population that is compared to the overall population, but the trip selection. As one can see, the results of this comparison are quite similar to the previous comparison. The only noteworthy difference is that this time females are slightly underrepresented compared to the Dutch population. The last four tables (table 9, 10, 11 and 12) show the distribution of the months in which the trips have taken place, the day on which the trips have taken place, the part of the day on which the trips have taken place and the daylight conditions under which the trips have taken place respectively.

Table 7: Gender distribution trip selection in relation to overall population

Gender	Trip selection		Overall population in 2019 (CBS, 2020)	
Male	74	51.0%	8581086	49.7%
Female	71	49.0%	8701077	50.3%
Total	145	100%	17282163	100%

Table 8: Age distribution trip selection in relation to overall population

Age	Trip selection		Overall population in 2019 (CBS, 2020)	
< 35	6	4.1%	7070532	40.9%
35-45	37	25.5%	2056681	11.9%
45-55	60	41.4%	2512575	14.5%
> 55	42	29.0%	5642375	32.6%
Total	145	100%	17282163	100%

Table 9: Trip frequencies – Month of the year

Month of the year	Absolute	Percentage
January	23	15.9%
February	15	10.3%
March	28	19.3%
April	14	9.7%
May	65	44.8%
Total	145	100%

Table 10: Trip frequencies - Part of the week

Part of the week	Absolute	Percentage
Weekday	129	89.0%
Weekend	16	11.0%
Total	145	100%

Table 11: Trip frequencies – Part of the day

Part of the day	Absolute	Percentage
Peak*	67	46.2%
Off peak	78	53.8%
Total	145	100%

*Peak is considered to be between 7.00-9.00 AM and 4.30-6.30 PM

Table 12: Trip frequencies – Amount of daylight

Amount of daylight	Absolute	Percentage
Daylight*	130	89.7%
No daylight	15	10.3%
Total	145	100%

*Daylight trips are all trips that start after sunrise and before sunset

Based on the descriptive statistics presented above, it can be concluded that also the trip selection shows similarities with the Dutch population based on the attributes age and gender. Again, some categories are underrepresented, but they are all present. In addition, these descriptive statistics show that the trip selection is diverse on multiple aspects. It includes trips made during cold and warm months, on weekends and on week days, during peak hours and off-peak hours and with and without daylight. Although chances are high that the numbers presented in the tables above will not match those of the Dutch population, many different kinds of trips are present. Therefore, based on the trip selection, the dataset is considered to be useful for further analysis. Though, the discrepancies between the trip selection and the Dutch population are something to consider when drawing the conclusions.

4.2.3. Descriptive statistics of the routes

The fact that the dataset that is used for this study includes 145 trips means that it includes a total of 290 routes. These are the 145 routes that the 48 respondents actually cycled from an origin to a destination and the 145 corresponding shortest routes between those same origins and destinations. This part presents the descriptive statistics of all 290 routes in two tables. The first table (table 13) presents the minimum and maxim attribute values for all routes. The second table (table 14) presents the mean attribute value and corresponding standard deviation for all routes.

Table 13: Minimum and maximum attribute values for all routes

	Driven routes		Shortest routes	
Attribute	**Minimum**	**Maximum**	**Minimum**	**Maximum**
Distance (Km)	1.16	7.23	1.12	6.98
BuiltUp (%)	32.84	100.00	19.08	100.00
HighUrban (%)	0.00	100.00	0.00	100.00
SlowTraffic (%)	65.79	100.00	51.06	100.00
QuietRoads (%)	19.04	100.00	32.55	100.00
SoftRoadSide (%)	0.00	71.13	0.00	87.74
NoParking (%)	12.29	100.00	14.78	100.00
Segregated (%)	11.31	100.00	14.62	100.00
HQPavement (%)	29.49	100.00	8.91	100.00
Exits(Abs)	0.00	219.00	0.00	179.00
Trees (Per100M)	2.19	17.71	1.35	21.15
SpeedBumps (Abs)	0.00	48.00	0.00	45.00
CyclistIntensity (Abs)	6.76	109.31	1.24	119.73
Av.Cyc.Speed (Abs)	15.21	20.86	15.03	22.59

Table 14: Mean attribute value and corresponding standard deviation for all routes

	Driven routes		Shortest routes	
Attribute	**Mean**	**Std. Dev**	**Mean**	**Std. Dev**
Distance (Km)	3.87	1.51	3.57	1.38
BuiltUp (%)	82.45	16.25	82.98	16.50
HighUrban (%)	78.82	23.25	79.79	22.45
SlowTraffic (%)	93.44	8.76	92.93	10.32
QuietRoads (%)	77.35	16.86	77.19	18.17
SoftRoadSide (%)	13.99	15.71	10.76	14.58
NoParking (%)	65.20	16.21	67.48	16.11
Segregated (%)	68.85	19.01	65.38	20.98
HQPavement (%)	75.32	15.55	71.92	17.53
Exits(Abs)	67.77	46.66	59.54	40.69
Trees (Per100M)	9.82	2.98	9.55	3.16
SpeedBumps (Abs)	9.38	9.02	9.83	8.21
CyclistIntensity (Abs)	41.20	21.18	39.06	22.73
Av.Cyc.Speed (Abs)	17.93	1.05	17.80	1.17

The descriptive statistics presented above show that all attributes are present in both the group with driven routes and the group with shortest routes and that there are considerable differences between the chosen routes and the shortest routes. Because of the fact that there are considerable differences between the chosen routes and the shortest routes, it can be concluded that comparing these two groups has the potential to yield interesting results. Just as with the sample population and the trip selection, the dataset is also considered to be useful for further analysis based on the route characteristics.

4.3. Statistical analysis

In order to generate appropriate insights for this study, a statistical analysis is carried out. As explained in the previous chapter, the aim of this analysis is to determine what the effect of various safety-related infrastructural attributes on the difference in distance between the chosen and shortest routes is.

4.3.1. Full regression models

Mostly regression analysis starts with a model that includes all available independent variables. Table 15 shows the estimation results of the full linear and the Tobit regression model. The detailed output of these regression models is presented in appendix C.

Table 15: Estimation results of the full linear and Tobit regression

		Linear regression		Tobit regression	
Model		Coefficient	Sig.	Coefficient	Sig.
1	(Constant)	.274**	.011	.267***	.009
	P_BuiltUp	-.001	.771	-.000	.759
	P_HighUrban	-.004	.216	-.003	.219
	P_SlowTraffic	.004	.367	.004	.319
	P_QuietRoads	.006*	.076	.005*	.073
	P_SoftRoadSide	.003*	.092	.003*	.065
	P_NoParking	-.003	.230	-.002	.186
	P_Segregated	-.001	.765	-.000	.819
	P_HQPavement	.000	.877	.000	.866
	Exits_Abs	.002**	.043	.001**	.025
	Trees_Per100M	.024**	.017	.024**	.010
	SpeedBumps_Abs	.002	.594	.002	.627
	Cyclist_Intensity	-.005***	.002	-.005***	.001
	AverageCyclistSpeed	-.007	.766	-.005	.798
	Gender	-.059	.244	-.064	.192
	Age	-.084	.106	-.093*	.063
	Month	-0.005	.920	-.015	.738
	Daylight	.006	.942	.013	.861
	Peak	.043	.398	.055	.274
	Weekend	.079	.322	.081	.302
	Sigma			.268***	.000
	Log Likelihood	-		-23.01	
	Mean prediction error	20.28%		20.42%	
	RMSE	0,2529701		0,252256	

Note: ***, **, * → Significance at, 99%, 95%, 90% interval

As can be seen from the table above, the two models that include all independent variables produce a relatively similar output. Both the coefficient values (size and direction) as well as the significance values are more or less the same. The goodness of fit that is represented by the mean prediction error and the root mean square error is also highly comparable.

Despite the fact that the output of the linear regression is very similar to the output of the Tobit regression, there is one important difference. The significance values of the Tobit regression are slightly better than those of the linear regression. As a result, in the Tobit regression, “age” is also significant and “distance” is significant within the 99% interval instead of the 95% interval.

The difference described above suggest that, when applied to our dataset, the Tobit model slightly outperforms the linear regression model. A result that is in line with the theory presented in paragraph 3.4. and therefore is expected. However, these are the full regression models. It is too early to conclude that the Tobit model not only theoretically fits the dataset better (due to the censored nature of the data), but also practically. The main reason for this are the high (above 0.5) correlations between some independent variables in the full models. These correlations might negatively impact the estimation process of the full models, which negatively impacts their suitability for answering our research question. Detailed information about attribute correlations of the full models can be found in appendix C.

In order to ensure that the regression models have the best model fit possible, adaption of the full regression models is required. The results of these adapted models are presented in the next part of this paragraph.

4.3.2. Adapted regression models

The endpoint of the regression analysis is an adapted model that is stripped of independent variables that are either insignificant or cause high correlations with other independent variables. Table 16 shows the estimation results of the adapted linear and Tobit regression model. The output of these regression models is presented in appendix C.

Table 16: Estimation results of the adapted linear and Tobit regression

		Linear regression		Tobit regression	
Model		Coefficient	Sig.	Coefficient	Sig.
1	(Constant)	.344***	.000	.340***	.000
	P_QuietRoads	.004***	.008	.004***	.007
	P_SoftRoadSide	.003*	.061	.003*	.056
	Exits_Abs	.002***	.002	.002***	.001
	Trees_Per100M	.024**	.013	.023**	.013
	Cyclist_Intensity	-.006***	.000	-.005***	.000
	Age	-.090*	.065	-.100**	.043
	Sigma			.276***	.000
	Log Likelihood	-		-27.54	
	Mean prediction error	21.17%		21.04%	
	RMSE	0,263187		0,260388	

Note: ***, **, * → Significance at, 99%, 95%, 90% interval

Just like the full regression models, the adapted regression models produce a relatively similar output. Again, both the coefficient values (size and direction) as well as the significance values are more or less the same. The only notable difference is that the significance values of the Tobit model are slightly better than those of the linear model which results in 'age' being significant within the 95% interval instead of the 90% interval. Also the goodness of fit of the models is highly comparable. However, also on this point the Tobit model seems to outperform the linear model slightly as the mean prediction error and the root mean square error suggest that the adapted Tobit model fits the data slightly better.

When it comes to comparing the model fit of the adapted models with that of the full models, the goodness of fit measurements seem to indicate that the adapted models perform slightly worse than the full models. However, based on the Likelihood Ratio Test presented in equation 5, this decrease is insignificant (Hensher et al., 2015).

$$LR = 2(LogL\ Unrestricted\ model - LogL\ Restricted\ model) \qquad \text{Equation 5.}$$

The Likelihood Ratio test uses the difference in Log Likelihood between the full model and the adapted model and relates this to the degree of restriction (decrease in predictors). Filling in this formula for the Tobit models results in a value of 9.06 . According to the Chi-squared table, the critical value for a decrease of 13 attributes (95 percent) is 22.36. Since 9.06 is way smaller than 22.36 it can be concluded that the model fit of the adapted model does not differ significantly from the model fit of the full model. Unfortunately, the linear regression models do not have a Log Likelihood value. However, since linear regression models are so similar to the Tobit models, it is assumed that the decrease in goodness of fit for the linear models is insignificant as well.

Overall, the regression output suggests that the adapted models perform slightly better than the full models and that the adapted Tobit model performs slightly better than the adapted linear model. However, the differences are just too small to be meaningful. Since the Tobit regression seems to perform slightly better and theoretically fits the dataset better, the adapted Tobit regression is used to interpret the coefficients.

4.3.3. Interpretation of the coefficients (adapted Tobit model)

The adapted Tobit model consists of 7 significant attributes including the constant 'Distance'.

Distance (Constant)

The 'constant distance' has a coefficient value of 0.340 which means that cyclists deviate on average 340 meters from the shortest route. This finding is pretty much in line with earlier research by Winters et al. (2010) and Dill & Gliebe (2008), who found that cyclists deviate on average 360 and 384 meters from the shortest route respectively. Therefore, an outcome of this magnitude was more or less expected.

A possible explanation for the relatively small difference between these findings could be the average trip lengths used in the different studies. The average trip length of the chosen routes that are used in this study is 3.87 Km, whereas Winters et al. (2010) and Dill & Gliebe (2008) used trips with an average distance of 5.33 Km and 6.92 Km on average respectively. It is logical that at longer distances the average deviation from the shortest route is larger than at shorter distances just because it is more difficult to determine the shortest route. In addition, a longer route generally means that there are more alternatives that may be more attractive than the shortest route.

Traffic intensity (P_QuietRoads)
The 'percentage of quiet roads' represents traffic intensity and this predictor has a coefficient value of 0.004 which means that for every additional percent of quiet roads cyclists deviate 4 meters more than average from the shortest route. The direction of this effect is in line with earlier research by Stinson and Bhat (2003), Sener et al. (2009) and Winters et al. (2010), who all found that cyclists prefer roads with lower traffic intensities. The positive sign for this attribute was therefore expected. However, the magnitude of the effect found in this study seems not to be in line with recent literature as it appears to be rather small. All the studies mentioned above found that traffic intensity has a major impact on route choice decisions and Stinson and Bhat (2003) even found that cyclists are willing to endure approximately 10% extra travel time in order to cycle on quieter roads. A possible explanation for the fact that the effect magnitude is smaller than expected is that the study area mainly consists of quiets roads. On average, the shortest routes consist for about 77% of quiet roads. Increasing this percentage even more might not outweigh the distance increase that comes with it. However, it must be noted that the different nature of the studies and differences in the way in which the attributes are operationalised make that comparing magnitude levels is very difficult.

Road sides (P_SoftRoadSides)
The 'percentage of soft road sides' predictor has a coefficient value of 0.003 which means that for every additional percent of soft road sides cyclists deviate 3 meters more than average from the shortest route. The effect of road sides on deviation from the shortest route or on route choice decision making in general has not been studied before, making it impossible to compare this finding with those of earlier investigations and making it hard to have founded expectations about the effect of this attribute. However, this does not mean that there were no expectations at all.

The literature review showed that road sides affect cycling safety and that safety affects the route choice behaviour of cyclists. Therefore, it is expected that the relatively safe soft road sides have a positive effect on the extent to which cyclists deviate from the shortest route. The positive sign for this attribute shown in the model output indicates that this expectation is in line with the model output.

When it comes to the size of the effect there were no expectations. However, based on the fact that the effect magnitude of P_SoftRoadSides is slightly smaller than that of P_QuietRoads, the size of the effect seems to make sense. The reason for this is that the number of quiet roads in the study area is a lot higher than the number of roads with soft road sides. This makes it easier for cyclists to choose an alternative route that is still pretty direct, but with more quiet roads than it is to choose a pretty direct route with more soft road sides. The fact that people add distance to their route in order to cycle on the relatively scarce roads with soft road sides underlines the significance of this attribute.

Entrances and exits (Exits_Abs)
The 'absolute amount of entrances and exits' has a coefficient value of 0.002 which means that for every additional entrance or exit along the route cyclists deviate 2 meters more than average from the shortest route. Just as with P_SoftRoadSides, the specific effect of this attribute on deviation from the shortest route or on route choice decision making in general has not been studied before. However, the effect of intersections in general on the route choice behaviour of cyclists has been studied before, which led to an expectation about the effect direction of this attribute.

The expectation for this attribute was that the coefficient value would have a minus sign. The main reason for this expectation is that the literature review revealed that a route is generally more likely to be chosen when the amount of intersections is lower. In addition, based on research from Stinson & Bhat (2003), it was expected that the shortest routes generally consist of more smaller (residential) streets with more exits and entrances as compared to the driven routes that were expected to consist more of segregated cycling facilities than run parallel to larger arterials.

The fact that 'Entrances and exits' attribute has a positive sign means that it does not match the expectation. A possible explanation for this is that the shortest routes are on average 340 meters shorter than the chosen routes, which means that the distance over which entrances and exits can be encountered is significantly smaller. As a result, it could be that the lower entrance and exit density on chosen routes is slightly outweighed by the fact that the shortest routes are on average 340 shorter. This theory would also explain the relatively small effect magnitude of this attribute.

Trees (Trees_Per100M)
The 'amount of trees per 100 metres of road' has a coefficient value of 0.23 which means that for every additional tree per 100 meters of route cyclists deviate 23 meters more than average from the shortest route. The direction of this effect is in line with earlier research by Krabbenborg et al. (2015), who found that the presence of trees along a route positively relates to that route being chosen. Whether the size of the effect is in line with earlier research is more difficult to say. Segadilha & Sanches (2014) found that the effect of trees along the route is moderate, but due to the different nature of our studies, it is impossible to compare the results.

Cyclist intensity (CycInt)
The 'average amount of cyclists encountered on a trip' has a coefficient value of -.005 which means that for every additional cyclist that is on average encountered on the route, cyclists deviate 5 meters less than average from the shortest route. This result indicates that cyclists prefer routes with lower cyclists intensities, which is unexpected because existing research

provides many indications that cyclists prefer higher cycling intensities. The most important indicator is provided by Larsen & El-Geneidy (2011) and Broach et al. (2012), who found that cyclists have a strong preference for segregated cycling facilities. Research by Larsen & El-Geneidy (2011) even shows that cyclists are willing to deviate from the shortest route if they can achieve greater separation from vehicle traffic. This general preference for segregated cycling facilities causes that cyclists intensities on cycle paths are higher than on roads that are an alternative for these cycle paths. Therefore, a direct consequence of the fact that cyclists prefer routes with segregated cycling facilities is that they indirectly also prefer routes with relatively high cyclist intensities.

A possible explanation for the unexpected result is that not all cyclists prefer separated cycling facilities and that this study focussed on cyclists that do not prefer separated cycling facilities. Research by Larsen & El-Geneidy (2011) seems to suggest that only recreational cyclists prefer segregated cycling facilities and thus higher cycling intensities. Research by Krabbenborg et al. (2015) even shows that some cyclists (mainly highly educated) tend to avoid busy cycle paths.

Age

The 'age of the respondent' predictor has a coefficient value of -.1 which means that older people (48-63 years) deviate 100 meters less than average from the shortest route and that younger people deviate 100 meters more than average from the shortest route. This is an expected result because recent literature has shown that older people generally add higher value to minimizing travel distance than younger people. This makes sense because older people are generally less healthy than younger people and might want to limit the physical effort they put into the trip. Another possible explanation for this result could be that older people just have slightly better knowledge of the transportation network which enables them to better determine the most direct path.

Insignificant attributes

As explained earlier in this chapter, some attributes that are incorporated in this study are found to be insignificant. In this study, this means that no evidence is found for a relationship between these attributes and the dependent variable 'difference in distance'. For the attributes 'Type of facility', 'Level of speed differences, 'Quality of pavement' and 'Parking' this is really unexpected as the literature review provides strong evidence for their role in the route choice behaviour of cyclists. For the attributes Land use, Urban density, Speed bumps, Average cycling speed, Gender, Weekend, Month, Peak and Daylight literature provides no evidence for their role in route choice decision making. However, based on logical reasoning, the insignificance of some of these attributes is also rather unexpected.

What the insignificance of these attributes could have possibly caused will be explained in the general discussion in the next chapter.

5. Conclusion

5.1. Introduction

The conclusion of this study consists of three parts that are presented in their own paragraph. Paragraph 5.2. presents the general conclusion of this study. This general conclusion provides an answer to the research question and fulfils the research objective by presenting recommendations for practice that can be used by transportation planners. Paragraph 5.3. presents the general discussion of this study. This section discusses the validity of the results as well as possible limitations of this study. Finally, in paragraph 5.4., recommendations for future research are presented.

5.2. General conclusion and recommendations for practice

In recent years, various studies have showed that cycling safety is related to the route choice behaviour of cyclists. However, no one yet studied how what this relationship exactly is about. This study contributes to filling this research gap by investigating how cycling safety affects the route choice behaviour of cyclists.

The objective of this study was to provide the Dutch transportation planners with information that will help them to better understand how the route choice behaviour of cyclists is related to cycling safety so that they can get more cyclists to use infrastructure that is safe for cyclists. The plan was to do this by providing insight into how various safety-related infrastructural attributes are related to the route choice behaviour of cyclists. This led to the formulation of the following research question:

> “What role do safety-related infrastructural attributes play in the extent to which cyclists deviate from the shortest route?”

By having compared 145 routes that are cycled in the Eindhoven region to their corresponding shortest routes and having estimated the coefficients of a Tobit regression model, this study presents the answer to this question. The answer is given by providing various insights into how safety-related attributes relate to the extent to which cyclists deviate from the shortest route and thereby influence the route choice behaviour of cyclists. Subsequently, based on these insights, an overarching conclusion is drawn that is also an advice for Dutch transportation planners and includes multiple recommendations.

First of all, this study shows that there are multiple safety-related infrastructural attributes that play a role in the extent to which cyclists deviate from the shortest route. These are: traffic intensity, road sides, entrances & exits, trees, and cyclists intensity. A higher percentage of softer road sides, a higher number of entrances & exits and a higher amount of trees along the route cause that cyclists are willing to deviate more than average from the shortest route. A higher traffic intensity, a higher cycling intensity and a higher age cause that cyclists are willing to deviate less than average from the shortest route. The main conclusion that can be drawn regarding the importance of these attributes relative to each other is that traffic intensity has a larger impact on the route choice behaviour of cyclists than road sides do.

Secondly, in contrast to what literature suggested, the results of this study show that when infrastructural attributes positively affect the route choice behaviour of cyclists they do not necessarily positively affect cycling safety as well. This became clear, because this study showed that encountering more entrances and exits (crossings) and higher cyclist intensities are positively related with the choice for a particular route, whereas recent literature clearly showed that higher levels of these attributes are negatively related with cycling safety.

Third and last, this study shows that age also plays a significant role in the extent to which cyclists deviate from the shortest route. By showing that older people are less willing to deviate from the shortest route than younger people, this study underlines the heterogeneity of the population. The effect of the attributes on route choice behaviour differs per individual and this is important to consider when designing bicycle infrastructure.

All in all, this study shows that it is possible for transportation planners to improve the cycling safety of a route while at the same time making the route more attractive for cyclists. Lowering traffic intensity and increasing the amount of soft road sides seem to be the easiest and most realistic design options that make infrastructure both more attractive for cyclists as well as safer. Planting more trees is an option to attract more cyclists to a route, but does not necessarily makes cycling safer. Furthermore, transportation planners need to keep in mind that the population is heterogeneous, which causes that not everyone's route choice behaviour is equally affected by the safety improving measures they take. This makes that it is important that transportation planners consider the composition of the population when designing safe and attractive bicycle infrastructure. Lastly, this study has highlighted multiple factors that should be included in travel models in order to make them reflect cycling behaviour more accurately. Given the fact that there is no literature available about the effect of road sides on route choice behaviour, chances are high that this attribute is not yet considered by at least some travel models. This makes the road side attribute particularly interesting for transportation planners.

5.3. General discussion

Despite the fact that the results of this study show many similarities with existing research, they do not completely line up. There are two main differences that stand out. First of all, some attributes of which current research suggest that they affect the route choice behaviour of cyclists ('Type of facility', 'Speed differences, 'Quality of pavement' and 'Parking'), were not found to be of significant influence in this study. Secondly and more importantly, the way in which some attributes affect the route choice behaviour of cyclist does not match the expectations created by existing literature. For two attributes ('Entrances and exits' and 'Cyclist intensity') positive effects were found were negative effects were expected and vice versa. There are several reasons that might explain the differences between the results of this study and those of existing literature.

A first possible cause of the discrepancies is the nature of the population sample that is used in this study. The sample drawn from the B-Riders project consisted of adults that are mostly between 30 and 60 years of age, that all have a job, are likely to be sporty and are experienced cyclists. As existing literature already pointed out, it could be possible that these, generally experienced, cyclists add great value to direct routes and are therefore less influenced by other attributes such as the level of segregation. In addition, certain attributes may have an opposite effect on the route choice behaviour of this specific group than they have on the route choice behaviour of the average, less experienced, cyclist. For example, it could be possible that where inexperienced cyclists like to cycle in places with high cycling intensities because it feels safer, experienced cyclists prefer to cycle in quiet places so that they can cycle faster.

A second reason is the relatively limited study area of this study. Since this study only focussed on the Eindhoven region, it only included one large city and not a lot of rural areas. Possibly, the unique way in which the infrastructure in the Eindhoven region is developed explains why for some attributes no significant relationship was found. For example, it could be possible that no significant relationship was found for the level of segregation, because Eindhoven has a lot of nice quiet roads that are almost just as good as a separated cycle path.

A third reason that might explain unexpected results of this study is the level of detail of the data files that are used and the way in which they are brought together. For most attributes, the data was very detailed. However, there were also attributes for which the data was not so detailed. In addition, due to the differences in the nature of the data files used, it was not always possible to combine them with 100% accuracy. Despite extensive manual and visual checks, it is possible that some road sections have not the most optimal attribute values. The lack of detail in combination with a certain percentage of not optimal attribute values could have played a role in finding unexpected results.

A fourth reason could be the way in which the attributes are operationalised. As explained before, it was decided to operationalise the attributes by using freely available data only. The result of this decision was that some attributes could not be exactly operationalised in the way of which existing research suggested that would be ideal. For example, quality of surface was operationalised based on the type of material instead of the actual quality as suggested by existing research.

A fifth and last reason that could explain the differences between the results of this study and those of existing literature is the sample size. As explained before, this study only used 145 unique trips made by 48 unique respondents. Although this sample is large enough to conduct a proper study, a larger sample might have resulted in fewer discrepancies between the results of this study and existing literature.

5.4. Recommendations for future research

On the basis of the discussion points presented in the previous paragraph, the first obvious recommendations for future research on this topic would be to:

- use a sample that represents the Dutch cycling population better;
- use a larger study area that is more diverse;
- use data files that are more detailed and achieve higher accuracy when combining them;
- use a larger sample that includes more respondents and more trips.

Furthermore, future research on this topic is recommended to incorporate intersections more extensively. This study only looked at crossing entrances and exits, but there are a lot more types of intersections in the Netherlands that differ from each other on the basis of their design (amount of lanes, separation of traffic, etc.) and how they regulate traffic (priority, traffic lights, etc.). Since existing literature suggested that these different types of intersections all have a different effect on the route choice behaviour and safety of cyclists, it would be very interesting to study these types of intersections as well.

A last recommendation is that future research incorporates travel time as such, instead of using distance as a proxy. Despite the fact that existing literature shows a linear relationship between travel time and distance, it might be more accurate to study travel time directly.

References

Allen, M.P. (1997). The problem of multicollinearity. In: Understanding Regression Analysis. Springer, Boston, MA

Amemiya, T. (1984). Tobit Models: A survey. Journal of Econometrics 24, Pg. 3-61. Doi: 10.1016/0304-4076(84)90074-5

ANWB. (2018a). Proactief meten veiligheid infrastructuur – CYCLERAP. Retrieved on 06-12-2019 from: https://www.anwb.nl/binaries/content/assets/anwb/pdf/belangenbehartiging/verkeer/cyclerap_factsheet_anwb_lr.pdf

ANWB. (2018b). Sterrensysteem voor veiliger fietsverkeer. Retrieved on 06-12-2019 from: https://www.anwb.nl/belangenbehartiging/verkeer/verkeersveiligheid/sterrensysteem-voor-veiliger-fietsverkeer

ANWB. (2020). Veilig op de fiets. Retrieved on 22-06-2020 from: https://www.anwb.nl/belangenbehartiging/verkiezingen/veilig-op-de-fiets

Bernhoft, I. M., & Carstensen, G. (2008). Preferences and behaviour of pedestrians and cyclists by age and gender. Transportation Research Part F: Traffic Psychology and Behaviour, 11(2), 83–95. Doi: 10.1016/j.trf.2007.08.004

Bollen, K. A. (1989). Structural Equations with Latent Variables. Wiley-Interscience Publication.

Bovy, P. H. L., and M. A. Bradley. (1985). Route Choice Analyzed with Stated-Preference Approaches. Transportation Research Record 1037, Pg. 11–20.

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, Pg. 409-432. Doi: 10.1007/s11116-011-9355-8

B-Riders. (2020). Fietsen naar het werk. Retrieved on 30-01-2020 from: http://www.briders.nl/b-riders-fietsen-naar-het-werk

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: 46, Pg. 1730–1740. Doi: 10.1016/j.tra.2012.07.005

CBS. (2017). Onderzoek Verplaatsingen in Nederland (OViN).

CBS. (2018). In 2017 meer verkeersdoden op de fiets dan in de auto. Retrieved on 29-09-2019 from https://www.cbs.nl/nl-nl/nieuws/2018/17/in-2017-meer-verkeersdoden-op-de-fiets-dan-in-de-auto

CBS. (2020). Retrieved on 04-02-2020 from: https://opendata.cbs.nl/statline/portal.html?_la=nl&_catalog=CBS

Cho, G., Rodríguez, D.A., Khattak, A.J. (2009). The role of the built environment in explaining relationships between perceived and actual pedestrian and bicyclist safety. Accident Analysis and Prevention: 41, Pg. 692-702. Doi: 10.1016/j.aap.2009.03.008

Chirstmas, S., Helman, S., Buttress, S., Newman, C. (2010). Cycling, Safety and Sharing the Road: Qualitative Research with Cyclists and Other Road Users. Road Safety Web Publication, 17.

Claasen, Y., Rienstra, S. (2017). Routekeuzegedrag van fietsers: meer dan alleen de snelste route telt. Antwerpen: Collocquium Vervoersplanologisch Speurwerk (CVS)

Cox, D. R. (2006). Principles of Statistical Inference. Cambridge University Press.

Daniels, S., Nuyts, E., Wets, G. (2008). The effects of roundabouts on traffic safety for bicyclists:

An observational study. Accident Analysis and Prevention: 40, Pg. 518-526. Doi: 10.1016/j.aap.2007.07.016.

Daniels, S., Wets, G. (2005). Traffic safety effects of roundabouts: a review with emphasis on bicyclist's safety. Proceedings of the 18th ICTCT-Workshop. Helsinki, Finland, www.ictct.org.

Dill, J., Gliebe, J., (2008). Understanding and Measuring Bicycling Behavior: a Focus on Travel Time and Route Choice. Oregon Transportation Research and Education Consortium, Portland, OR. Doi: 10.15760/trec.151

Doherty, S.T., Aultman-Hall, L., Swaynos, J., (2000). Commuter cyclist accident patterns in Toronto and Ottawa. Journal of Transportation Engineering: 126, Pg. 21–26.

Dollisson, J., Cox, W., Zheng, Z. (2013). Comparison of Motorists' and Cyclists' perception of bicycle safety. In 36th Australasian Transport Research Forum (ATRF) Annual Conference, 2 - 4 October 2013, QUT Gardens Point, Brisbane, Australia.

Eindhoven Open Data. (2020). Retrieved on 20-03-2020 from: https://data.eindhoven.nl /pages/home/?flg=nl

Elvik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. Accident Anal. Prevent: 41(4), Pg. 849–855. Doi: 10.1016/j.aap.2009.04.009

Fabriek, E., De Waard, D., & Schepers, P. (2012). Improving the visibility of bicycle infrastructure. International Journal of Human Factors and Ergonomics: 1(1), Pg. 98-115. Doi: 10.1504/IJHFE.2012.045274

Feng, T., & Timmermans, H. J. P. (2018). Integrated data imputation and annotation tool: trace annotator. In P. Kiefer, H. Huang, N. van de Weghe, & M. Raubal (Eds.), Adjunct Proceedings of the 14th International Conference on Location Based Services (Pg. 221-222). Zurich: ETH Zürich. Doi: 10.3929/ethz-b-000225623

FHWA. (1998). The Bicycle Compatibility Index – Implementation manual. Retrieved on 06-12-2019 from: https://safety.fhwa.dot.gov/tools/docs/bci.pdf

Field, A. (2009). Discovering statistics using SPSS. Third edition.

Fietsersbond. (2020). Snelfietsroutes. Retrieved on 22-06-2020 from: https://www.fietsersbond.nl/ons-werk/mobiliteit/snelfietsroutes/

Fietstelweek. (2020). Retrieved on 23-03-2020 from: http://fietstelweek.nl/data/

Gemeente Utrecht. (2020). Dafne Schippersbrug. Retrieved on 23-06-2020 from: https://www.anwb.nl/belangenbehartiging/verkeer/verkeersveiligheid/sterrensysteem-voor-veiliger-fietsverkeer

Götschi, T., Castro, A., Deforth, M., Miranda-Moreno, L., Zangenehpour, S. (2018). Towards a comprehensive safety evaluation of cycling infrastructure including objective and subjective measures. Journal of Transport and Health: 8, Pg. 44-54.

Greene, W.H. (2016a). LIMDEP Version 11 Economic modelling guide.

Greene, W.H. (2016b). LIMDEP Version 11 reference guide.

Heinen, E., Van Wee, B., Maat, K. (2010). Commuting by bicycle: an overview of the literature. Transport Reviews: 30 (1), Pg. 59–96.

Hendriks, T. (2018). Fietsongevallen, een analyse van oorzaken. Verkeerskunde: 2018-6, Pg. 134-136. Retrieved on 13-11-2019 from: http://www.verkeerskunde.nl/Uploads/2018/6/Fietsongelukken--Ton-Hendriks.pdf

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). Applied Choice Analysis (Second Edi).

Cambridge University Press. Doi: 10.1017/CBO9780511610356
Hood, J., Sall, E., Charlton, B. (2011). A GPS-based route choice model for San Fransisco California. Transportation Letters: The International Journal of Transportation Research: 3(1), Pg. 63-75. Doi: 10.3328/TL.2011.03.01.63-75
Hummel, T. (2001). Land use planning in Safer Transportation Network Planning – Safety principles, planning framework and library information.
Huisman, D. J., Hengeveld, J. (2014). Opmaak van een methode voor fietstellingen in de provincie Antwerpen.
ITF (International Transport Forum) – OECD. (2018). Cycling Safety – Summary and conclusions – Roundtable 168.
Isaksson-Hellman, I. (2012). A study of bicycle and passenger car collisions based on insurance claims data. If Insurance Company P&C Ltd.
Jacobsen, P.L. (2003). Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Injury Prevention: 9(3), Pg. 205-209. Doi: 10.1136/ip.9.3.205.
Jain, H., Tiwari, G., Zuidgeest, M. (2010). Evaluating bicyclists comfort and safety perception. 12th WCTR, July 11-15, 2010 – Lisbon, Portugal.
Johnson, M., Newstead, S., Oxley, J., Charlton, J. (2013). Cyclists and op vehicle doors: Crash characteristics and risk factors. Safety Science: 59, Pg. 135-140. Doi: 10.1016/j.ssci.2013.04.010.
Kaplan, S., Prato, C.G. (2015). A Spatial Analysis of Land Use and Network Effects on Frequency and Severity of Cyclist–Motorist Crashes in the Copenhagen Region. Traffic Injury Prevention: 16 (7), Pg. 724-731. Doi: 10.1080/15389588.2014.1003818.
KiM (Kennisinstituut voor Mobiliteitsbeleid). (2018). Kerncijfers Mobiliteit 2018.
Kim, J. -K., Kim, S., Ulfarsson, G. F., & Porrello, L. A. (2007). Bicyclist injury severities in bicycle–motor vehicle accidents. Accident Analysis & Prevention: 39, Pg. 238–251. Doi: 10.1016/j.aap.2006.07.002
Krabbenborg, L., Annema, J.A., Snellen, D. (2015). De invloed van bebouwde omgeving op fietsen in voortransport. Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk, Antwerpen, België.
Kroeze, P., Sweers, W., Schepers, P. (2010). Fietsers en oversteekongevallen. Retrieved on 13-11-2019 from: http://www.verkeerskunde.nl/fietsongevallen
Lankhuijzen, R., Ruijs, K., Orsouw van, M. (2016). Drukte op fietspaden in Zeeland, eindrapport. XTNT, Utrecht.
Larsen, J., El-Geneidy, A. (2011). A travel behavior analysis of urban cycling facilities in Montreal Canada. Transportation Research Part D 16 (2), Pg. 172–177. Doi: 10.1016/j.trd.2010.07.011
Lawson, A.R., Pakrashi, V., Ghosh, B., Szeto, W.Y. (2013). Perception of safety of cyclists in Dublin City. Accident Analysis and Prevention: 50, Pg. 499-511. Doi: 10.1016/j.aap.2012.05.029
Lim, C., Sen, P.K., Pedadda, S.D. (2012). Accounting for Uncertainty in Heteroskedasticity in Nonlinear regression. Stat Plan Inference 142(5), Pg. 1047–1062. Doi: 10.1016/j.jspi.2011.11.003
Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). Stated Choice Methods - Analysis and Applications. UK: Cambridge University Press.

Lusk, A.C., Furth, P.G., Morency, P., Miranda-Moreno, L.F., Willett, W.C., Dennerlein, J.T. (2011). Risk of injury for bicycling on cycle tracks versus in the street. British Medical Journal 17(2), Pg. 131-135. Doi: 10.1136/ip.2010.028696

Manton, R., Rau, H., Clifford, E., Fahy, F. (2013). Is cycling safe? Using mental mapping to unpack perceptions of cycling safety. Proceedings of the ITRN, September 5-6, 2013. Trinity College Dublin.

Ng, A., Debnath, A.K., Heesch, K.C. (2017). Cyclist' safety perceptions of cycling infrastructure at un-signalized intersections: Cross-sectional survey of Queensland cyclists.

Openstreetmap. (2020). Retrieved on 25-03-2020 from: https://www.openstreetmap.org

Parkin J.; Wardman, M.R.; Page, M. (2007). Models of Perceived Cycling Risk and Route Acceptability. Accident Analysis and Prevention: 39, Pg. 364-371. Doi: 10.1016/j.aap.2006.08.007

Reurings, M.C.B., Vlakveld, W.P., Twisk, D.A.M., Dijkstra, A. et al. (2012). Van fietsongeval naar maatregelen: kennis en hiaten; Inventarisatie ten behoeve van de Nationale Onderzoek agenda Fietsveiligheid (NOaF). R-2012-8. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam.

Rietveld, P., Zwart, B., van Wee, B., van den Hoorn, T. (1999). On the relationship between travel time and travel distance of commuters – Reported versus network travel data in the Netherlands. The Annals of Regional Science 33: Pg. 269-287

Rijksoverheid. (2018). Meer mensen op de fiets. Retrieved on 29-09-2019 from https://www.rijksoverheid.nl/onderwerpen/fiets/fietsbeleid

Rijkswaterstaat. (2020). Index of/geodata/dmc/nwb-wegen/geogegevens/shapefile. Retrieved on 20-01-2020 from: https://www.rijkswaterstaat.nl/apps/geoservices/geodata/dmc/nwb-wegen/geogegevens/shapefile/

Saylor Acadamy. (2020). Utility Theory. Retrieved on 30-01-2020 from:https://saylordotorg.github.io/text_risk-management-for-enterprises-and-individuals/s07-01-utility-theory.html

Schepers, P. (2008). De rol van infrastructuur bij enkelvoudige fietsongevallen. Rijkswaterstaat Dienst Verkeer en Scheepvaart.

Schepers, P. & den Brinker, B. (2011). What do cyclists need to see to avoid single-bicycle crashes?. Ergonomics: 54:4, Pg. 315-327. Doi: 10.1080/00140139.2011.558633

Schepers, P., Hagenzieker, M., Methorst, R., van Wee, B., Wegman, F. (2014). A conceptual framework for road safety and mobility applied to cycling safety. Accident Analysis and Prevention: 62, Pg. 331-340. Doi: 10.1016/j.aap.2013.03.032

Schepers, P., Kroeze, P.A., Sweers, W., Wüst, J.C. (2011). Road factors and bicycle-motor vehicle crashes at unsignalized priority intersections. Accidents Analysis and Prevention: 43, Pg. 853-861. Doi: 10.1016/j.aap.2010.11.005.

Schepers, P., Twisk, D., Fishman, E., Fyhri, A., Jensen, A. (2017). The Dutch road to a high level of cycling safety. Safety Science: 92, Pg. 264-273. Doi: 10.1016/j.ssci.2015.06.005.

Schepers, J.P., Voorham, J. (2010). Oversteekongevallen met fietsers – Het effect van infrastructuurkenmerken op voorrangskruispunten. Rijkswaterstaat Dienst Verkeer en Scheepvaart.

Schneider, A., Hommel, G., Blettner, M. (2010). Linear regression analysis—part 14 of a series on evaluation of scientific publications. Dtsch Arztebl Int 107(44), Pg. 776–82. Doi: 10.3238/arztebl.2010.0776

Schoon, C.C., Blokpoel, A. (2000). Frequentie en oorzaken van enkelvoudige fietsongevallen – Een ongevallenanalyse gebaseerd op een enquête onder fietsslachtoffers. Rapportnummer: R-2000-20.

Sener, I.N., Eluru, N. & Bhat, C.R. (2009). An analysis of bicycle route choice preferences in Texas, US. Transportation: 36, Pg. 511–539. Doi: 10.1007/s11116-009-9201-4

Snizek, B., Nielsen, T.A.S., Skov-Petersen, H. (2013). Mapping bicyclists' experiences in Copenhagen. Journal of Transport Geography: 30, Pg. 227-233. Doi: 10.1016/j.jtrangeo.2013.02.001

Sørensen, M., Mosslemi, M. (2009). Subjective and Objective Safety – The effect of road safety measures on subjective safety among vulnerable road users. TØI report 1009/2009

Stinson, M., & Bhat, C. (2003). Commuter Bicyclist Route Choice: Analysis Using a Stated Preference Survey. Transportation Research Record: 1828(1), Pg. 107–115. Doi: 10.3141/1828-13

Stone, M. and Gosling, R. (2008). Attitudes to Cycling 2008: Research Report. London: Transport for London.

Thomas, B., DeRobertis, M. (2013). The safety of urban cycle tracks: A review of the literature. Accident Analysis and Prevention: 52, Pg. 219-227. Doi: 10.1016/j.aap.2012.12.017

Vandebona, U. and Kiyota, M. (2001). Safety Perception Issues Related to Pedestrians and Cyclists. Tasmania: Department of Infrastructure, Energy and Resources.

van der Waerden, P. J. H. J., Willems, R., & van Dongen, R. P. (2018). Voorkeuren van fietsers met betrekking tot de hoogte en plaatsing van bomen langs fietsinfrastructuur in het stedelijke gebied. 1-13. Paper gepresenteerd op Colloquium Vervoersplanologisch Speurwerk, Amersfoort, Nederland.

van Goeverden, K., Godefrooij, T. (2010). Ontwikkeling van het fietsbeleid en –gebruik in Nederland. Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 25 en 26 november 2010, Roermond.

Van Overdijk, R. P. (2016). The influence of comfort aspects on route- and mode-choice decisions of cyclists in the Netherlands. Technical University Eindhoven, The Built Environment.

Vanparijs, J., Int Panis, L., Meeusen, R., de Geus, B. (2015). Exposure measurement in bicycle safety analysis: A review of the literature. Accident Analysis and Prevention: 84, Pg. 9-19. Doi: 10.1016/j.aap.2015.08.007

VeiligheidNL. (2016). Fietsongevallen in Nederland – SEH-behandelingen 2016. Rapport nummer 679.

Weijermars, W., Korving, H., van Schagen, I., Goldenbeld, Ch., Bos, N., Stipdonk, H. (2016). Monitor verkeersveiligheid 2016 – Toename verkeersdoden en ernstig verkeersgewonden.

Wijlhuizen & G.J., Aarts, L.T. (2014). Monitoring fietsveiligheid – Safety Performance Indicators (SPI's) en een eerste opzet voor een gestructureerd decentraal meetnet.

Wijlhuizen, G.J., van Petegem, J.W.H., Goldenbeld, Ch., van Gent, P., de Bruin, J., Commandeur, J.J.F., Kars, V. (2016). Doorontwikkeling CycleRAP-instrument voor veiligheidsbeoordeling fietsinfrastructuur - Doelmatigheid handmatige intensiteitsmetingen, betrouwbaarheid beoordelingen infrastructuur en validiteit van het CycleRAP-instrument.

Winters, M., Teschke, K., Grant, M., Setton, E.M., Brauer, M. (2010). How far out of the way will we travel? Built environment influences on route selection for bicycle and car travel. Transportation Research Record (2190), 1–10. Doi: 10.3141/2190-01

Yang, C., & Mesbah, M. (2013). Route Choice Behaviour of Cyclists by Stated Preference and Revealed Preference. Australasian Transport Research Forum 2013 Proceedings, (October). Retrieved on 28-11-2019 from: http://atrf.info/papers/2013/2013_yang_mesbah.pdf

Appendix A – Background information on cycling

This appendix elaborates on the cycling situation in the Netherlands and provides some additional information on the pros and cons of cycling.

Cycling in the Netherlands

When you ask someone to name a few things that are typical Dutch, cycling is very likely to be one of them. Contrary to the idea that all Dutch people wear wooden shoes, this stereotyping is justified. Dutch people cycle a lot and cycling is an essential part of the Dutch identity. News outlets seem to indicate that there are many reasons why cycling became so popular in the Netherlands that it got woven into the Dutch culture. However, they almost all suggest that the most important reason why cycling could ever become as popular as it is today in the Netherlands is that the Netherlands is a very small-scaled, densely populated and extremely flat country. Also the fact that the Dutch government started to invest seriously in cycling infrastructure from the '70s onwards and other impulses like the oil crisis of 1973 are mentioned in various news outlets as reasons for the enormous popularity of cycling in the Netherlands today. Some of the more recent reasons of which various news outlets suggest that they contribute to today's cycling popularity are the health benefits associated with cycling and environmental awareness.

Despite the fact that cycling is already very big in the Netherlands, the popularity of cycling is still growing. According to the Kennisinstituut voor Mobilitietsbeleid (KiM) (2018), who base their information on CBS (2017), in 2017, Dutch people travelled approximately 15,5 billion kilometres by bicycle. That is an increase of 1.5 billion kilometres as compared to 2005, when they travelled around 14 billion kilometres. This large number of kilometres travelled by bicycle makes cycling the second most important means of transport in the Netherlands. In 2016, cycling was responsible for 27% of daily mobility in the Netherlands. Only the car was more important and accounted for 47% of daily mobility. Third place was for walking with 18% (KiM, 2018). According to ITF (2018), who bases his data on Castro and Götschi (2018), ITF (2013) and the IRTAD database, the Netherlands is world leader when it comes to cycling kilometres per person per year. On average, Dutch people travel 891 kilometres a year by bicycle and are followed at a large distance by Denmark (547), Germany (439) and Belgium (279). This comparison underlines once more how extremely popular cycling is in the Netherlands and that cycling is a considerable mode of transport within Dutch mobility.

When it comes to the future of cycling in the Netherlands, the Dutch government has a clear ambitions. The Rijksoverheid (2018) explains on their website that their ambition for coming period of government is to get 200.000 extra commuters by bicycle or by bicycle in combination with public transport. This ambition is part of their overarching goal of having 20% more cycling kilometres in the Netherlands by 2027. They furthermore explain that in order to achieve their goal, the Dutch government has made 100 million euros available for coming period of government and municipalities and provinces are adding another 245 million. The available millions will be used to construct and adapt fast cycle routes and to build and expand bicycle parking facilities.

The benefits of cycling

The advantages of cycling are undeniable and diverse. First, there are the health benefits. Recent studies have shown that cycling offers cardiorespiratory fitness benefits and that it reduces cardiovascular risk factors (Oja et al., 2011; Warburton, 2006). Cycling also helps to reduce the risk of getting type 2 diabetes and some types of cancer (Celis-Morales et al, 2017; Warburton, 2006; Nijland, 2017). Woodcock et al. (2014) showed with their study that these benefits get larger as people get older. According to ITF (2018), this is mostly due to the fact that older people more often than younger people suffer from the diseases which cycling helps preventing. In addition to these physical health benefits, cycling can also have mental health benefits. Woodcock et al. (2014) found that cycling can contribute to reducing depression.

Second, cycling has large environmental benefits. Using a bicycle instead of a car leads to lower greenhouse gas emissions and cleaner air. According to Harms & Kansen (2018), switching from a car to a bicycle saves 150 grams of CO_2, 0.2 grams of NO_x and 0.01 grams of particulate matter per kilometre. They furthermore explain that cars are used for 3.6 billion short trips (< 7.5 km) annually and replacing all these short trips by cycling would save about 2 megatons of CO_2, 2.6 kilotons of NO_x and 0.13 kilotons of particulate matter each year.

Third, cycling has some significant practical benefits. Due to the fact that the Netherlands is a small and densely populated country with a lot of areas that are not very suitable for car traffic (e.g. old inner-city areas), getting somewhere by bicycle is sometimes a lot faster than getting there by car. In addition to this, it is often a lot easier to park a relatively small bicycle than it is to park a car. Furthermore, cycling is just a lot cheaper than driving a car. Bicycles are a lot cheaper to purchase, maintain and insure compared to cars and bicycles also do not require fuel to ride on.

Fourth and last, cycling has some serious economic benefits. The economic benefit that is most often mentioned is congestion relief. Congestion can be reduced by a modal shift from car to cycling, because bicycles require less space on the road than cars do (Litman, 2019). The reason why congestion relief is an economical benefit is because congestion costs money and the less congestion there is, the lower the economic damage will be (Panteia, 2018). Another lesser-known economic benefit of cycling is job creation. According to Blondiau et al. (2016), cycling creates jobs because bicycles need to be manufactured, sold and serviced, but also because cycling infrastructure has to be created. However, the largest contributor to cycling related jobs is cycling tourism. Blondiau et al. (2016) estimate that today over 650 thousand jobs are linked to cycling in the European Union. They furthermore state that doubling the modal share of cycling could add over 400 thousand new jobs to this already high number.

The drawbacks of cycling

Despite the fact that cycling has many benefits compared to car driving, it also has some drawbacks. These drawbacks can roughly be divided into two categories, namely: Disadvantages of cycling itself and barriers to choosing the bicycle as a means of transport.

When it comes to disadvantages of cycling, the most important one is that when you are on a bicycle, you are very vulnerable. A bicycle does not offer any protection such as a seat belts, crush zones or air bags and having an accident on a bicycle can therefore easily lead to serious injury or even death. CBS (2018) explains on their website that in 2017, 613 road fatalities occurred in the Netherlands. Of these 613 road fatalities, 206 were cyclists and 201 were car drivers/occupants. This number is more or less the same, however, the distance travelled by car in 2017 year was almost 10 times greater (KiM, 2018). This means that cycling is more unsafe per kilometre travelled then being a car driver/passenger. CBS (2018) furthermore explain that in the last twenty years the number of cycling fatalities has hardly decreased in the Netherlands and fluctuates around 200 casualties a year. The number of road fatalities among car drivers, however, has fallen sharply in the last twenty years. Where more than 600 car drivers/passengers died in 1998, only 200 died in 2017. According to the data provided by CBS (2018), 2017 was the first year in which the number of cyclists who lost their lives in traffic was higher than the number of car drivers/passengers who lost their lives in traffic. Also when it comes to road users getting seriously injured, cycling scores very poorly compared to other means of transport. According to Weijermars et al. (2018) in 2017, 64% of all seriously injured road users in the Netherlands were cyclists, followed at a large distance by motor/scooter drivers (19%), car drivers/occupants (11%) and pedestrians (5%).

A second disadvantage of cycling is that when you are on a bicycle, you are likely to breathe in a significant amount of polluted air. De Hartog et al. (2010) explain that despite the fact that cyclists breath cleaner air than car drivers do, the higher minute ventilation due to physical activity causes that cyclists ultimately inhale more harmful substances than car drivers do. They estimate that switching from car to bicycle on average results in a loss of 21 life days per person.

When it comes to barriers to choosing the bicycle as a means of transport, the most important one is the fact that as a cyclist you are directly influenced by weather conditions. Research by Rietveld et al. (2012) shows that of all modes of transport, cycling is the most sensitive to weather conditions and that temperature and rainfall have the largest influence on bicycle use. Because of this sensitivity to weather conditions there is a chance that cyclists sometimes arrive wet or sweaty at their destination. As a result, cycling is in some situations seen as a less comfortable means of transport which makes people prefer to go, for example, by car.

A second barrier to choosing the bicycle as a means of transport is the lack of bicycle parking facilities. According to Stinson & Bath (2004) and Hunt & Abraham (2007), bicycle use depends to a certain extent on whether the bicycle can be safely stored at the destination. Their studies show that the presence of bicycle parking facilities has a positive effect on cycling. However, there are not enough bicycle parking facilities in the Netherlands. For example, around public transport stops (Fietsersbond, 2017). This lack of parking facilities is confirmed by the ambition of the Dutch government to invest in bicycle parking facilities, which is explained on their website (Rijksoverheid, 2018).

A last barrier to choosing the bicycle as a means of transport is that it requires some skill and a certain amount of endurance. First of all, as a cyclist, you have to be able to maintain your balance and manoeuvre your bicycle. Secondly, you must have sufficient endurance to be able to reach your destination in a comfortable way. For the majority of people this will not cause any difficulties, but for elderly people or people with a handicap this could definitely be a problem.

All in all, cycling has some clear drawbacks compared to driving a car. However, these drawbacks are generally outweighed by the health, environmental, practical and economic benefits that cycling has to offer. De Hartog et al. (2010) explain that on average, the health benefits of cycling alone already outweigh the risks when compared to car driving. When it comes to the question if air pollution can negate the health benefits of cycling, Tainio et al. (2016) explain that the benefits from cycling generally outweigh health risks from air pollution. Only in a small number of cities with an extremely high particulate matter concentration the risks of cycling could outweigh the benefits. Whether the benefits of cycling are large enough to overcome the barriers to cycling is highly person dependent. However, looking at the above statements, the answer is in most cases very likely to be yes.

References

Blondiau, T., van Zeebroeck, B., Haubold, H. (2016). Economic benefits of increased cycling. Transportation Research Procedia 14. Pg. 2306-2313. Doi.org/10.1016/j.trpro.2016.05.247

Castro, A. , Kahlmeier, S., Götschi, T. (2018). Exposure-adjusted road fatality rates for cycling and walking in European countries, International Transport Forum Discussion Paper

CBS. (2017). Onderzoek Verplaatsingen in Nederland (OViN).

CBS. (2018). In 2017 meer verkeersdoden op de fiets dan in de auto. Retrieved on 29-09-2019 from https://www.cbs.nl/nl-nl/nieuws/2018/17/in-2017-meer-verkeersdoden-op-de-fiets-dan-in-de-auto

Celis-Morales, C.A., Lyall, D.M., Welsh, P., Anderson, J., Steell, L., Guo, Y., Maldonado, R., Mackay, D.F., Pell, J.P., Sattar, N., Gill, J.M.R. (2017). Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study. BMJ 2017;357:j1456. Doi: https://doi.org/10.1136/bmj.j1456

de Hartog, J.J., Boogaard, H., Nijland, H., Hoek, G. (2010). Do the health benefits of cycling outweigh the risks? Environmental Health Perspectives: 118, Pg. 1109-1116.

Fietsersbond. (2017). Fietsparkeren – Bij stations en haltes. Retrieved on 13-12-2019 from: https://files.fietsersbond.nl/app/uploads/2017/11/01073729/Rapport-fietsparkeren.pdf

Harms, L., Kansen, M. (2018). Cycling Facts. Netherlands Institute for Transport Policy Analysis – KIM – Ministry of Infrastructure and Water Management.

Hunt, J. D., Abraham, J. E. (2007) Influences on bicycle use. Transportation: 34(4), Pg. 453–470. Doi: 10.1007/s11116-006-9109-1

ITF – OECD. (2013). Cycling, Health and Safety, OECD Publishing, Paris

ITF – OECD. (2018). Cycling Safety – Summary and conclusions – Roundtable 168.

KiM (Kennisinstituut voor Mobiliteitsbeleid) (2018). Kerncijfers Mobiliteit 2018.

Litman, T. (2019). Smart Congestion Relief - Comprehensive Evaluation of Traffic Congestion Costs and Congestion Reduction Strategies. Victoria Transport Policy Institute.

Nijland, H. (2017). Fietsen leidt tot langer en gezond leven. Planbureau voor de leefomgeving.

Oja, P., Titze, S., Bauman, A., de Geus, B., Krenn, P., Reger-Nash, B., Kohlberger, T. (2011). Health Benefits of Cycling: A systematic review. Scandinavian Journal of Medicine & Science in Sports. Pg 1-14. Doi: 10.1111/j.1600-0838.2011.01299.x

Panteai. (2018). Economische Wegwijzer 2018. Retrieved on 15-10-2019 from https://www.tln.nl/actueel/nieuws/PublishingImages/Paginas/Fileschade-vrachtverkeer-stijgt-naar-1komma3-miljardeuro/Panteia%20Economische%20Wegwijzer%202018%20nov18.pdf

Rietveld, P., Sabir, M., van Ommeren, J. (2012). Fietsen door weer en wind: Een analyse van de invloed van weer en klimaat op fietsgebruik. Tijdschrift Vervoerswetenschap: 48 -4, Pg. 46-59.

Rijksoverheid. (2018). Meer mensen op de fiets. Retrieved on 29-09-2019 from: https://www.rijksoverheid.nl/onderwerpen/fiets/fietsbeleid

Stinson, M. A., Bhat, C. (2004). Frequency of bicycle commuting: internet-based survey analysis. Transportation Research Record: 1878, Pg. 122–130. Doi: 10.3141/1878-15

Tainio, M., de Nazelle, A.J., Götschi, T., Kahlmeier, S., Rojas-Rueda, D., Nieuwenhuijsen, M.J., de Sá, T.H., Kelly, P., Woodcock, J. (2016). Can air pollution negate the health benefits of cycling and walking? Preventive Medicine. Pg. 233-236. Doi: 10.1016/j.ypmed.2016.02.002

Warburton, D.E., Nicol, C.W., Bredin, S.S. (2006). Health benefits of physical activity: the evidence. CMAJ. 2006;174:801–809. Doi: 10.1503/cmaj.051351

Weijermars, W., Goldenbeld, Ch., de Goede, M., Moore, K., Mons, C., Bijleveld, F.D., Bos, NM. - SWOV. (2018). Monitor Verkeersveiligheid 2018 – Achtergrondinformatie en onderzoeksverantwoording. R-2018-16A

Woodcock, J., Tainio, M., Cheshire, J., O'Brien, O., Goodman, A. (2014). Health effects of the London Bicycle sharing system: health impact modelling study. BMJ 2014;348:g425 Doi: 10.1136/bmj.g425

Appendix B – Choice behaviour & Utility theory

This appendix discusses choice behaviour and aims to clarify the process of making choices. First, the basic principles of choice making are described. Second, the utility theory is explained.

The basic principles of choice making

People make choices every day. Relatively simple choices, such as deciding on what we will have for dinner, but also more difficult choices that, for example, relate to our work or study. Some of these choices are made consciously. Especially the important ones. However, there are also many choices that we make subconsciously. These are mostly choices that are less important and are made very often. Since choosing refers to the ability to choose between multiple options, people need at least two options to choose from in order to be able to make choices.

Cyclists also constantly make choices. For example, when they make decisions about what route they will take. There are several theories regarding route choice behaviour. However, the choice theory that is most often used in route choice modelling is the utility theory.

The utility theory

The Saylor Academy (2020) explains on their website that the utility theory is originally developed in economics and aims to explain the observed behaviour of individuals. In other words, this theory is developed to find out what value people attach to certain attributes in order to explain why they make certain choices. According to Louviere et al. (2000), the utility theory assumes that individuals always try to choose the alternative that will benefit them most. A principal that is referred to as utility maximisation. Within route choice modelling research, utility maximisation means that travellers choose the route of which they think that it suits their preferences best, given its attributes.

When it is assumed that utility maximisation takes place and that the choice for a specific alternative within a choice set is based on observed and unobserved (hidden) attributes, the utility (U_{ai}) for choosing alternative a by individual i can be expressed as shown in equation 1. In this equation V_{ai} represents the observed attributes for utility and ε_{ai} represents a random component that reflects the unobserved attributes (Hensher et al., 2015).

$$U_{ai} = V_{ai} + \varepsilon_{ai} \tag{1}$$

Given the fact that this equation contains a random component, the observed utility (V_{ai}) is not equal to the actual utility (U_{ai}). The observed utility of alternative a for individual i can be defined as a function of k variables x_{aik} with parameter estimates β as shown in equation 2 (Hensher et al., 2015).

$$V_{ai} = f(x_{aik}, \beta) \tag{2}$$

This equation is often translated to a linear function that represents the observed utility by assuming a linear relationship between the observed utility (V_{ai}) and the parameter estimates (β_k) multiplied with the attribute variables (x_{aik}) (Hensher et al., 2015). This function is shown in equation 3.

$$V_{ai} = \sum_{k=1}^{K} \beta_k \, x_{aik} \tag{3}$$

When the utility of the attributes and alternatives are known, it is possible to calculate the probability that individual *i* chooses alternative *a* over alternative *b* (within the same choice set) by using the behavioural model shown in equation 4 (Train, 2002).

$$P_{ai} = prob(U_{ai} > U_{bi} \;\forall\, b \neq a) \tag{4}$$

References

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). Applied Choice Analysis (Second Edi). Cambridge University Press. Doi: 10.1017/CBO9780511610356

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). Stated Choice Methods - Analysis and Applications. UK: Cambridge University Press.

Saylor Acadamy. (2020). Utility Theory. Retrieved on 30-01-2020 from: https://saylordotorg.github.io/text_risk-management-for-enterprises-and-individuals/s07-01-utility-theory.html

Train, K. (2002). Discrete Choice Methods with Simulation. Cambridge University Press. Doi: 10.1017/CBO9780511753930

Appendix C – Correlations and regression output

Correlations

		Dist	PBuilt	PHigh	PSlow	PQuie	PSoft	PNoP	PSeg	PHQS	ExAbs	Trees	SBAbs	CycInt	CycSp
Dist	Pearson Correlation	1	-,161	-,186*	,208*	,244**	,155	,025	,121	-,140	,098	,170*	,010	-,366**	-,018
	Sig. (2-tailed)		,053	,025	,012	,003	,062	,762	,146	,094	,242	,042	,907	,000	,830
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
PBuilt	Pearson Correlation	-,161	1	,408**	-,043	-,134	-,173*	-,203*	-,103	,019	-,186*	-,115	,094	,032	-,230**
	Sig. (2-tailed)	,053		,000	,603	,109	,037	,014	,216	,822	,025	,169	,262	,702	,005
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
PHigh	Pearson Correlation	-,186*	,408**	1	-,042	-,097	-,187*	-,260**	-,090	-,072	-,031	-,085	,272**	,165*	-,006
	Sig. (2-tailed)	,025	,000		,617	,245	,025	,002	,281	,388	,714	,310	,001	,048	,944
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
PSlow	Pearson Correlation	,208*	-,043	-,042	1	,570**	,076	,399**	,475**	-,040	-,328**	-,043	-,252**	-,296**	-,194*
	Sig. (2-tailed)	,012	,603	,617		,000	,364	,000	,000	,637	,000	,608	,002	,000	,019
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
PQuie	Pearson Correlation	,244**	-,134	-,097	,570**	1	,127	,599**	,840**	,114	-,463**	-,042	-,351**	-,360**	-,106
	Sig. (2-tailed)	,003	,109	,245	,000		,129	,000	,000	,173	,000	,617	,000	,000	,206
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
PSoft	Pearson Correlation	,155	-,173*	-,187*	,076	,127	1	,206*	,116	,057	-,107	-,009	,017	,043	,219**
	Sig. (2-tailed)	,062	,037	,025	,364	,129		,013	,165	,499	,200	,919	,841	,605	,008
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
PNoP	Pearson Correlation	,025	-,203*	-,260**	,399**	,599**	,206*	1	,596**	,193*	-,522**	,014	-,369**	-,121	-,037
	Sig. (2-tailed)	,762	,014	,002	,000	,000	,013		,000	,020	,000	,867	,000	,146	,654
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
PSeg	Pearson Correlation	,121	-,103	-,090	,475**	,840**	,116	,596**	1	,442**	-,436**	-,024	-,482**	-,170*	,024

	Sig. (2-tailed)	,146	,216	,281	,000	,000	,165	,000		,000	,000	,771	,000	,041	,776
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
PHQS	Pearson Correlation	-,140	,019	-,072	-,040	,114	,057	,193*	,442**	1	-,027	-,070	-,319**	,361**	,097
	Sig. (2-tailed)	,094	,822	,388	,637	,173	,499	,020	,000		,752	,400	,000	,000	,243
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
ExAbs	Pearson Correlation	,098	-,186*	-,031	-,328**	-,463**	-,107	-,522**	-,436**	-,027	1	,006	,378**	,087	,132
	Sig. (2-tailed)	,242	,025	,714	,000	,000	,200	,000	,000	,752		,945	,000	,298	,115
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
Trees	Pearson Correlation	,170*	-,115	-,085	-,043	-,042	-,009	,014	-,024	-,070	,006	1	,042	,008	,043
	Sig. (2-tailed)	,042	,169	,310	,608	,617	,919	,867	,771	,400	,945		,612	,922	,605
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
SBAb	Pearson Correlation	,010	,094	,272**	-,252**	-,351**	,017	-,369**	-,482**	-,319**	,378**	,042	1	,112	,090
	Sig. (2-tailed)	,907	,262	,001	,002	,000	,841	,000	,000	,000	,000	,612		,181	,284
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
CycIn	Pearson Correlation	-,366**	,032	,165*	-,296**	-,360**	,043	-,121	-,170*	,361**	,087	,008	,112	1	,033
	Sig. (2-tailed)	,000	,702	,048	,000	,000	,605	,146	,041	,000	,298	,922	,181		,690
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
CycSp	Pearson Correlation	-,018	-,230**	-,006	-,194*	-,106	,219**	-,037	,024	,097	,132	,043	,090	,033	1
	Sig. (2-tailed)	,830	,005	,944	,019	,206	,008	,654	,776	,243	,115	,605	,284	,690	
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
Gen	Pearson Correlation	,030	-,055	-,150	,097	,219**	,087	,188*	,118	-,174*	-,074	,047	-,078	-,214**	,083
	Sig. (2-tailed)	,717	,510	,072	,245	,008	,296	,024	,158	,036	,379	,573	,351	,010	,320
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
Age	Pearson Correlation	-,117	-,088	-,059	-,061	-,021	-,197*	-,065	-,044	-,047	,106	,008	,095	-,077	,064
	Sig. (2-tailed)	,160	,294	,478	,463	,800	,018	,439	,597	,575	,203	,922	,257	,357	,444
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
Mon	Pearson Correlation	-,070	-,072	,033	-,058	-,056	,022	,063	-,018	,017	-,004	,046	,028	,034	,154

	Sig. (2-tailed)	,402	,393	,696	,485	,506	,792	,449	,828	,842	,964	,580	,735	,687	,065
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
Light	Pearson Correlation	-,001	-,044	,086	,031	,015	,008	-,096	-,030	-,104	,044	-,062	,136	,017	,172*
	Sig. (2-tailed)	,989	,596	,302	,709	,858	,925	,251	,722	,212	,602	,455	,104	,838	,039
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
Peak	Pearson Correlation	-,084	,011	,034	-,100	,052	-,040	,077	,096	,125	-,198*	-,141	,003	,142	,098
	Sig. (2-tailed)	,318	,895	,689	,232	,532	,635	,359	,250	,134	,017	,091	,972	,088	,242
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145
Wknd	Pearson Correlation	,079	-,067	-,082	,070	-,031	,047	,018	-,045	-,043	-,012	,059	-,107	,088	-,030
	Sig. (2-tailed)	,346	,426	,327	,405	,711	,573	,826	,594	,611	,883	,478	,199	,293	,717
	N	145	145	145	145	145	145	145	145	145	145	145	145	145	145

Correlations – continued

		Gen	Age	Mon	Light	Peak	Wknd
Dist	Pearson Correlation	,030	-,117	-,070	-,001	-,084	,079
	Sig. (2-tailed)	,717	,160	,402	,989	,318	,346
	N	145	145	145	145	145	145
PBuilt	Pearson Correlation	-,055	-,088	-,072	-,044	,011	-,067
	Sig. (2-tailed)	,510	,294	,393	,596	,895	,426
	N	145	145	145	145	145	145
PHigh	Pearson Correlation	-,150	-,059	,033	,086	,034	-,082
	Sig. (2-tailed)	,072	,478	,696	,302	,689	,327
	N	145	145	145	145	145	145
PSlow	Pearson Correlation	,097	-,061	-,058	,031	-,100	,070
	Sig. (2-tailed)	,245	,463	,485	,709	,232	,405
	N	145	145	145	145	145	145
PQuiet	Pearson Correlation	,219**	-,021	-,056	,015	,052	-,031
	Sig. (2-tailed)	,008	,800	,506	,858	,532	,711
	N	145	145	145	145	145	145
PSoft	Pearson Correlation	,087	-,197*	,022	,008	-,040	,047
	Sig. (2-tailed)	,296	,018	,792	,925	,635	,573
	N	145	145	145	145	145	145
PNoP	Pearson Correlation	,188*	-,065	,063	-,096	,077	,018
	Sig. (2-tailed)	,024	,439	,449	,251	,359	,826
	N	145	145	145	145	145	145
PSeg	Pearson Correlation	,118	-,044	-,018	-,030	,096	-,045
	Sig. (2-tailed)	,158	,597	,828	,722	,250	,594
	N	145	145	145	145	145	145
PHQS	Pearson Correlation	-,174*	-,047	,017	-,104	,125	-,043
	Sig. (2-tailed)	,036	,575	,842	,212	,134	,611
	N	145	145	145	145	145	145

ExAbs	Pearson Correlation	-,074	,106	-,004	,044	-,198*	-,012
	Sig. (2-tailed)	,379	,203	,964	,602	,017	,883
	N	145	145	145	145	145	145
Trees	Pearson Correlation	,047	,008	,046	-,062	-,141	,059
	Sig. (2-tailed)	,573	,922	,580	,455	,091	,478
	N	145	145	145	145	145	145
SBAbs	Pearson Correlation	-,078	,095	,028	,136	,003	-,107
	Sig. (2-tailed)	,351	,257	,735	,104	,972	,199
	N	145	145	145	145	145	145
CycInt	Pearson Correlation	-,214**	-,077	,034	,017	,142	,088
	Sig. (2-tailed)	,010	,357	,687	,838	,088	,293
	N	145	145	145	145	145	145
CycSp	Pearson Correlation	,083	,064	,154	,172*	,098	-,030
	Sig. (2-tailed)	,320	,444	,065	,039	,242	,717
	N	145	145	145	145	145	145
Gen	Pearson Correlation	1	,059	,074	,106	-,050	,051
	Sig. (2-tailed)		,484	,374	,204	,550	,540
	N	145	145	145	145	145	145
Age	Pearson Correlation	,059	1	,081	,061	,042	-,213**
	Sig. (2-tailed)	,484		,334	,470	,620	,010
	N	145	145	145	145	145	145
Mon	Pearson Correlation	,074	,081	1	-,190*	-,014	-,101
	Sig. (2-tailed)	,374	,334		,022	,869	,227
	N	145	145	145	145	145	145
Light	Pearson Correlation	,106	,061	-,190*	1	-,003	-,025
	Sig. (2-tailed)	,204	,470	,022		,970	,766
	N	145	145	145	145	145	145
Peak	Pearson Correlation	-,050	,042	-,014	-,003	1	-,238**
	Sig. (2-tailed)	,550	,620	,869	,970		,004
	N	145	145	145	145	145	145
Wknd	Pearson Correlation	,051	-,213**	-,101	-,025	-,238**	1
	Sig. (2-tailed)	,540	,010	,227	,766	,004	
	N	145	145	145	145	145	145

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Regression output full linear model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,561[a]	,315	,210	,27484

a. Predictors: (Constant), Wknd, ExAbs, Light, Trees, PSoft, CycInt, Mon, Gen, PBuilt, Age, Peak, PHQS, CycSp, PSlowT, PHighU, SBAbs, PNoPa, PQuiet, PSeg

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4,332	19	,228	3,018	,000[b]
	Residual	9,442	125	,076		
	Total	13,775	144			

a. Dependent Variable: Dist

b. Predictors: (Constant), Wknd, ExAbs, Light, Trees, PSoft, CycInt, Mon, Gen, PBuilt, Age, Peak, PHQS, CycSp, PSlowT, PHighU, SBAbs, PNoPa, PQuiet, PSeg

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	,274	,106		2,597	,011
	PBuilt	-,001	,003	-,027	-,292	,771
	PHighU	-,004	,003	-,114	-1,245	,216
	PSlowT	,004	,004	,086	,905	,367
	PQuiet	,006	,003	,324	1,789	,076
	PSoft	,003	,002	,142	1,696	,092
	PNoPa	-,003	,002	-,133	-1,206	,230
	PSeg	-,001	,003	-,058	-,300	,765
	PHQS	,000	,002	-,018	-,156	,877
	ExAbs	,002	,001	,218	2,049	,043
	Trees	,024	,010	,186	2,416	,017
	SBAbs	,002	,004	,052	,535	,594
	CycInt	-,005	,002	-,303	-3,146	,002
	CycSp	-,007	,024	-,026	-,299	,766
	Gen	-,059	,050	-,096	-1,172	,244
	Age	-,084	,051	-,130	-1,630	,106
	Mon	-,005	,049	-,008	-,100	,920
	Light	,006	,081	,006	,073	,942
	Peak	,043	,051	,070	,848	,398
	Wknd	,079	,080	,081	,995	,322

a. Dependent Variable: Dist

Regression output adapted linear model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,520[a]	,271	,239	,26978

a. Predictors: (Constant), BirthYear01, Trees_Per100M, P_QuietRoads, P_SoftRoadSide, Cyclist_Intensity, Exits_Abs

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3,731	6	,622	8,543	,000[b]
	Residual	10,044	138	,073		
	Total	13,775	144			

a. Dependent Variable: Distance

b. Predictors: (Constant), BirthYear01, Trees_Per100M, P_QuietRoads, P_SoftRoadSide, Cyclist_Intensity, Exits_Abs

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,344	,040		8,589	,000
	P_QuietRoads	,004	,002	,238	2,679	,008
	P_SoftRoadSide	,003	,002	,142	1,889	,061
	Exits_Abs	,002	,001	,265	3,195	,002
	Trees_Per100M	,024	,009	,183	2,514	,013
	Cyclist_Intensity	-,006	,001	-,322	-4,085	,000
	BirthYear01	-,090	,048	-,139	-1,860	,065

a. Dependent Variable: Distance

Regression output full Tobit model

```
--------------------------------------------------------------------------
--
Limited Dependent Variable Model - CENSORED
Dependent variable                 DIST
Log likelihood function       -23.01553
Estimation based on N =     145, K =  21
Inf.Cr.AIC  =      88.0 AIC/N =      .607
Model estimated: May 26, 2020, 13:44:05
Threshold values for the model:
Lower =       .0000    Upper = +infinity
ANOVA  based fit measure =       .179622
DECOMP based fit measure =       .226757
--------+-----------------------------------------------------------------
--
```

DIST	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Primary Index Equation for Model					
Constant	.26797***	.10389	2.58	.0099	.06435	.47160
PBUILT	-.00082	.00266	-.31	.7593	-.00603	.00440
PHIGHU	-.00344	.00280	-1.23	.2195	-.00894	.00205
PSLOWT	.00427	.00429	1.00	.3197	-.00414	.01267
PQUIET	.00557*	.00311	1.79	.0735	-.00053	.01166
PSOFT	.00370*	.00201	1.84	.0652	-.00023	.00763
PNOPA	-.00299	.00226	-1.32	.1862	-.00742	.00144
PSEG	-.00064	.00279	-.23	.8195	-.00610	.00483
PHQS	-.00036	.00216	-.17	.8661	-.00461	.00388
EXABS	.00171**	.00076	2.23	.0258	.00021	.00320
TREES	.02493**	.00969	2.57	.0101	.00593	.04393
SBABS	.00211	.00434	.49	.6275	-.00641	.01062
CYCINT	-.00532***	.00163	-3.26	.0011	-.00852	-.00212
CYCSP	-.00588	.02301	-.26	.7982	-.05097	.03921
GEN	-.06443	.04942	-1.30	.1923	-.16130	.03243
AGE	-.09346*	.05028	-1.86	.0630	-.19200	.00508
MON	-.01599	.04796	-.33	.7388	-.11000	.07801
LIGHT	.01390	.07934	.18	.8610	-.14162	.16941
PEAK	.05519	.05047	1.09	.2742	-.04374	.15412
WKND	.08101	.07860	1.03	.3027	-.07304	.23507
	Disturbance standard deviation					
Sigma	.26800***	.01644	16.30	.0000	.23578	.30022

```
--------+-----------------------------------------------------------------
--
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------
--
```

Regression output adapted Tobit model

```
-----------------------------------------------------------------------------
--
Limited Dependent Variable Model - CENSORED
Dependent variable                   DIST
Log likelihood function        -27.54569
Estimation based on N =     145, K =   8
Inf.Cr.AIC  =     71.1 AIC/N =      .490
Model estimated: May 26, 2020, 13:47:39
Threshold values for the model:
Lower =       .0000    Upper = +infinity
ANOVA  based fit measure =       .152455
DECOMP based fit measure =       .191762
--------+--------------------------------------------------------------------
--
        |                 Standard            Prob.      95% Confidence
    DIST|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
--
        |Primary Index Equation for Model
Constant|     .34007***      .04112     8.27  .0000      .25947    .42067
  PQUIET|     .00427***      .00160     2.67  .0076      .00113    .00740
   PSOFT|     .00361*        .00189     1.91  .0567    -.00010    .00732
   EXABS|     .00203***      .00063     3.24  .0012      .00080    .00325
   TREES|     .02393**       .00963     2.48  .0130      .00506    .04281
  CYCINT|    -.00556***      .00140    -3.96  .0001    -.00831   -.00280
     AGE|    -.10000**       .04963    -2.01  .0439    -.19728   -.00272
        |Disturbance standard deviation
   Sigma|     .27652***      .01697    16.30  .0000      .24327    .30978
--------+--------------------------------------------------------------------
--
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
--
```