Analysing and modelling of influencing attributes on transfer walking times for metro transfers

Berend van Voorst tot Voorst

Analysing and modelling of influencing attributes on transfer walking times for metro transfers

by

Berend van Voorst tot Voorst

to obtain the degree of Master of Science

Civil Engineering, Track: Transport & Planning

at the Delft University of Technology,

to be defended publicly on Wednesday 31th of May, 2023 at 14:00.

Student number:	4466039	
Project duration:	1 April, 2022 - 31 May, 2023	
Thesis committee:	Dr. ir. W. Daamen,	TU Delft, chair
	Dr. ir. Y. Yuan,	TU Delft, daily supervisor
	Dr. ir. N. van Oort,	TU Delft
	Ir. L. Pardini Susacasa,	Royal HaskoningDHV
	Ir. B. Donners,	Royal HaskoningDHV

This thesis is confidential and cannot be made public until 31th of May, 2023.

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Cover image is obtained from RET, photo by Rick Keus.

TU Delft

Royal HaskoningDHV
Enhancing Society Together

Preface

This thesis marks the end of my student career at the TU Delft, especially my master's in Transport and Planning. In these seven years, I have learned a lot personally and scientifically. Furthermore, I kept involved in improving the education system, with two years of experience in the Faculty Student Council and three years as a student assistant, which helped me further develop. Funny side note, the covid-19 pandemic allowed me to complete my master's without doing any examination at the TU campus, but all from my student room. Since I was small, I had an interest in transport, which grew during my masters when being taught about the extensive world of public transport. Therefore, there was no doubt I would perform a public transport-related thesis.

Firstly, I thank Royal HaskoningDHV for providing the opportunity for this research. Most of all, I would like to thank my three supervisors. Firstly, Barth Donners for setting up the thesis proposal, for his constructive feedback, and for keeping me focused on the thesis topic. Also, a big thanks to Laura Pardini Susacasa for providing guidance through my research. And lastly, Nigel Birch, for his ideas, advice and feedback during my thesis process. For all the meetings I had with all of your three, I appreciate your eagerness to support me and improve my thesis.

Secondly, I would like to thank my TU Delft supervisors for their assistance. Foremost, I would like to thank my daily supervisor Yufei Yuan for always sparking me with new insights to continue with my research and commenting on the report sections. Our meetings always gave me new motivation or findings to continue or a way out when I was stuck on a certain part. Special thanks go to Niels van Oort for being my external supervisor. Your open view and constructive feedback really helped me to remember the general application of my research to the public transport aspect. Finally, I would like to thank my chair of the committee, Winnie Daamen. Although we only met a few times, your superior knowledge of walking behaviour gave me the essential tools to set up this research.

Additionally, I would like to thank the RET, the Rotterdam metro operator, for allowing me to collect the walking times at Beurs metro station. The collaboration was pleasant between us, and I am happy to hear that my data support the future design options of the station. The very friendly station managers gave valuable information about the daily operations and hassles at the station.

Finally, a special thanks to my family, friends and housemates who supported me during my thesis. I did enjoy the afternoon walks with my housemates to relieve the thesis frustrations and the occasional dinners and gatherings with friends to provide the essential distraction to the thesis life. Moreover, all holidays with family or friends helped me reset and continue with new energy on the research.

Berend van Voorst tot Voorst
Delft, May 2023

Executive summary

Ridership increases in current public transport systems due to the significant growth of cities, leading to crowding and unreliability issues. In metro networks, the goal is to minimise journey times and provide quick, safe and easy transfers between metro lines. These transfers rely on more than the waiting time for the next metro but also on the required walking time for passengers to walk from one to the other platform. Understanding the transfer walking times in current (metro) stations helps to design better transfer stations in the future or improve existing ones. Therefore, the primary focus of this study is the transfer walking time of passengers.

The transfer walking time depends on more than the walking distance between the platforms and the station layout. One of the important factors in the transfer walking time is the variance in passengers and, thus, in walking behaviour and, eventually, in the magnitude of the transfer walking time. Besides the variance in walking behaviour, capacity bottlenecks in a metro station could lead to crowding or queues and, thus, require more time for passengers to pass that element. Lastly, passengers choose their preferred mode to move a vertical level in a metro station. Models already exist to predict transfer walking time variation regarding station layout or the crowding effect but lack the inclusion of passenger characteristics. Currently, limited research is done to include the effect of passenger variation or choices on the transfer walking time, but the topic has been recommended as a valuable addition.

This research aims to understand how and to what extent specific attributes could influence transfer walking times for metro-to-metro transfers. The attributes are mainly related to passenger characteristics and station layout elements. The research objective is to highlight the most important attributes and investigate how these attributes could be used to improve transfer walking time estimations or models. To fulfil the research objective, the following main research question should be answered: ***What are the significant passenger characteristics and station attributes that affect the transfer walking time in metro transfers and to which extent can this information be used to improve current estimations of transfer walking times?***

The research scope limits the transfer layout requirements as shown in Figure 1, with a corridor and a vertical transport point (VTP). The corridor length or the sum of all corridor lengths has a minimum because it expects a large variation in walking time and discards cross-platform transfers. The VTP should have at least two vertical transport modes to investigate the effect of different travel speeds per mode, and most previous research only included one mode. To answer the main research question, different methods have been applied. Firstly, a literature review on all station layouts and passenger attributes which might influence the transfer walking time and the current models to estimate the transfer walking time. Moreover, this study collected data on the transfer walking time at Beurs metro station in Rotterdam. The walking times were analysed and provided information about the most significant factors. In the last step, the transfer walking time has been modelled with the included significant attributes along one of the presented ways of the literature. All in all, these steps provided conclusions to the research objective.

The literature review provided insights into the primary factors from the passengers and the station layout on the walking time. The passenger walking speed depends on gender, group size, trip purpose and crowding in a segment. Furthermore, passengers prefer the shortest route in a station and the escalator over the stairs as a vertical transport mode in stations. Recent studies suggest investigating the moment of making because passenger behaviour or trip purpose could differ between weekdays, weekends or holidays. Moreover, the transfer walking time is already different for different time-frames for train-metro transfers, which might also be the case for metro transfers. Moreover, in a VTP, only stairs and escalators are considered. The level of crowding or queue before the escalator affects the mode choice between the escalator and stairs. Probably due to the low assumed usage, the lift is excluded as a vertical transport mode in previous studies. However, certain passengers, the disabled, those with walking issues or those with large luggage, rely on the lift and could be excluded when

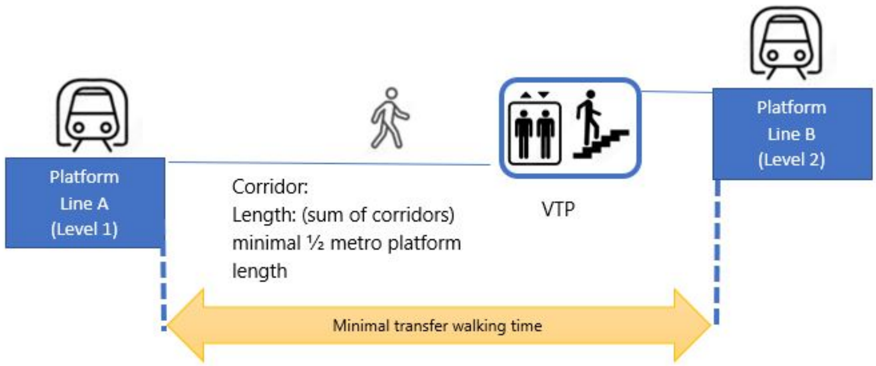


Figure 1: Minimal layout requirement for transfer path (TP) in this research.

estimating the transfer walking time. Therefore, the lift is included as a vertical transport mode in this study.

Current models on the transfer walking time use the transfer path distance only or include a lognormal walking time distribution to estimate the walking time of passengers. A different model approach uses data from Automatic Fare Collection and Automatic Vehicle Location and assumptions on the walking speed to estimate the variance in transfer walking times. Most studies on the metro transfer walking time modelling recommend including passenger attributes to better understand the desired transfer time, primarily to ensure feasible transfers for all passengers at the end of the day. When including different passenger or layout attributes on the walking speed or time, a suitable model approach is with multiple linear regression.

Four data collection methods are available or used to collect passenger movements in stations or walking times: video footage, Bluetooth, a combination of AFC/AVL data and covert observations. Regarding the research objective and possible variables influencing the transfer walking time, covert observations are the preferred method for collecting the walking time. Covert observation collects the data without informing the participants.

The data collection was performed at metro station Beurs, Rotterdam, the Netherlands, with the knowledge of possible influencing attributes. A transfer path, and thus the transfer walking time, starts when a passenger exits the metro and stops when the passenger steps onto the desired platform. The station has five transfer paths, and each is split into segments. In each segment, the observer randomly picks a passenger at the segment starting point and collects the passengers walking time and the related variables as identified from the literature review. The categories of the variables are given in Table 1. Four categories are collected in all segments, while two are specific to the VTP segment, and only one is at the platform location. The variables are based on the literature for modelling the transfer walking time or the passenger walking speed. The collection focused on passengers who were mainly familiar with the station by collecting only samples which walked most directly through a segment without hesitation. At various moments during the day and week, the walking times were collected at Beurs station in order to consider the demand variety. The collection's reflection showed that considerable congestion or crowding was absent in the station.

Table 1: Used variables per category in the data collection.

Gender	**Luggage**	**Group size**	**Crowding**	**VTP mode**	**VTP waiting**	**Alighting location**
Male Female	No/small Large	Alone In group	Free Crowded	Stairs Escalator Lift	Direct boarding Wait to board	Front Middle Back
All segments				VTP segments		Platform segment

The data analysis gave the following results on the collected walking times. Too few samples with large luggage were collected for some segments because these samples were primarily absent during the whole collection period. Therefore, the conclusions were treated as less convincing for those segments in the luggage category. For samples walking in groups, these were primarily duos, and the large luggage items were mainly strollers and bicycles. In most segments, the walking time had a lognormal distribution, while some segments had a normal distribution of the walking time. The reason for the difference remains unclear why some had normal walking time distribution. The assumption is that all segments have a lognormal walking time in order to perform the same statistical test for each segment and category.

Through statistical tests, the categories of group size, vertical transport mode, waiting for VTP and the alighting location from Table 1 had significantly different walking times. Passengers in a group required 20-30 % more walking time than passengers walking alone. The difference between stairs and escalators was insignificant. In contrast, according to the median walking time per mode, lift users required around 45-60 seconds longer per VTP. At the platform segment, the medians of each alighting location add or subtract 20 seconds from the walking time to leave the platform compared to the other alighting areas. Furthermore, there is an insignificant difference between peak and off-peak walking time in Beurs station.

The walking times were normalised regarding the segment length into passing speeds to compare with literature findings and to check if specific station attributes have an additional effect. The passing speed is always lower or close to the walking speed due to possible moments a sample had to slow down in the collected walking times. Station elements such as the presence of fare gates or the direction of vertical movement impact the passing speed beside the difference in walking time for group size, vertical transport mode, condition to board a vertical transport mode or the alighting location. The last category does have a significant accuracy error because the exact distance between the exit and the alighting location per sample was not recorded. Therefore, the distance was averaged per alighting location, resulting in a significant considerable variation in the passing speed for the alighting location closest to the exit. Therefore, only the passing speeds at the furthest alighting parts were assumed to be the most representative of the passing speed analysis.

The passing speeds were analysed along three segment types: corridor, VTP and the platform. The overall median passing speed ranged between 1.12-1.37 m/s in the corridors. The passing speed in VTP segments is around 0.90 m/s for the stair modes, but these segments include a corridor part. If a queue is present and the passenger has to wait to board a stair mode, the passing speed is around a third lower. The passing speed is approximately 0.20 m/s for passengers using the lift. The passing speed at the platform varies between 1.10 and 1.20 m/s, which is lower than the passing speed in the corridors. The recommendation is to use different passing speeds for each segment type, when modelling a metro station.

A Monte Carlo simulation translated the walking time per segment into estimating the total transfer walking time per significant variable for the data analysis. The underlying assumption was that each variable walking time was lognormal distributed. Compared to the overall estimated transfer walking times, walking in a group increases the time by around 15-40%, in line with the data analysis results at a segment level. The differences are more considerable when comparing each furthest and closest part of the platform to the exit. The walking time is between 20% and 60% longer for passengers who alight far away from the exit than those who alight close to the exit. Passengers who must wait to board a vertical transport mode require, on average, around 30 % to 60 % more time to transfer than those who can board directly. The difference depends on the number of VTPs to pass in the transfer path. The difference becomes larger when more VTPs are present in a transfer path. The most significant category is the vertical transport mode. Between a stair mode and a lift mode, the transfer walking time is at least 55% longer when taking the lift. With two lifts in the transfer path, the transfer walking time becomes at least double as long as compared to passengers who only take stairs or escalators.

Based on these findings, it is suggested to determine the transfer walking time for the following categories besides the general transfer walking time: passengers depending on the lift, walking in a group and whether waiting to board a vertical transport is present in the busier periods. Especially, the passengers relying on the lift might take a metro later after the transfer compared to the other passengers as their transfer walking time could be double as long.

The final step is to model the walking time with related attributes from the data collection and the discovered ones from the passing speed analysis. The combination of the data collection setup and the discovered multiple linear model approach in the literature review to estimate the effect of a particular combination of attributes on a walking speed led to the decision to use the multiple linear regression model approach to predict the transfer walking time. As previous literature divided the segments into three types, and these types were used to analyse the passing speed, the split will also be present in the modelling. The walking time is estimated with a multiple linear regression model for three segment types of a transfer: corridors, VTP and the platform segments. The included attributes are from the data collection and station layout elements to see if more or different ones are significant compared to those found in the data analysis.

Two models are made per segment type, one as a walking time model and the other as a passing speed model. The walking time model is part of the main research objective and can be calibrated directly with the collected walking times per segment. The passing speed model is the generalised version and has the favour in practical applications to model passenger behaviour in stations. The model approach was to generate first the walking time model per segment type with the data collection attributes. The second step involved adding mainly layout-related attributes to the model. The last step is backward elimination to remove insignificant or correlated variables from the model. The remaining walking time model should have only significant variables. The passing speed model was fitted with the same significant attributes as the walking time model. If the passing speed model had insignificant attributes, these were removed with backward elimination to end up with a model of only significant attributes.

Before the modelling, the segments were divided into model calibration and validation segments. The calibration segments were used to determine the attribute magnitude, and the validation segments were used to test the predictive power of the models. The validation was done at a qualitative and quantitative level. The qualitative method checks a segment's estimated walking time distribution with the observed walking time distribution. The quantitative method involved calculating the error between the observed and estimated walking time for the combination of attributes as observed per sample. The validation process concludes whether the walking time and passing speed models were usable. For the passing speed model validation, the estimated passing speeds are translated to walking time using the length of the validation segment. Both models could be used for the corridor and platform segment type, with a slightly smaller error in the walking time model. For the VTP segment, only the walking time model is recommended because of the smaller difference between the observed and estimated walking time compared to the walking time estimated through the passing speed model.

The walking time models predict a mean, lower and upper bound per combination of attributes by introducing variable X_O, which indicates whether the lower, mean or upper walking time will be calculated. In the passing speed models, the variable only represents the mean or upper bound of the passing speed because otherwise, the variable turns all remaining attributes insignificant when considering a lower bound.

The resulting walking time model (WT_c) for corridor segments is given in Equation 1, with variables as the length L, group size X_{GR}, luggage X_{LU}, presence of fare gates X_{FG}, whether the segment is the last one before the platform X_{LS}, and including the outlier variable X_O to represent an upper or lower bound of the walking time per combination of attributes. Besides the length, the passing speed model has the same attributes included.

$$WT_C(s) = -2.119 + 0.771L + 0.61X_{GR} + 1.981X_{LU} + 3.673X_{FG} + 2.712X_{LS} + 6.119X_O \quad (1)$$

For the vertical transport segments, the walking time model (WT_V) in Equation 2 is split in horizontal L_H and vertical transport length L_V. Furthermore, vertical transport mode X_{VT} only differs between stairs modes (escalator and stairs) and the lift. Moreover, the waiting condition X_{WA} to board and the location of the horizontal part X_P play a role. The passing speed has significantly fewer attributes, only group size, vertical transport mode and waiting condition. The reason could be the lower variance in passing speed compared to the walking time and the influence on the significant attributes.

$$WT_V(s) = 0.481L_h + 1.815L_v + 2.459X_{GR} + 43.711X_{VT} + 4.633X_{WA} + 4.563X_{LU} + 9.944X_O + 2.409X_P \quad (2)$$

The last walking time model for the platform (WT_P) is given in Equation 3, with only the new variable X_A describing the relative distance to the exit and the group size as relevant attributes. The length uses the average walking distance for each 1/3 of the platform regarding the exit location and platform part. In the passing speed model, the same attributes are present except for the length.

$$WT_P(s) = 0.763L + 2.58X_{GR} - 8.21X_A + 15.075X_O \tag{3}$$

All variable descriptions and states are given in Table 2. A final remark is that all models have been made and validated for short segments up to 50 meters or platforms of 90 meters. Furthermore, the models have not been tested to predict the total transfer walking time because the exact transfer walking times were not collected due to the collection set-up. Lastly, the walking time model estimates the effect of attributes independent of the length. Therefore, in longer segments, the difference between different attributes could become smaller and thus insignificant.

Table 2: Variables explanation in the walking time models for the segment types.

Variable	Description	State
L	Length of the corridor or platform alight walking distance (m)	
L_h	Horizontal length of flat part in vertical transport segment (m)	
L_v	Horizontal length of vertical transport mode (m)	
X_A	Platform part	0 Middle 1/3 to exit, -1 closest 1/3 to exit, 1 Furthest 1/3 to exit
X_{FG}	Fare gates	0 None, 1 Present
X_{GR}	Group size	0 Walks alone, 1 In a group
X_{LU}	Luggage size	0 No/small, 1 Large
X_{LS}	Last segment	0 None, 1 Before platform
X_O	Walking outlier	0 Average, -1 Quickest 20 % (upper bound) , 1 Slowest 20 % (lower bound)
X_P	Location L_h	0 Unclear, -1 before VTP, 1 after VTP
X_{VT}	VTP mode	0 Stairs/escalator, 1 Lift
X_{WA}	Waiting for VTP	0 Direct boarding, 1 Wait to board
X_{WI}	Segment width	0 $<$ 3.5 m, 1 $\geq$ 3.5 m

The answer to the main research question is as follows; three attributes influence the transfer walking time the most: the group size, vertical transport mode choice, the waiting condition to board the mode, especially between the lift and stair modes, and the passengers' alighting position at the platform. Crowding had an insignificant effect on the transfer walking time from the observations and within the models. The method to model these significant attributes on the transfer walking time is splitting the transfer path into three segment types: corridors, VTP and platform segments. After that, the walking time can be estimated with the generated walking time or passing speed models per segment. The models predict a mean and lower bound walking time per combination of attributes. Additionally, the walking time model estimates an upper bound of the walking time.

The most considerable contribution is the insight into metro transfer behaviour based on different passenger characteristics, choices or layout elements. The main results are the significant categories influencing the transfer walking time. The effect of group size, alighting location and the vertical mode choice are the most important, not only on a transfer segment level but also on the complete metro transfer. Furthermore, this study included the elevator as a vertical transport mode in stations, besides stairs and escalators. Lastly, capturing all significant attributes in walking time and passing speed models to estimate the walking time per combination of attributes.

However, the largest limitation of this study and given conclusions is that the data collection was performed at only one metro station. Furthermore, the level of crowding was insignificant according to the observed and modelled walking times. Therefore, these results are assumed to be invalid for metro stations with extensive crowding because the walking time can increase through crowding. Moreover, the generated models have not been used yet to predict a total transfer walking time.

The findings of this study have the following practical suggestions for metro schedulers or public transport planners. Firstly, the transfer walking time should be differentiated for different passenger groups. At least the difference between passengers walking alone or in a group and between escalator/stair users and lift users. The difference between the groups on the transfer walking time can be derived using the generated walking time or passing speed models. These estimated walking times are used as input for the metro schedule to check if all transfers are feasible for all passengers. An application for passengers is to provide an improved route planner when using the metro. The passenger can fill in their characteristics or preferred vertical transport mode. The planner will subsequently present a realistic transfer walking time and check which next metro the passenger should take. Eventually, the planner provides a representative travel time when using the metro for all passenger categories.

The recommendations for further studies are split into collecting additional variables and modelling the transfer walking time for future research on the transfer walking time.

- In future data collection on transfer walking times: Trip purpose, the difference between passengers familiar and unfamiliar in the metro station, passenger age, departure information for the next metro, and disabled passengers are proposed variables to check if those also significantly impact the transfer walking time.
- The Monte Carlo simulation assumed all the walking times per segment of having a lognormal distribution. While some segments had normal or uniform walking time distributions. Therefore, future studies could do the Monte Carlo simulation with the actual walking time distribution type and check whether the resulting transfer walking times are significantly different compared to assuming only lognormal walking time distribution per segment.
- A total transfer walking time has not been estimated with the combination of all three models. Therefore, the recommendation is to estimate a complete transfer walking time with the models.
- Calibrate the walking time and the passing speed models with more different vertical transport elements, such as the travel speed or capacity. Furthermore, situations of extensive crowding, which increase walking time, could be added to improve the models and to be more representable. These calibration steps could be done with walking time data from different metro transfer stations. The next step could be to validate the models for other rail or public transport transfers.
- Use a non-linear approach to model the walking time in future studies. A non-linear approach could be used to generate a walking time model that includes the effect of the significant attributes depending on the length.

Contents

List of Figures

List of Tables

List of Abbreviations

AFC Automatic Fare Collection

AVL Automated Vehicle Location

EB Eastbound

LOS Level of Service

MAE Mean absolute error

NB Northbound

PS Passing speed

RMSE Root mean squared error

SB Southbound

TP Transfer path

TWT Transfer walking time

VT Vertical transport

VTP Vertical transport point

WB Westbound

WT Walking time

List of Symbols

List of Greek symbols

α Level of significance

β Constant to fit in models per variable

γ_1 Skewness

μ Population mean

σ Population standard deviation

List of Latin symbols

d Desired accuracy

H_0 Null hypothesis

H_1 Alternative hypothesis

L Segment horizontal length for corridors and platforms

L_h Corridor length part in vertical transport segment

L_v Horizontal vertical transport mode length in vertical transport segment

N Population size

n Calculated sample size

n' Non-adjusted sample size

n_c Calculated categorical sample size

q_n Percentile

p Probability

ps Passing speed

ps_c Estimated passing speed in corridor segments

ps_p Estimated passing speed at platform segments

ps_v Estimated passing speed in vertical transport segments

R^2 Coefficient of determination

s Sample standard deviation

$s.e$	Standard error
WT_c	Estimated walking time in corridors segments
WT_o	Observed walking times in data collection
WT_p	Estimated walking time at platforms
WT_v	Estimated walking time in vertical transport segments
X_A	Alighting location of sample at platform segment
X_B	Dummy: Vertical transport develops queues.
X_{CO}	Completness of vertical transport presence in segment
X_E	Dummy: Number of exits at the platform
X_{FG}	Dummy: Fare gates presence
X_G	Dummy: Gender of sample
X_{GR}	Dummy: Group size of sample
X_{LU}	Dummy: Presence of large luggage
X_{LS}	Dummy: If the segment is the last one before the platform or not
X_O	State if sample is outlier
X_P	location horizontal part in vertical transport segment
X_S	Direction of vertical movement of vertical transport segment
X_{SS}	Dummy: if vertical transport segment consists of only stair modes
X_{VT}	Dummy: Vertical transport mode
X_{WA}	Dummy: Waiting condition vertical transport
X_{WI}	Dummy: Segment width
$\bar{x}$	Sample mean
Z_a	Z-score

1 Introduction

More and more people will live in cities worldwide in the future, according to International Organization for Migration (2015). All these new inhabitants want to have reliable mobility inside the town. European cities should provide more sustainable mobility by providing more train and public transport services (European Commission, 2020). The demand for these public transport systems will increase to levels where crowding inside stations and vehicles could lead to more unreliable and longer journeys. Expanding public transport infrastructure, especially for rail modes, is expensive. Optimising the current use of stations is a more intelligent alternative in the short and long run.

The metro is one of the highest rail capacity modes (UITP, 2019), whereas in European cities, already large metro networks exist. Passengers avoid making transfers where possible (Cascajo et al., 2017) because transfers are penalised for being inconvenient. However, in more extensive metro networks, a transfer sometimes is required to reach your destination.

When modelling metro or public transport networks, minimising transfer times is an often used objective in these models for determining timetables in the tactical planning stage of a network (Liu et al., 2021). These timetables help reduce transfer times for passengers. However, these times should be at least long enough to walk from one platform to the other to complete the transfer. Due to crowding in the station or the complex transfer route layout, the transfer time could vary inside a station. New metro transfer station designs include optimal transfer experiences for passengers due to building cross-platform layouts or platforms directly above each other, minimizing the walking time and, thus, the complexity of the transfer. Nevertheless, optimal designs cannot always be built in all stations through different motives. The result is that passengers are still required to walk between platforms. Long walking distances in the station negatively impact the passenger experience of a metro transfer (Lin et al., 2022). Understanding transfer walking times in current stations help to design better transfer stations in the future or improve existing ones.

However, the transfer walk depends not only on the route layout or the available time. One of the important factors is the variance in passengers and, thus, in walking behaviour. Not only directly from the passengers themselves but also due to the layout of the station elements. Capacity bottlenecks in a station could lead to queues and, thus, require more time for passengers to pass that element. Furthermore, some passengers have trouble leaving the metro because of other passengers blocking the exit who wish to continue with the metro. These are examples of delays out of the influence of the passenger but will increase the required time to reach the desired metro platform. The passengers themselves play a role because their walking speed is largely affected by different aspects, such as gender, age, trip purpose, the physical environment and more (Bosina and Weidmann, 2017). The walking speed and walking time are reciprocally related. A lower walking speed results in a longer walking time. Lastly, passenger choices also play a role in their transfer, especially when vertically moving between levels in a station. In metro stations, escalators are the preferred mode to move up or down (Lin et al., 2022). However, some passengers might use the stairs or rely on the lift. Each mode might have a different travel time and thus influence the transfer time of a passenger. Therefore more research is required on the influence of certain passenger and station layout variables on the transfer walking time. In collaboration with Royal HaskoningDHV (RHDHV), an engineering and consulting firm, a research master thesis has been written. RHDHV provides consultations and solutions for their rail clients. This research contributes to a better understanding of stations as hubs in larger rail networks, resulting in better consulting with their customers.

The remainder of this introduction chapter discusses the problem statement, with the obtained research objective and the following research questions in section 1.1. The set research scope and case are presented in section 1.2 and the expected relevance of this research in section 1.3. Finally, section 1.4 present the reports' outline.

1.1. Problem statement

Methods already exist to predict the walking time of passengers in a metro station making a transfer, called transfer walking times. While previous research on the transfer walking time can estimate variation in the transfer walking time, most only included station layout elements. These elements are used only to predict the level of crowding and the possible increase in walking time. Furthermore, current models use Automatic Fare Collection (AFC) and Automatic Vehicle Location (AVL) to predict the transfer walking time with assumptions on the walking distribution. The models also lack validation from observed walking time in the metro stations. Limited research is done to include the effect of passenger variation or choices on the transfer walking time. In contrast, most studies on transfer walking time recommend including the effect of passenger attributes on the walking time distribution. The importance of including passenger behaviour or choices besides station layout elements is given in chapter 2.

1.1.1. Research objective

This research aims to understand how specific attributes might influence the transfer walking times for metro-to-metro transfers. The motivation for including only metro-to-metro transfers is the low headway of this rail mode, thus having minimal waiting times. The transfer walking time is assumed to be nominative for the total transfer time for passengers. Furthermore, the transfer walk in the metro stations is more complex (Du et al., 2009) than for other mode transfers. The transfer attributes are mainly related to passenger characteristics and station layout elements. This research aims to highlight the most important ones and investigate how these attributes could be used to improve transfer walking time estimations or models.

1.1.2. Research questions

The research objective leads to the following research question:

What are the significant passenger characteristics and station attributes that affect the transfer walking time in metro transfers and to which extent can this information be used to improve current estimations of transfer walking times?

Sub-questions

Firstly, possible elements influencing the transfer walking time estimations should be understood. Not only from the passenger and station layout but also possibly different ones. Furthermore, present methods for obtaining the transfer walking time should be analysed, which could include the influencing variables. The following sub-questions are framed:

1. What passenger's characteristics or decisions, station layout or other elements play a role in the transfer walking time?
2. How are metro transfer walking times currently estimated or collected?

The transfer walking time might depend on a large number of variables. However, the key is to determine the most important ones. Therefore, the next sub-question is derived:

3. Which attributes have the most significant effect on the transfer walking time?

With the most important attributes found, improved estimations can be made to represent the transfer walking time for the different significant aspects. The estimation helps to predict transfer walking times for other metro transfers. Resulting in the last sub-question:

4. How can the information of the significant attributes be used to estimate transfer walking times better?

1.2. Research scope

This section discusses the research scope, including the variables of transfer walking time used in this research and the limitations of the transfer station layout. Based on the problem statement in the previous paragraph, there is a gap in gathering more information on passenger types and station layout characteristics on the metro transfer walking time. Furthermore, this research focuses on only metro-to-metro transfers.

1.2.1. Scope on metro transfer station layout

Metro transfer stations exist in different layouts, from cross-platform transfers to large stations with platforms at various spatial and vertical locations. This research focuses on a specific transfer station layout, where passengers must walk a significant distance to make a transfer. The term transfer path describes the transfer route in a station in this research. This research's minimal transfer path (TP) requirements are given and motivated below, including a visual representation in Figure 1.1.

- Each transfer path should at least have one corridor to cross. In that way, a large variation in walking time is expected, and cross-platform transfers are discarded. The minimal length of this corridor is a half platform length or the sum of multiple corridors. A metro platform (in Europe) is around 100-120 m long (Guerrieri, 2023). Thus the corridor length should be at least 50-60 m.

- The transfer path should at least have 1 vertical transport point (VTP) because then the effect of vertical movement can be observed in the transfer walking time.

- The vertical transport point should have at least 2 different vertical transport modes because each mode might have a different travel time to move a vertical level. Previous studies on transfer walking time only include one vertical transport mode (Du et al., 2009; Zhou et al., 2016).

- The vertical transport point should at least clear a floor level in height. The assumption is around 3-4 meters.

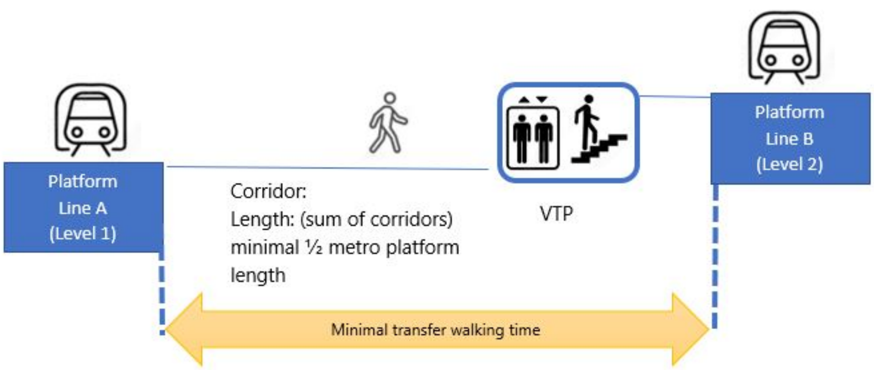


Figure 1.1: Minimal layout requirement for transfer path (TP) in this research.

1.2.2. Case study station

The station choice in this research is metro station Beurs in Rotterdam. It is the only metro station in the Netherlands that fulfils the requirements from Figure 1.1 and has multiple transfer layouts. Therefore, this station helps extend the knowledge of the effect of transfer layout.

1.3. Relevance

The objective is to determine the effect of passenger and station characteristics on the transfer walking time for metro-metro transfers. The results of this research are useful for different stakeholders, such as public transport authorities, researchers, transport planners and passengers.

1.3.1. Societal relevance

The largest benefit of this research is for the public transport authorities with operating metro systems. The main result is an improved understanding of transferring attributes when passengers are walking to their connecting metro. Therefore, transfer walking time could be better estimated for different passengers or in different service operations. Moreover, the result helps to match the data for trip profiles from AFC and AVL data (Eltved et al., 2021; Zhu et al., 2020). Or metro timetables could be improved to ensure feasible transfers (Yin et al., 2021). Moreover, future stations could be designed for an optimal transfer walking experience. All in all, to provide a better metro service to passengers and improve public transport attractiveness. Lastly, insights into metro-to-metro transfers are currently lacking in the Netherlands. When analysing transfer behaviour or modelling travel trips, the metro is often combined with bus and tram as a city public transport (Kouwenberg et al., 2019).

The public transport authority of Rotterdam, the RET, has the largest benefit because the transfer walking time is collected in one of their metro stations. Therefore, providing valuable data on passenger movement in the station. The RET could use this data and information to make metro station operation decisions.

1.3.2. Scientific relevance

This works studies the effect of passenger and station layout attributes on the (transfer) walking time in metro stations. Therefore, a new extension on the knowledge of transfers in stations. Moreover, the knowledge of passengers' movements in the station improves because this research combines walking characteristics and interaction with vertical transportation. Using different passenger groups in this research, the average population walking time in the (metro) station is better understood.

The results and conclusions of this research on transfer walking times could help to understand metro transfer better. This research will provide an overview of transfer walking time distribution from platform to platform. The distribution might help to estimate the waiting time for transferring passengers and indirectly determine the total transfer time for each passenger in a transfer station. For public transport models or transfer models, this information helps to improve these models. For example, to better schedule metro/train departures in the station to make transfers achievable for most passenger types (Yin et al., 2021) or for the most critical transfers inside the station. Lastly, the results of this study might lead to new insights in information about passenger movements in a station. For RHDHV, the main relevance is possible new insights, which require updating their pedestrian models to simulate walking behaviour in stations. Currently, information is lacking to include different passenger types to model more accurate behaviour of passengers in a station environment.

1.4. Report outline

The report's structure is as follows: The literature review on the transfer walking times and passenger or station layout attributes is given in chapter 2. Furthermore, the chapter presents the research gap on the transfer walking time. Chapter 3 discusses the methodology of this research for answering the sub-questions and the main research question formed by the literature review and research gap. The previous chapter's knowledge gaps provide a basis for the data collection method of the transfer walking times in chapter 4. After the collection, the walking times are analysed in chapter 5 specific for the walking time but also in a generalised form. Besides the data analysis, the chapter presents a simulation tool to give the complete transfer walking time. Hereafter, the walking time is modelled from the collected data in chapter 6 with two different models and per transfer segment type. Finally, based on the results from the walking time collection method, analysis and model, the conclusions, main limitations and recommendations for further studies are discussed in chapter 7.

2
Literature review

This chapter should answer the first two sub-questions through a literature review on transfer walking time attributes. Firstly, in section 2.1, a conceptual model shows the elements which could influence the transfer walking time. Secondly, the transfer walking distribution type is presented in section 2.2. Next, there is a description of current models to estimate the transfer walking time in section 2.3. The last part of the literature review in section 2.4 present the data collection method to determine a transfer walking time. The chapter ends with the literature review conclusions and the following research gap for this study in section 2.5.

2.1. Elements of transfer walking time

A transfer walking time depends on various elements. Therefore, a conceptual model presents all these elements and their possible relationships. The model is given in Figure 2.1 and split into a passenger and layout side. The starting point of the conceptual model is the transfer walking time study of Zhou et al. (2016). The study six main attributes which affect the transfer walking time, the station layout, the transfer path elements and total transfer walking distance as the infrastructural attributes of the transfers. The total transfer path distance is one of the most significant attributes of the transfer walking time (Du et al., 2009; Zhou et al., 2016), which depends on the metro station layout. corridor, the vertical transport and the platform elements. The corridor is a horizontal passage in the metro station, vertical transport is a location where passengers move a level up or down, and the platform is where passengers exit the metro and start their transfer. The alighting location of the passenger at the metro also influences the transfer path distance because those alighting close to the exit have a shorter distance to clear the platform than the other passengers, especially when only one platform is present (Zhou et al., 2016).

Based on these layout elements, two factors can be derived to introduce a certain level of walking time distribution: the crowding and passengers' walking speed according to Zhou et al. (2016). The crowding level depends on the passenger demand in the station and the effective capacity of an element. Furthermore, the walking speed and crowding are related as the demand becomes higher, the walking speed will drop significantly (Z. Chen et al., 2016; Daamen et al., 2005; Zhu et al., 2020). However, the walking speed itself is already dependent on a large number of attributes from the passenger (Daamen et al., 2005; Weidmann, 1993; Bosina and Weidmann, 2017).

A passenger may choose between different vertical transport modes at the vertical transport point. Already the choice between stairs and escalator depends on the level of crowding (van den Heuvel et al., 2015) or height to clear (Li et al., 2014). Each vertical transport mode could have a different travel speed and, thus, require a different time to clear. This aspect relates in two directions between the choice of the passenger because the passenger might choose the vertical transport time on the expected travel time, and the choice of the passenger affects the time to use the vertical transport to the transfer walking time. As walking speed and vertical transport have various influencing attributes, which affect the transfer walking time, these are elaborated in the following paragraphs.

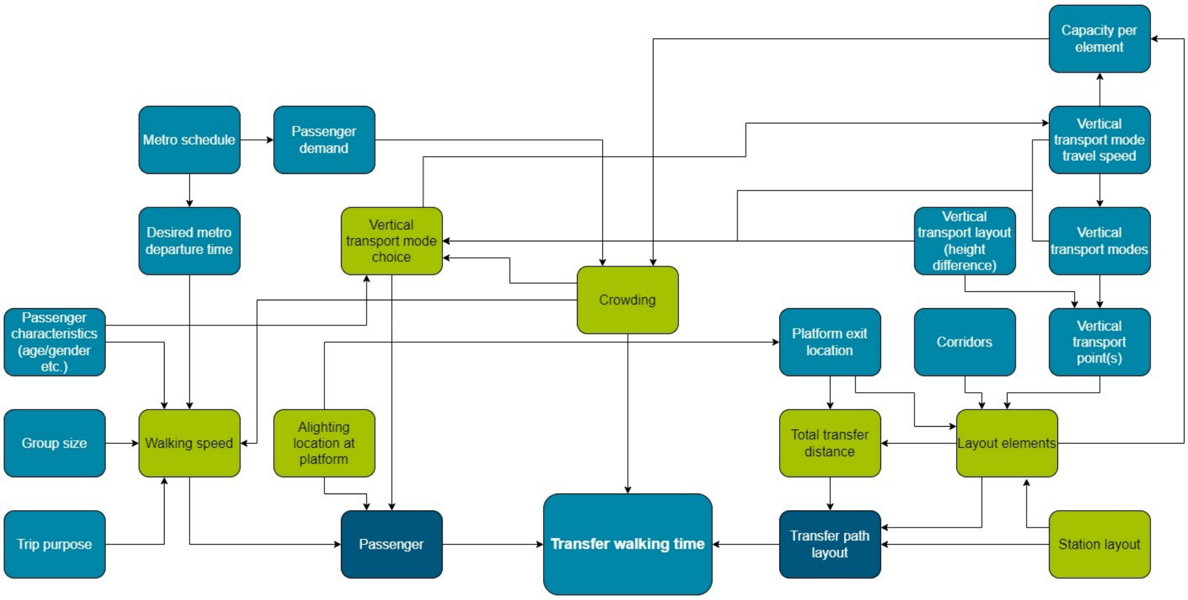


Figure 2.1: Conceptual model of the transfer walking time.

2.1.1. Walking speed

According to Weidmann (1993), the walking speed depends on various factors. The same study proposed an average walking speed of 1.34 m/s because that value is the mean walking speed in the studied literature. This value has been accepted as valid for the walking speed of a single commuter person (Bosina and Weidmann, 2017). However, according to Daamen and Hoogendoorn (2007), the walking speed in the Netherlands is slightly higher at 1.40 m/s. When analysing walking speeds, Bosina and Weidmann (2017) proposes standard categories to use in research. The study's main ones, gender, group size and trip purpose are presented in Table 2.1, including the proposed walking speed per category. Trip purpose shows a variation in the walking speed. Still, the assumption is that most metro passengers are commuters because most public transport trips in the Netherlands are educational or work-related (Centraal Bureau voor de Statistiek, 2021). However, the significantly different walking speeds between gender and group size could be studied, whether they apply to transfer walking times.

Table 2.1: Proposed standard walking speeds (m/s) by Bosina and Weidmann (2017).

Standard walking speed:	Person walks alone as a commuter	1.34 m/s
Gender	Male	1.39 m/s
	Female	1.29 m/s
Group size	1	1.34 m/s
	2	1.21 m/s
	3	1.12 m/s
	4 or more	1.05 m/s
Trip purpose	Commuter	1.49 m/s
	Shopping	1.08 m/s
	Event	1.11 m/s

Moreover, the walking speed shows high variation for certain passenger types (Z. Chen et al., 2016) in Table 2.2 when ascending a stairway in a transfer metro station. Not only for the deviation in the walking speed but also for the mean value per passenger category. Those results for the walking speed are from a normative evening peak hours situation. The attributes of male, middle-aged and heavy congestion were statistically significant in this research, and the walking speed distribution is lognormal. At last, the free walking speed also depends on the level of congestion, and if the opposite flow exists in the corridor (Daamen and Hoogendoorn, 2007), otherwise, the free walking speed significantly drops. Especially when a flow of people crosses the main flow direction. Therefore, the level of crowding is also a significant, influential factor in the walking speed and should be covered in this study on the transfer walking time.

Table 2.2: Walking speed when ascending stairs in transfer metro station (Z. Chen et al., 2016).

Categories	Average (m/s)	Std. dev. (m/s)
Male	0.90	0.36
Female	0.71	0.13
Young	0.84	0.29
Middle-aged	0.69	0.12
Single	0.82	0.26
In group	0.61	0.06
Smooth	0.86	0.32
Slight congested	0.72	0.14
Heavy congested	0.67	0.09

2.1.2. Vertical transport mode choice

One station layout-specific variable has already been heavily studied: the choice of passengers between stairs and escalators as a vertical transport mode. Passengers prefer the route with the shortest time to leave the platform (Daamen et al., 2005) and therefore take the first available vertical transport point. Especially when the waiting time is over 45 seconds, the likelihood is 50% to choose an alternative vertical transport point (van den Heuvel et al., 2015) when an alternative is present. Moreover, the vertical transport's height and the escalator's waiting time affect the decision between the escalator or the stairs (Li et al., 2014). The conclusion is that passengers prefer the escalator over the stairs because of these three studies, mainly for the comfort of not walking up or down. The lift is excluded as a vertical transport mode, probably due to the low expected usage. However, some passengers might rely on the elevator to transfer, so this study includes the lift as a vertical transport mode.

2.1.3. Effect of the moment making the transfer

Based on the literature review on general metro-to-metro transfer-related topics in Lin et al. (2022), the moment of transferring is recommended to investigate further. Especially the difference in time of year between weekdays and holidays. Furthermore, this could also explain the walking speed difference on trip purpose from Bosina and Weidmann (2017) because passengers make trips with a different purpose on other days compared to weekdays. The study of Fujiyama and Cao (2016) found that the moment of making a transfer significantly affects the transfer walking time when connecting between the train and metro. Compared to an early weekday morning and an afternoon at the weekend, the transfer walking time could be twice as long as shown in Figure 2.2. The study did not include the complete transfer from platform to platform but from the train fare gate to the metro fare gate. The larger difference at the weekend could be that passengers do not directly transfer to the metro but have an activity in between. Transfer between train to metro is out of scope in this research, but the recommendation is to collect the transfer walking times at least on weekdays and weekends. Then this study might confirm if the same conclusion from Fujiyama and Cao (2016) applies to metro-to-metro transfers and confirm the hypothesis from Lin et al. (2022).

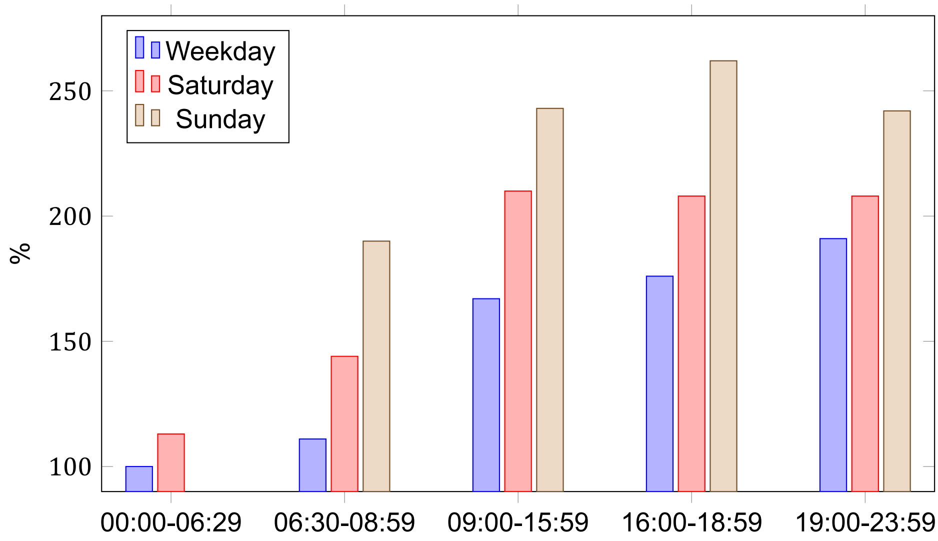


Figure 2.2: Relative transfer walking time compared to the base value (100%) of weekday before 6:29, adjusted from Fujiyama and Cao (2016).

2.2. Transfer walking time distributions

Three studies (Du et al., 2009; Zheng et al., 2014; Zhu et al., 2020) found different transfer walking time distributions in metro stations. All found that the transfer walking time follows a lognormal distribution. None of these studies includes any information from a passenger or layout side. The distributions of (Du et al., 2009 and Zheng et al., 2014) are based on manually collected data. However, the study of Zheng et al. (2014) only collected 20 samples to estimate the distribution, which is too few to be statistically accurate enough (Field, 2013). The resulting fitted lognormal parameters from Du et al. (2009) are given in Figure 2.3. The results yield a median transfer walking time of around 3 minutes off-peak and 5 minutes in peak conditions. At the 80th percentile, the walking time has increased to 8 minutes and 13 minutes for off-peak and peak, respectively. Therefore, the difference between peak and off-peak might be significant for metro transfers, in line with the findings from Fujiyama and Cao (2016).

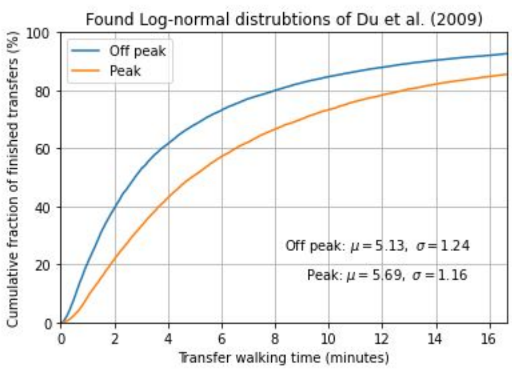


Figure 2.3: Transfer walking time (cumulative) lognormal distributions based on results of Du et al. (2009).

The method of Zhu et al. (2020) uses AFC and AVL data and assumes a uniform walking behaviour and thus walking speed in this model because the study recognises the multiple factors affecting the walking speed, which are hard to model. Based on the AFC data, the walking speed per passenger is estimated through the percentile compared to the fastest and slowest walker. The effect of the transfer path layout is neglected in this study. However, the estimated walking time data showed a clear peak and right-skewed distribution in all nine transfer stations. The peak for the highest frequency of transfer walking time ranged from 20 to 300 seconds. For the 300-second peak, the tail reached 850 seconds.

In other words, that is almost 15 minutes of walking time. The study fitted three distributions for right-skewed data on transfer walking time, the lognormal, Weibull and Gamma distributions. All distributions have a R^2 > 0.80, but the lognormal and Gamma distributions had an R^2 > 0.90 in all stations. The researchers mention that the lognormal distribution was the best fit for all cases, thus for (very) short and longer transfer walking time.

Using a transfer walking distribution rather than a fixed walking time especially helps to ensure successful transfers at the end of metro service (Y. Chen et al., 2019) or even between different rail modes (Long et al., 2020). Both studies modelled the last metro scheduling with one of the inputs being transfer walking time distribution (Y. Chen et al., 2019) or with groups having different walking times (Long et al., 2020). Along the different objectives or model functions, both conclude that using a distribution is better than a fixed value.

2.3. Transfer walking time estimation models

Transfer walking times in metro or public transport systems can be modelled in various ways. Models exist to predict the pure walking time and the total transfer time, which the walking time could be assumed or extracted from. First, an overview is given on the transfer walking time models, followed by models using the main transfer time. Lastly, a model approach is presented, which could cover the effect of including attributes related to the transfer walking time.

2.3.1. Transfer walking time models

From (Du et al., 2009) and (Zhou et al., 2016), different estimation models have been found for transfer walking time in metro stations. The model from Du et al. (2009) is based on transfer path length and the mean transfer walking time, whereas Zhou uses passenger flow, corridor length and escalator capacity. Both are given in Figure 2.4. A second-degree polynomial is fitted through the observed transfer walking times of Du. The R^2 is 0.90, which indicates a perfect linear fit. The researcher also tested a (linear) first-degree polynomial, and the R^2 is 0.85, which is still a very good estimation. The conclusion is that the transfer walking time will increase linearly with the transfer path length. For Zhou et al. (2016), only a linear fit is tested based on the observed transfer walking time and the measured length of the transfer path for all components. The fit is poor, only $R^2 = 0.20$ because for each station 2 samples were followed to collect the transfer walking time to validate Zhou's model. The same transfer path had two transfer walking times. Therefore, it is easy to see the difference in crowding. The observed transfer paths are shorter in Zhou et al. (2016) than in Du et al., 2009, but the transfer walking time is double as long. With a transfer path distance of 100-200 m, the estimated transfer walking time is between 4-7 minutes according to the model of Zhou et al. (2016), while Du et al. (2009) estimates a transfer walking time of only 1.5-3 minutes.

The model of Zhou et al. (2016) provides accurate transfer walking time based on the passenger flow in the station to see the effect of crowding on the walking time. However, this model only considers escalators as a vertical transport mode. While in other transfer stations, escalators could be absent, and other vertical transport modes might have a different travel time compared to the escalator. Therefore this model does not estimate the transfer walking time for all passengers.

2.3.2. Transfer modelling in public transport systems

Transfers in public transport systems consist of two parts. The walking time and waiting time of the transfer. Research for modelling public transport networks has focused on reducing the waiting part (Liu et al., 2021) or ensuring successful transfers. At the same time, others investigate the effect of transfer time on variance for journey times (Dixit et al., 2019). For a minimal transfer time in metro networks, the passengers should be able to arrive at the next platform before the next metro departs. Accurate transfer walking times are vital for ensuring successful transfers for the last metro of the day (Y. Chen et al., 2019). Nowadays, more studies are also focusing on transfer-related attributes on the route choice of passengers to investigate route choice behaviour (Raveau et al., 2014). The presence of escalators in a transfer station is a factor in determining a route within the public transport network (Nielsen et al., 2021).

Nowadays, there is a shift in research methods for determining the transfer walking time due to available data on Automatic Fare Collection (AFC) and Automatic Vehicle Location (AVL). The current practice is

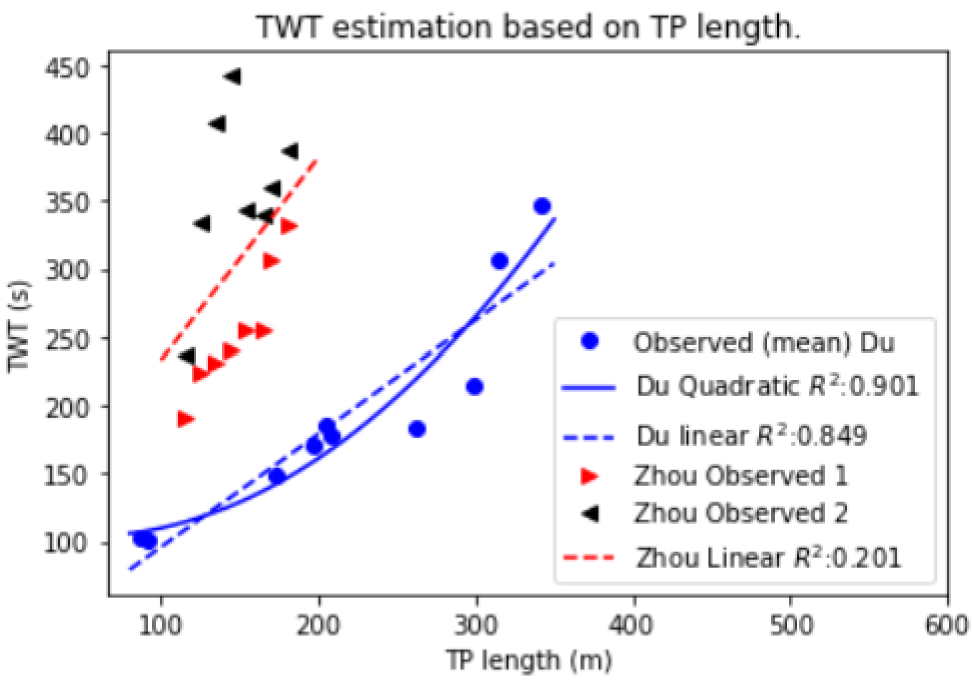


Figure 2.4: Estimation transfer walking time (TWT) through transfer path (TP) length.

to estimate the transfer walking time through AFC and AVL data rather than collect transfer walking time manually (Eltved et al., 2021; Ensing, 2022; Fujiyama and Cao, 2016; Zhu et al., 2020). However, these estimations might only partly cover the total transfer walking time because the are form fare gate to fare gate rather than from platform to platform. Two estimation models are not validated with observed transfer walking times (Eltved et al., 2021; Zhu et al., 2020). Both mention that collecting the observed transfer walking time is too costly and time-consuming. Therefore, collecting the transfer walking times is essential to validate these models. Moreover, Ensing (2022) collected the total transfer time for train-tram transfers with AFC and AVL data. The drawback of the method is the lack of separation between the exact walking and waiting time. Therefore, the exact transfer walking time is unknown with this method per sample.

However, the transfer walking time estimations with station layout variables from Zhou et al. (2016) or with AFC/AVL sources lack information from the passenger characteristics or choices during the transfer walk. While these attributes already influence their walking speed (Bosina and Weidmann, 2017) or choice of vertical transport (mode) choice (van den Heuvel et al., 2015 & Li et al., 2014). Different aggregate parameters could help predict passenger types' transfer walking and waiting times. Improved information on the transfer walking time helps to better schedule metro timetables to ensure feasible transfers, especially at the end of metro service (Y. Chen et al., 2019).

2.3.3. Modelling attributes related to the transfer walking time

One model approach could be suitable to model the effect of different attributes on the (transfer) walking time. The multiple linear regression model is used in the study of Z. Chen et al. (2016) to represent different categories related to the walking speed of passengers. The model uses different passengers and crowding-related variables to predict walking speed in a transfer corridor in a metro station. This model approach seems a good approach to estimate the transfer walking time for different passenger and layout attributes. Furthermore, the transfer path could be split into three specific segment types because Zhou et al. (2016) modelled three types separately; the corridor, vertical transport and the platform. Each segment might influence the related attributes differently, and each transfer metro station is different in layout. However, the best method would be a model for the complete transfer path because then the effect of the attributes could be presented in a general manner.

2.4. Existing data collection techniques

Various data collection methods exist on passengers' movement in stations. Daamen et al. (2015) distinguishes the procedures on the level of measurement objective and perspective for the data collection on passengers' movements. The microscopic perspective follows individuals (trajectories), whereas the macroscopic collects the traffic flow without individual information. Moreover, the local measurement collects data at a fixed spot, while the global measurement is on movements through a bounded network. The research objective is to collect transfer walking time, a point-to-point distance and time to cover in the station layout, thus a global measurement. The transfer walking time differs for passengers. Therefore, extra information on passenger route and choice are only collectable at the microscopic level. At a macroscopic level, only the value of the transfer walking time could be collected. However, with videos from different locations, it might be an alternative to collect the transfer walking time of individuals. Table 2.3 shows the already used data collection methods in the literature, especially on transfer walking time attributes, in relation to the measurement objective and perspective. These tools will be explained further in the following paragraphs.

Table 2.3: Classification for the data collection on passengers' movements in stations, based on Daamen et al. (2015).

		Measurement objective	
		Local (Fixed spot, short path)	**Global (Longer path)**
Measurement perspective	**Microscopic**	Video (Li et al., 2014)	Covert observations (Du et al., 2009 & Zheng et al., 2014) Bluetooth & Wifi (van den Heuvel et al., 2016) Bluetooth, Wifi & Infrared sensors (van den Heuvel et al., 2015)
	Macroscopic	Not applicable for transfer walking time	AFC & AVL (Eltved et al., 2021; Ensing, 2022 ;van den Heuvel and Hoogenraad, 2014; Zhu et al., 2020)

2.4.1. Data collection tools/techniques

Four data collection tools and techniques have successfully studied passenger movements in stations. These are video footage, AFC/AVL, Bluetooth tracking and covert observations. The following section explains each tool or technique and mentions their strengths and weaknesses for transfer walking time collection.

Video footage (tool)

This method uses video recording from surveillance cameras in the station or temporarily installed cameras. From the videotapes, the walking time and relevant characteristics are derived. Video recordings in the station are possible to collect data on walking time (Daamen et al., 2005) or choice of vertical transport mode and crowding (Li et al., 2014) or passenger characteristics during the transfer walk (Zhuang et al., 2018). Especially when the station already has many cameras through the station for safety purposes, the required equipment and, thus, setup costs are low. Furthermore, the collection time is short because older camera tapes are usable. The transfer walking time could be reviewed for all moments in the station, thus providing data on different operation levels. Moreover, Bosina and Weidmann (2017) proposes this method as the most favourable for collecting walking time or speeds. Nevertheless, privacy and analysis are the main disadvantages. Firstly, the research should receive approval from the public transport authority or police to use the recording, and the processing of the videos must follow strict privacy laws. Secondly, passengers' privacy is in danger because the footage is stored somewhere and might show passengers' trip patterns and individual characteristics. Besides, analysing all video recordings to collect transfer walking time on samples is time-consuming.

AFC & AVL (tool)
The AFC and AVL data method uses the check-in/out time of the passenger ticket when passing a fare gate or ticket validation device. Combined with a public transport vehicle's arrival or departure time, the walking time could be estimated for the transfer time. or the entry/exit time in a station. The combination of AFC and AVL as data sources have been used to collect transfer times for multimodal trips (Eltved et al., 2021; Ensing, 2022) or metro trips (Zhu et al., 2020). The advantages are the large data set or samples provided and the little equipment required to collect this data. Besides, the transfer walking time could be estimated for multiple transfer metro stations. Therefore, the tool is cost-effective. However, there are some significant setbacks. Firstly, the method is only usable in AFC systems with obligatory check-in and check-out per mode. Besides, this method primarily determines the total transfer time and requires assumptions on the walking speed (Zhu et al., 2020), making this method unfavourable for transfer walking time collection. Furthermore, the method is most suitable for determining the entry or exit walking time because the passenger passes a fare gate. During a transfer, passengers remain in the fare system. Secondly, the method does not provide any information about the passenger type making the transfer, including the level of crowding. Furthermore, AFC is passenger privacy sensible data. The data set should first be aggregated and anonymised before the data analysis. The data set could be large and much time is going into the data analysis. One of the most prominent disadvantages is the requirement to validate transfer walking time results from AFC/AVL with manual observations, recommended by Eltved et al. (2021). Therefore this method is unfavourable to use for transfer walking time attributes collection.

Bluetooth tracking (tool)
Tracking passengers with a Bluetooth signal is a different method to collect passenger movements in stations (van den Heuvel et al., 2015) because most passengers have devices with Bluetooth. By installing Bluetooth sensors at strategic locations in the stations, the system tracks the Bluetooth signal of mobile phones and the time. With infrared sensors to count the number of passengers, it gives an even better estimation of passenger flow (van den Heuvel et al., 2016). Therefore, this method could collect the transfer walking time. The main advantage is the potential to collect many samples in the station. The drawback is the implementation of strict privacy regulations in the data analysis according to van den Heuvel et al. (2013). Even the passengers' information per sample is in-collectable because it only tracks the movement. Moreover, this method does not detect all passengers (groups) because some might not have Bluetooth devices.

Covert observations (technique)
The final tool insists on following or collecting passenger behaviour without informing the chosen passenger, the covert observation method. (Du et al., 2009; Zhou et al., 2016; Zheng et al., 2014) used the covert observation method in metro transfers: The observer chooses a passenger at the arrival platform and follows the sample until the transfer is complete. When the metro doors open, a timer starts. When the chosen passenger reaches the new platform, the timer stops. Moreover, during the walk, the observer might write down some characteristics of the passenger or infrastructure. Therefore, this method's main advantage is collecting more information than only the transfer walking time. Specific characteristics of each sample could be collected. These help to analyse differences in transfer walking time based on passenger types. Besides, the covered observation method ensures normal behaviour of the passengers and minimal influence on the results, thus providing accurate results on transfer walking time. The largest disadvantage is the collection time because each transfer path in the station requires a minimal number of samples for accuracy. Or more observers for more cost do the collection in less time, or only one observer does all the samples and requires more time to complete the collection. Furthermore, to capture the effect of the moment on making a transfer, samples should be collected on different days or times, increasing the collection time.

All the tools and techniques of the previous paragraph are usable in two different observation situations. The first is in real-world observations, and the second is in an experimental setup.

Real world observations

All of the tools from the previous paragraph are used in operating metro stations. The advantage is the uninfluenced behaviour of passengers moving around in the station. Therefore, any collected empirical data is representative of the total population passing through the station. Moreover, the number of samples to collect is unlimited because many passengers use the station daily. A drawback of real-world observations is to capture all possible operating scenarios of a station because of possible limited collection time. Some operation events have a low probability but could significantly impact the transfer walking time, such as intense crowding due to an event nearby. Furthermore, there is a limitation to test scenarios in the station because public transport authorities prefer the uninterrupted operation of a station. Introducing adjustments in a station for a case of transfer walking time takes a lot of work to achieve.

Experiment in mock-up setup

An option is to build a mock-up or simplified transfer station and study the transfer walking time through analysing video tapes from the experiment. Studies analysing walking behaviour rely on this experimental method of Daamen and Hoogendoorn (2007). The method allows running scenarios to see the effect on transfer walking time in a short timeframe. Furthermore, privacy is not an issue because the experiment uses voluntary participants. However, the voluntary participants risk being an unrepresentative group of the total population, which could influence the results of the data analysis. One of the main drawbacks is the setup time and costs, mainly through the equipment to build the mock-up station and the requirement to use one of the mentioned data collection tools. Furthermore, passenger behaviour is in question because the experiment setup might influence how the samples move around compared to the real world. Besides video recordings, covert observations could also be used as a data collection tool in this scenario.

2.4.2. Data collection tool/technique choice in this study

In Table 2.4, all data collection tools and techniques are assessed on certain aspects. The assessment has general points when comparing different data collection methods or is about the research objective to collect the transfer walking time and the related attributes. Using AFC/AVL or Bluetooth is assumed unusable to collect the transfer walking time as both cannot the actual walking time or the passenger behaviour. Therefore, these methods are unfavourable to use in this research. The other two methods, videos and covert observations score neutral or positive on most assessment points. However, the largest drawback of using videos is the passengers' privacy because of strict privacy laws in the Netherlands. Using and storing videos for data processing has significant privacy risks, and public transport operators would not allow this method. Therefore, covert observations are the most favourable method because it has only moderate disadvantages and can fulfil the research objective of collecting passenger attributes which could influence the walking time. The only slight drawbacks of the method are the required time to perform the collection and the possible relatively small sample size compared to the other methods.

The assessment of the data collection situation shows that performing in the real world favours is the best approach. On all points the expected score is negative, except for collecting attributes of the transfer walking time and the passenger privacy. Setting up an experimental transfer station is too costly and time-consuming for this research. Besides, the participants in the experiment might show influenced walking behaviour because of the setup. Therefore, this study will collect the transfer walking time in an operating metro station, through the method of covert observations.

Table 2.4: Comparison used data collection tools in previous studies and score as positive (+), neutral (0) or negative (-).

Collection tool	Videos	AFC & AVL	Bluetooth	Covert observations
Privacy samples/passengers	-	0	0	+
Availability	0	0	0	+
Set-up/Equipment effort	+	+	-	+
Analyzing time/effort	0	-	-	0
Collect (actual) TWT	+	-	-	+
Collect attributes on TWT	+	-	-	+
Collect (normal) passengers behaviour and/or choices	+	-	0	+
Data storage	-	0	-	0
Data size	+	0	+	0
Number of samples	+	+	+	0
Situation data collection	**Real world observations**		**Experiment**	
Privacy samples/passengers	0		0	
Availability	+		-	
Set-up/Equipment effort	+		-	
Collection time	+		-	
Analyzing time/effort	0		0	
Collect (actual) TWT	+		-/0	
Collect attributes on TWT	+		+	
Collect (normal) passengers behaviour and/or choices	+		-	
Number of samples	+		-	

2.5. Conclusions form literature study and knowledge gap

Regarding the framework in Figure 2.1, most studies focused on the station layout and metro service side of the transfer walking time. The passenger side is an unobserved side from the walking time, while it has one of the essential attributes, such as the walking speed and vertical transport mode choice. The following conclusions follow from the literature review:

- The walking speed mainly depends on gender, group size and level of crowding.
- Passengers take the shortest route when leaving the platform.
- Congestion before the escalator affects passenger choice on vertical transport between escalator or stairs.
- In vertical transport mode choice studies, the lift is left out.
- The transfer walking time might depend on the moment of the day.
- The transfer walking time follows a lognormal distribution.
- Collection methods for transfer walking times use AFC/AVL data sources, videotapes or covert observations in operating stations.
- Currently, transfers or transfer walking times are modelled through disaggregate AFC/AVL data sources.
- Transfer walking distributions are preferred over fixed walking times for modelling the last connections of the day between metro or other rail modes.
- The best model approach for the transfer walking time and the related attributes could be a combination of a multiple linear estimation model and dividing the transfer into specific segments.

The research gap in Figure 2.5 follows from the literature conclusions. A study on the effect of transfer walking time distribution regarding at least variables such as gender, group size and crowding from the walking speed perspective. From the stations' layout, the lift should be included in the vertical mode choice besides the stairs and escalator. With these variables, the transfer walking time could be better determined for certain passenger groups or compared with the found distributions through the AFC/AVL models. Besides, in that way, metro scheduling could be improved, especially for the last metros of the day. The best method of collecting the attributes from the literature study and the walking time is using covert observations in an operating metro station because mainly of moderate privacy problems, easy setup and having uninfluenced behaviour of passengers. The following methodology from the literature review and research objective is given in chapter 3.

The answer to the first research question is that from a passenger's perspective, the walking time depends on the walking speed, which is related to the gender, group size, the trip purpose of a passenger and the presence of crowding. Moreover, the vertical mode choice relates to the walking time from a station layout perspective. Between stairs and escalators, passengers prefer the escalator. Lastly, most passengers use the closest exit of a platform. From a time perspective, the moment of making the transfer might have an influential effect because of the already significant difference in walking times between train-metro transfers.

The transfer walking times are mostly collected through covert observations. Videotapes and tracking through Bluetooth are also possible tools for collecting passenger movements in a station, but have privacy and data storage risks when using this method. The present way of obtaining the transfer walking time is through AFC and AVL data sources estimations. However, the drawback of this method is the lack of validation with observed walking times in a station. All in all, this is the main conclusion to the second research question.

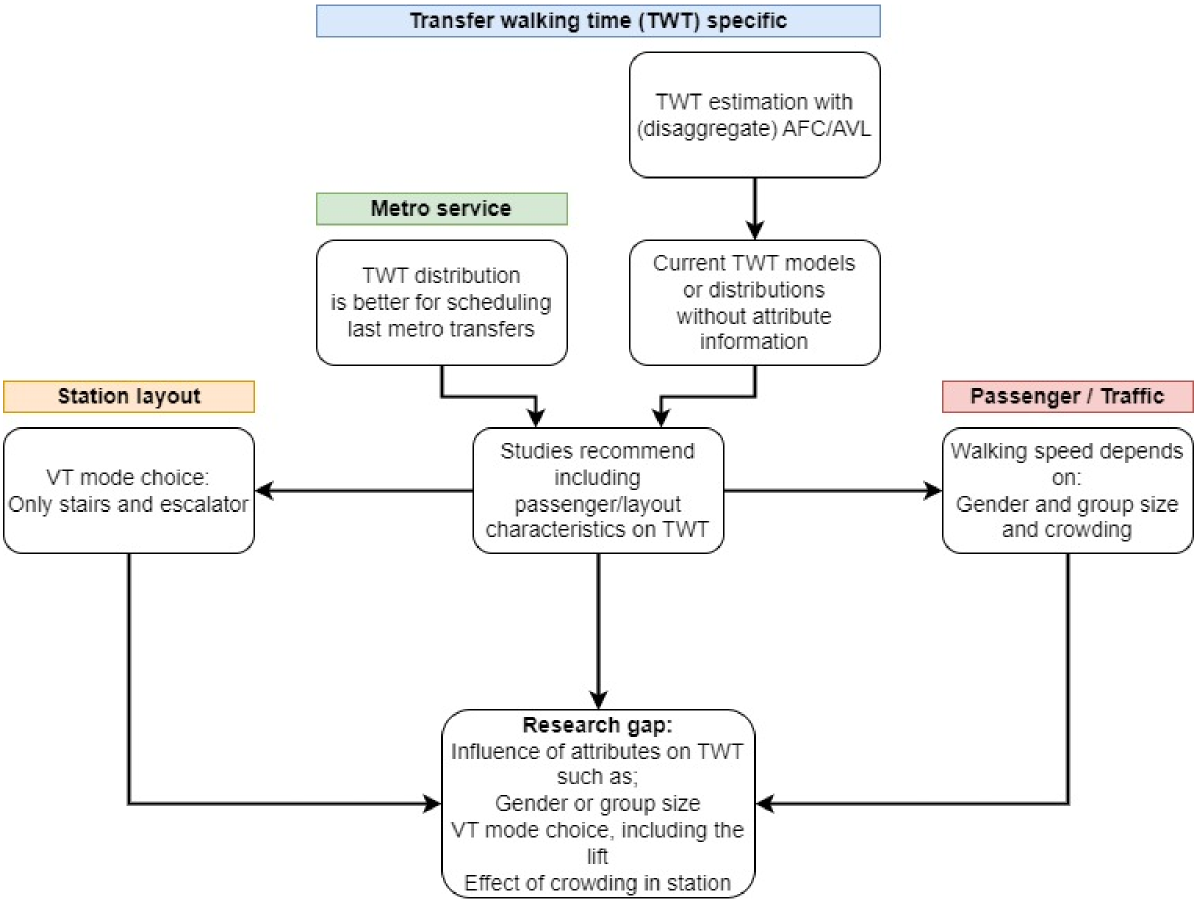


Figure 2.5: Research gap from the literature review and usage in this research.

3

Methodology

This chapter presents the methodology for this research to answer the main and sub-research questions. Figure 3.1 presents the framework for this research. The methodology is divided into four blocks: Literature review in section 3.1, data collection in section 3.2, data analysis in section 3.3 and modelling in section 3.4. Each block is explained in the following sections.

3.1. Literature review

The first part of the research answered the first two sub-questions and decided on the used data collection method in this study based on a scientific literature review in chapter 2. Using the conceptual model in Figure 2.1 on the transfer walking time, important attributes were found from a passenger and station layout perspective. Therefore, the first research questions could be answered.

The second part of the literature answered the current estimation methods of the transfer walking time. Furthermore, the tools to collect the transfer walking time were discussed. The conclusions from these two parts answer the second research question. Combined with the found attributes, a research gap was found to transfer walking times. Based on the gap, this research is split into three phases after the literature review: the data collection and analysis of the transfer walking time, followed by the modelling of the transfer walking time. First, walking time data should be collected and analysed to determine the significant attribute of certain variables on the transfer walking time. With the information on important variables of the walking time, a model can be made to predict the effect of each variable.

3.2. Data collection of transfer walking times

The chosen method is performing a covert observation from the research gap and the available collection tools for capturing passenger movements in chapter 2. The initial plan is to collect the walking time and the influential variables from the literature review in line with the method from Du et al. (2009). However, during this phase, some additional variables might be included because, from the literature, the list of variables is limited or indirectly follow from other variables. During the collection setup, there might be unobserved variables which have not been covered yet by literature. These variables could be included in the collection and the research. The data collection is the source of information to answer the third research question in the next phase, the data analysis.

The collection is performed at the one station as mentioned in section 1.2 because it meets the minimal requirements of a transfer path. However, performing the collection at only one station is a limitation because it could make the conclusions in the data analysis and the remainder of the research very specific to only this station. Therefore, the results of this research could be biased. Furthermore, the collection will present limitations and assumptions, which are going to be used when analysing the results.

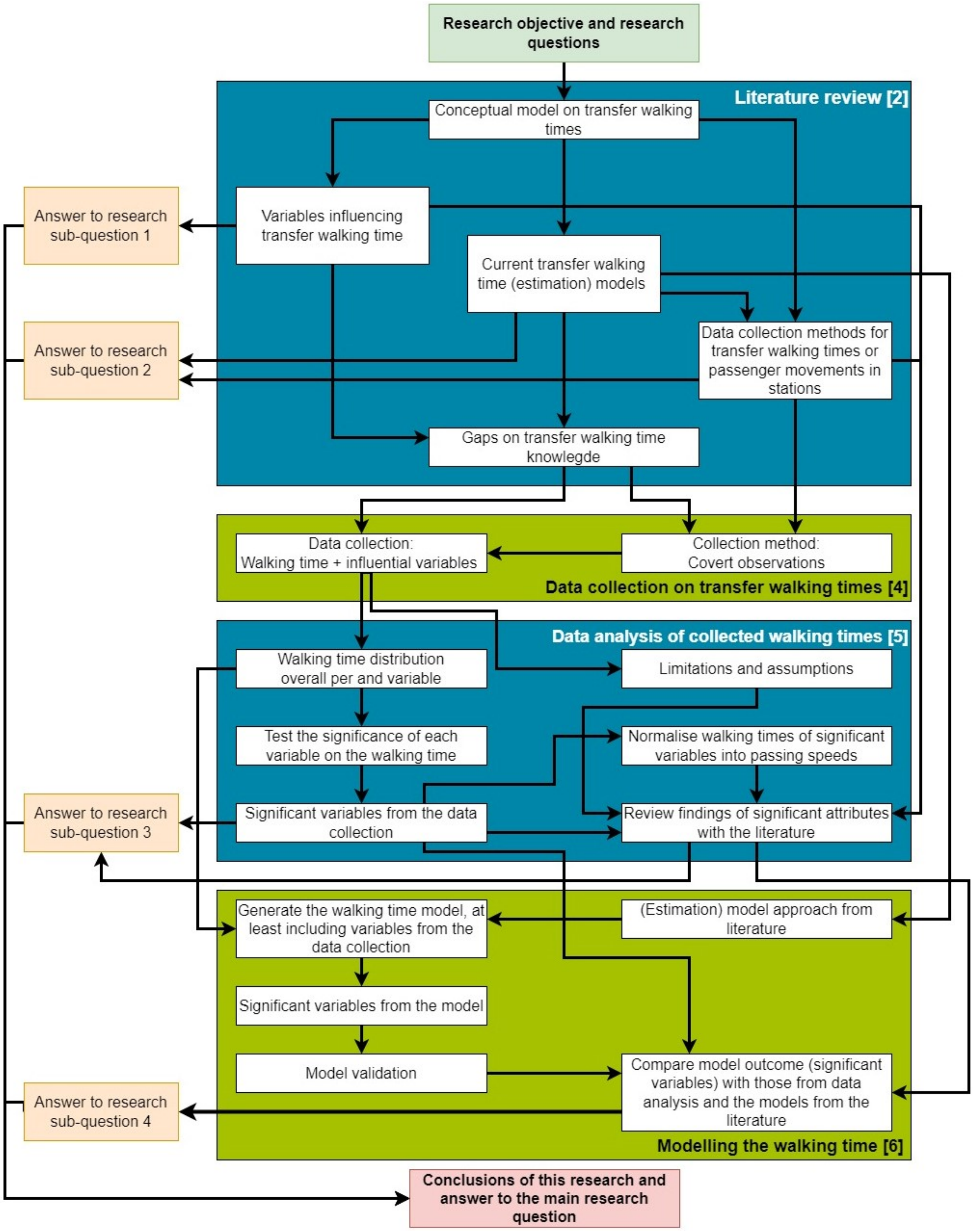


Figure 3.1: Framework of this research, including the chapter numbers as in the square brackets.

3.3. Data analysis of collected transfer walking times

The raw data collected in the previous phase is analysed in this phase. The data analysis is performed with Python code on Juypter Notebook software. The first step is to obtain descriptive statistics of the overall walking time per transfer path and observe the general walking time distribution. According to the literature, the expected walking time distribution is lognormal. The conclusion of distribution type comes from the skewness parameter, the Kolomogorov-Smirnov test and the qualitative observation of the walking time histograms. The distribution type determines the statistical test (Field, 2013) for detecting the significance between the variables. If a lognormal walking time distribution is expected, then the assumption is that non-parametric tests will be used to test the significance. The exact statistical test depends on the number of attributes to compare simultaneously. All tests are performed on a 95 % significance level ($\alpha = 0.05$). The result of this phase is the significant variables of the walking time for metro-metro transfers.

The significant variables on the walking time are then related to the ones found in the literature, and the walking times are generalised into passing speeds. The assumption is that the walking time might include some hidden waiting time, and, thus cannot represent the walking speed of passengers. Therefore, the term passing speed is used rather than walking speed. After the variables review of the literature, the main conclusion can be given to the third research question. Passing speeds are close to the walking speed of passengers because the assumption is that the walking time of samples might include parts when the sample is walking significantly slower due to crowding (Daamen and Hoogendoorn, 2007). Therefore, passengers' walking speed, the normalised variable of the walking time, is always assumed to be lower or equal to their preferred walking speed. These passing speeds will be compared to the studied literature on walking speed to test if the hypothesis on passing speeds is true.

3.4. Modelling the transfer walking time

The final part of this study consists of modelling the transfer walking time for answering the final sub-question. The exact model procedure will depend on the data collection setup and the results of the data analysis. However, the model type could align with the method of Zhou et al. (2016), which split the transfer into three segment types: corridors, vertical transport and platforms and made a separate model for each segment type. As the planned data collection has variables to collect per sample besides the walking time, the suitable model type might be a multiple linear regression model (Field, 2013) to estimate the effect of certain variables on the transfer walking time, in line with the approach from Z. Chen et al. (2016). The model input is all transfer walking time samples from the data collection to see if the model predicts the same or additional significant variables as the data analysis because the model might highlight additional variables, which had an insignificant difference from the covert observations in section 3.2. Ultimately, the model is validated. The exact method depends on the model type by comparing the model estimation of the transfer walking time with the observed walking time of one of the transfer paths.

When the model gives the significant variables, those are compared with those found in the data analysis and the literature from the models. if the model procedure of Zhou et al. (2016) is used, then there might be different significant variables at a segment type level. Based on the model validation and the comparison with the literature, the answer can be given to the fourth sub-question. The answers to all sub-questions are the answer to the main research question and, thus, the conclusion of this research.

4 Data collection

This chapter helps to answer the third sub-question, which attributes contribute the most to the transfer walking time. Therefore data collection is performed. The structure of the chapter is as follows. Firstly, an overview of the data collection objective and its relation to the research questions is given in section 4.1. Next is the data collection setup in section 4.2. The tasks before the collection start are given in section 4.3, and after that, the collection pilots performed are discussed in section 4.4. The chapter ends with the data collection reflection and limitations in section 4.5.

4.1. Objective data collection

The data collection helps to provide evidence of which attributes might influence the transfer walking time. Furthermore, as stated in the scope, two elements are already part of a transfer path. A corridor and a vertical transport point each might have a different role in the time to traverse each element. The transfer walking time attributes from the literature review are the time of day, passengers' characteristics on the walking speed and the crowding level. Therefore, these items are included in the data collection. Furthermore, the lift should be part of the vertical transport mode because data is lacking for that mode. The research objective is to study the attributes influencing transfer walking time. To collect normal behaviour of samples, the tool of choice is the covert observation technique in real-world observations. Moreover, the method has been used successfully in previous studies on transfer walking times. However, the privacy of passengers is a top priority with this technique.

The covert observation method could collect the transfer walking time when the setup has a clear description. The setup plan should include the exact covert observations method because the accuracy of the transfer walking time depends on it. Moreover, this method required approval from various (public transport authorities) instances before the collection period, which is explained in subsection 4.2.2. Using the covert observation method assures naturalistic passenger behaviour. However, there is a risk that passengers could behave slightly differently because the collection is done in person. Besides, the stations should be open and accessible to perform covert observation. Lastly, this method allows for the validation of future models on the (transfer) walking time because of the realistic observed walking behaviour.

4.2. Setup data collection

This section discusses the setup for the data collection. Firstly, the used method of covert observations is discussed. Followed by a description of the case study metro station transfer layout. Hereafter, the used variables to collect are given, which might influence the transfer walking time.

4.2.1. Method of data collection

For two reasons, the research uses an adjusted version of the covert observation method of Du et al. (2009); Zheng et al. (2014). Firstly, both studies followed a transferring passenger for the complete transfer path each time. Therefore this method is time-consuming and privacy-invasive because each passenger has to be followed for a long stretch. Furthermore, in the chosen station, the transfer path overlaps with the exit route. Therefore, the probability is significant that the observer would follow an exiting passenger than a transferring person. All in all, an adjusted covert observation method will be used in this research.

The method cuts a transfer path into sections wherein the data is collected separately. In Figure 4.1, an example is given on the segment division of a transfer path with the minimal requirements from the research scope in section 1.2. Because of the layout, the segments could be different in opposite directions of the same transfer path. The first segment is at the arrival platform of the metro or train because the starting time of a transfer walking time is when the passenger leaves the arriving metro or train. The last segment will be the last corridor or vertical transport before the other platform. There is no additional section at the desired/departure platform because the transfer is complete when the passenger steps onto the platform. In this way, the transfer path route has the same starting and finishing point as the study of Du et al. (2009). The exact number of sections depends on the station and transfer path layout.

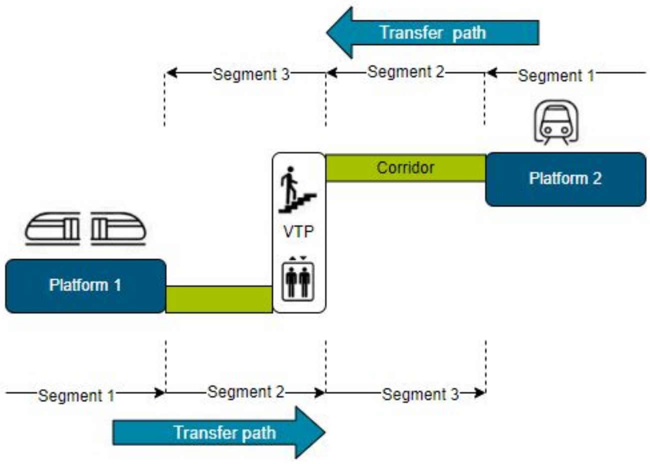


Figure 4.1: Example of a transfer path split into segments for both directions.

In this research, the researcher is the observer in the data collection. The observer stands or sits at a fixed point in a segment, with a good view of the complete segment. In each segment, the observer randomly selects a passenger. The observer clocks the time of the passenger walking through the segment. During or after the timing, the observer notes some (passenger) characteristics which are explained in subsection 4.2.3 and subsection 4.2.4. A precondition is that the sample walks continuously and only stops due to crowding or short activity. The short activity involves cases such as orientation to find the correct route in the station, tying a shoe or waiting for a group to be complete. Longer activities in the range of buying a ticket or visiting a shop in the station will result in discarding that sample in the collection.

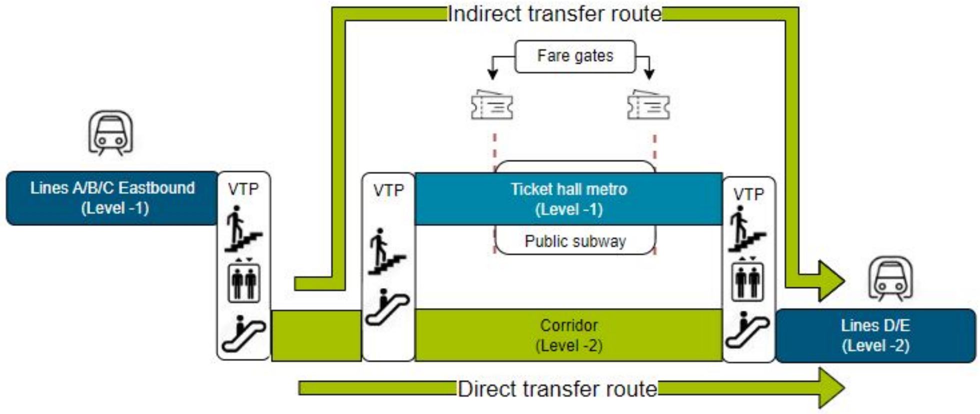
4.2.2. Location for data collection

This paragraph gives the station's motivation to collect the data. The Netherlands has only two metro stations with the transfer elements from section 1.2 Centraal Station in Amsterdam and Beurs in Rotterdam. Beurs has multiple transfer paths with different layouts, while Centraal Station only has one transfer path. Using multiple transfer paths in the data collection ensures a more generalised data analysis and modelling of the transfer walking time. Therefore, the data is collected at Beurs station in Rotterdam, the Netherlands. In the metro network of Rotterdam, as shown in Figure 4.2, the station is the only point in the metro network to change directly to one of the five metro lines.

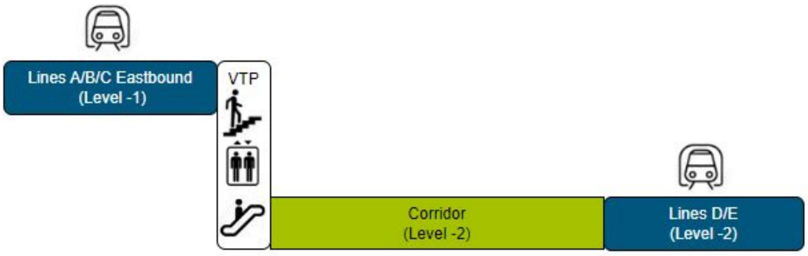
Figure 4.2: Location Beurs station in the metro network of Rotterdam.

The station choice also has societal relevance because the public transport authority of Rotterdam, the RET, is interested in Beurs's (transfer) walking time results. Plans for rebuilding these stations are slowly starting (MRDH, 2023) because the station already has some capacity issues "with 60,000 metro transfer passengers per weekday" (H. Kranenburg, personal communication, August 11, 2022). Collecting the transfer walking time and the corresponding attributes helps the RET understand the current transfer and potentially optimise it. Therefore, the RET approved the data collection at Beurs station.

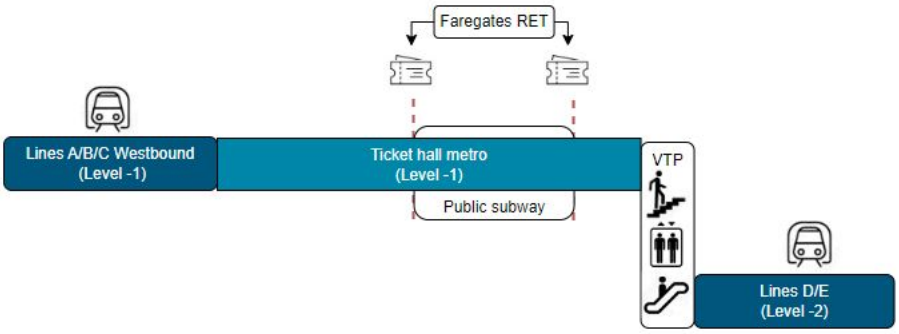
The station has two levels. Level -1 contains Lines A, B, and C platforms in an east-west direction and the main ticket hall. Under the A/B/C lines at Level -2 are the Lines D and E platforms in a north-south direction. From above, the station layout has the form of an upside-down T and is symmetrical. Therefore, the transfer paths from and to Lines D and E platforms have the same layout. A schematic overview of the station layout is in Appendix B. Four main transfer routes exist based on the signage in the station. Two from A/B/C Eastbound (EB) to lines D/E, one from/to lines A/B/C Westbound (WB) from/to lines D/E and one from lines D/E to lines A/B/C Eastbound. Figure 4.3 shows all transfer routes in Beurs, including the elements schematically. From Lines A/B/C Eastbound, there are two routes (1 and 2). One is the most direct, while the other involves two additional vertical transport points and passing through the ticket hall. The remaining transfer routes (3 and 4) have a corridor and one vertical transport point.



(a) Beurs transfer 1, 2: Lines A/B/C Eastbound → Lines D/E with direct (1) and indirect route (2).



(b) Beurs transfer 3: Lines D/E ↔ Lines A/B/C Eastbound.



(c) Beurs transfer 4: Lines A/B/C Westbound ↔ Lines D/E.

Figure 4.3: Schematic (officially signed) transfer routes in Beurs station with elements.

Each transfer route in Beurs has the minimal requirement of section 1.2, having a corridor and one vertical transport point. Furthermore, passengers can choose between three vertical transport modes for each transfer, the escalator, stairs or the lift. All transfer paths are divided into segments in the pre-pilot collection. Each path consists of at least three segments from the transfer paths split up in Table 4.1. Due to overview constraints during the pre-pilot, the transfer is split into two routes, one by the escalator and one by the lift or stairs. The exact segment boundaries are given in Appendix B.

Table 4.1: Transfer path segment numbers.

Transfer path	Segments (excluding the platform)
A/B/C Eastbound to D/E Direct	OV1→OV2→OV3→OV4
A/B/C Eastbound to D/E Indirect	OV1→OVI2→OVI3→WV2 →WV3
A/B/C Westbound to D/E	WV1→WV2→WV3
D/E to A/B/C Eastbound	VO1 →VO2 →VO3 →VO4
D/E to A/B/C Westbound by Lift/Stairs	VO1 →VWL →VW3→VW4
D/E to A/B/C Westbound by Escalator	VWR1 →VWR2 →VW3 →VW4

4.2.3. Passenger types considered

The passenger types considered in this research are based on the recommendation of standard conditions of the walking speed studies setup from Bosina and Weidmann (2017) and the collected walking speed for different groups in the study of Z. Chen et al. (2016).

Gender

Firstly, an essential factor is gender (Bosina and Weidmann, 2017; Z. Chen et al., 2016) because females walk slower than males. Therefore, gender is included in this research because a different walking speed gives a different walking time. Only samples with clear gender-specific characteristics will be collected to avoid guessing the gender wrong.

Group size

Secondly, the group size of a passenger group shows variation in walking speed. In Z. Chen et al. (2016) made the difference between walking alone or accompanied, and the result was that the last one showed a lower variance than walking alone. While Bosina and Weidmann (2017)) propose walking speeds based on the group size because the walking speed varies significantly between groups 2-3 to groups of 4 or more. Therefore, only when a clear group is recognisable by the observer than the group size is noted because, during crowding, it is hard to see clear groups.

Walkability

Thirdly, none of the studies of (Bosina and Weidmann, 2017; Z. Chen et al., 2016) takes into account the effect of disabled passengers with walking difficulties, for example, passengers in an (electrical) wheelchair or who require a walking stick to walk, especially older passengers might fall in this category. While these groups might move significantly slower, increasing the minimal transfer walking time for them and might choose a different vertical transport mode compared to non-disabled passengers. Besides, the RET would like to collect more walking information about disabled passengers in their metro stations. Therefore the difference in physical ability is included in this data collection.

Luggage

Lastly, the influence of having luggage only plays a role where heavy luggage is usual (Bosina and Weidmann, 2017) on the walking speed. Small backpacks or suitcases almost never impact the passenger's walking speed. However, stations near large shopping areas might lead to more passengers with large shopping bags, especially at weekends. Thus, only the difference between having large luggage or not is present in the data collection. The large luggage items to look for are strollers, multiple shopping bags, bikes and large suitcases. The RET allows regular bikes in the metro during off-peak moments, while foldable bikes are always allowed (RET, 2022b). All in all, Table 4.2 presents the passenger types to study in this data collection plan.

Table 4.2: Planned Passenger variables to use this data collection.

Gender	Group size	Walk-ability	Luggage
Male	Alone	Non-disabled	No or small luggage
Female	In group	Disabled	Large luggage

4.2.4. External factors considered in the data collection

Not only passengers' characteristics but also some external factors on the transfer walking time are included in the collection. These are crowding, vertical transport mode choice and alighting point.

Crowding in terms of Level of Service (LOS)

During the transfer walk, a passenger might encounter crowding for various reasons, such as capacity restrictions of a station element or high demand in a section. In crowding, the walking speed of the passenger could be lower than the desired walking speed (Z. Chen et al., 2016; Daamen and Hoogendoorn, 2007). Therefore, it could take longer to clear a certain segment for that passenger. Only the information if a sample encounters (local) crowding in a segment is noted in this research. The level of crowding is scored on a qualitative scale because of the chosen data collection method. The observer should distinguish certain levels of crowding easily. Fruin (1992) provides an overview of crowding through six different LOS from A to F. For simplicity, the number of LOS is three in this research because of an easier qualitative difference between crowding levels for the observer. The level definitions are free, crowded and queued, depending on the distance between the sample and other passengers.

Vertical transport (VT) mode choice and waiting time

In the segments with vertical transport, the VT mode choice of the sample will be noted. Passengers choose their preferred VT mode during the transfer walk. Sometimes a sample might encounter a waiting time because of high demand exceeding the capacity of a VT mode or for the arrival of a VT mode (in the case of the lift). Then the waiting time to board a VT mode is noted as well. Lastly, the waiting condition is part of the collection. A sample could board a Vt mode directly or has to wait. A difference between obligatory and voluntary wait is part of this research. The compulsory wait is through capacity restrictions of a VT mode, whereas the voluntary wait is the choice of the passenger. This voluntary wait is mainly between stairs and escalators. Passengers have a preference for escalators over stairs. A sample could wait in queue for the escalator while the adjacent stairs are free-flow without anyone on it. In that case, the waiting condition will be noted as a choice of the sample.

Alighting location and time

The first segment of each transfer path is at the alighting platform as stated in Figure 4.1. In Rotterdam, most metro services are with six carriages, so the platform is divided into three parts, where each one representing two carriages. Besides the alighting location, the starting time of the samples is different than at the other segments. The timer will start when the metro doors open, not only to be in line with the study of Du et al. (2009) but also to include the effect on passengers having trouble exiting the metro. In this way, the walking time better represents the population of passengers exiting a metro and leaving the platform.

4.2.5. Removal of certain categories and final ones in the data collection

Each planned variable to note from the previous paragraphs for each sample in subsection 4.2.3 and subsection 4.2.4 are checked if there were feasible to collect during the pre-pilot and pilot collection at Beurs station. All categories are again given in Table 4.3. The reason to drop walkability is the expected very low number of passengers with a physical disability using the station. Collecting a sufficient sample size of that variable in each segment requires a significantly longer collection time. Therefore, this category is not used in this data collection, but if a disabled passenger is present, that one will be collected and a remark will be given that the passenger has a disability.

After the pre-pilot in Beurs, three variables were dropped again: ”Queue”, ”Waiting time” and ”Wait (choice)” The number of samples that encountered crowding level ”Queue” was very limited at the platform. It was completely absent in the corridors while it was peak period. Furthermore, the observer had difficulties seeing the difference between ”Crowded” and ”Queue” with the set conditions. Therefore, the variable ”Queue” is no longer part of the collection. The other variables (”Free flow” & ”Crowded”) remain with their conditions. Moreover, the difference between waiting as a choice and due to capacity was hard to see by the observer and very limited samples waited for the escalator as a choice. All in all, a reason to drop this variable and ”Wait (must)” is simplified to ”Wait”.

Collecting the waiting time before boarding a vertical transport mode separately was very hard during the pre-pilot collection because of the exact start of the wait, especially for the stairs and escalator.

Even with the definitions from Table 4.4 for the capacity constraint, the starting point for a sample is an issue for the observer. Therefore, this item was dropped and will not be used further in the collection. The variable describing the waiting condition already helps to determine if a sample has to wait.

Table 4.3: Control variables part of the data collection.

Category	**Variable**	**Considered**	**Pre-Pilot**	**Pilot**	**Actual collection**	**Reason exclusion**
Gender	Male	✓	✓	✓	✓	
	Female	✓	✓	✓	✓	
Luggage	No/small	✓	✓	✓	✓	
	Large Male	✓	✓	✓	✓	
Walk-ability	Non-disabled	✓				Insufficient sample size disabled
	Disabled	✓				
Groupsize	Alone	✓	✓	✓	✓	
	In group	✓	✓	✓	✓	
Crowding level	Free walking	✓	✓	✓	✓	Boundary between crowding and queuing
	Crowded	✓	✓	✓	✓	
	Queue	✓	✓			
Vertical transport mode choice		✓	✓	✓	✓	
Waiting condition for boarding VT	Direct	✓	✓	✓	✓	Determining waiting as a choice
	Wait	✓	✓	✓	✓	
	Wait (choice)	✓	✓			
Waiting time for boarding VT		✓	✓			Starting point of waiting

The final selection of variables to collect during the data collection with a description is given in Table 4.4. Besides the walking time, four main categories are part of all segments, two specifically for the vertical transport segment and one only for the platform segments. The observer himself determines the best category choice per attribute. Therefore there is a risk of biased results because the observer might have a false stereotype about certain expressions of samples. Nevertheless, the observer will try to choose randomly as best as possible to ensure a valid transfer passengers population in the data.

Table 4.4: Final variables to use in the data collection with definitions.

Item	Definition
Walking time	
Time	The time (in seconds) it takes for a sample to walk through a segment.
Gender	
Male	The (clear) male appearance, based on face, clothing and body form.
Female	The (clear) female appearance, based on face, clothing and body form.
Luggage	
Small	Up to two (different) luggage pieces and items are smaller than 55 x 30 x 20 cm.
Large	More than two different luggage pieces and/ or an item larger than 55 x 30 x 20 cm.
Group size	
Alone	Sample walks alone or group size is indeterminable.
In group	Sample walks in a clearly recognize group.
Crowding	
Free	No other passengers around the sample within a radius of more than 2 meters.
Crowded	Other passengers are within a 2-meter range of the sample.
Vertical transport mode (only in segments with vertical transport modes)	
VT Mode	Stairs, escalator or lift
Waiting condition (only in segments with vertical transport modes)	
Wait	Waiting 1) due to capacity restrictions. 2) Waiting for the escalator to change the preferred direction. 3) Lift doors do not open directly as the lift has not arrived.
Direct	Sample can board desired vertical transport mode directly.
Alighting location (only at platform segment)	
Front	In travel direction, the front two metro carriages, roughly the front $1/3$ of the platform.
Middle	In travel direction, the middle two metro carriages, roughly the middle $1/3$ of the platform.
Back	In travel direction, the back two metro carriages, roughly the back $1/3$ of the platform.

4.3. Tasks before collection

Before the start of the collection, the minimal sample sizes are determined, and moments to perform the collection are planned. Furthermore, the limitations and assumptions during the collection are presented.

4.3.1. Minimal sample size

The walking time's minimal sample size (n) can be calculated through two methods. Firstly, the sample size is calculated with the method of (Richardson et al., 1995). The expected error is set at 5 %, which gives a Z value of 1.96 in Equation 4.1. Transfer passengers numbers are large in a station. Thus, the population size N is set at 1000 by the researcher's choice. The transfer walking time distribution parameters of Du et al. (2009) are used to determine the sample size. The corresponding values are the mean ($\mu = 5.13$) and the standard deviation ($\sigma = 1.24$). The required sample size for the walking time is 84, calculated with Equation 4.1-Equation 4.3 from Richardson et al. (1995).

$$s.e(\mu) = \frac{error * \mu}{Z_\alpha} = \frac{0.05 * 5.13}{1.96} = 0.13 \tag{4.1}$$

$$n' = \frac{\sigma^2}{s.e^2} = \frac{1.24^2}{0.13^2} \cong 91 \tag{4.2}$$

$$n = \frac{n'}{1 + \frac{n'}{N}} = \frac{91}{1 + \frac{91}{1000}} = 84 \tag{4.3}$$

According to Dekking et al. (2005), the sample size is calculated through Equation 4.4. The study of (de Dios Ortúzar and Willumsen, 2011) states that the value for desired accuracy d is the researcher's choice. Previous studies showed that the transfer walking time follows a lognormal distribution and depends on many variables. Therefore, the choice for the confidence interval is 25 % within the mean. In other words, the value for d will be 0.25. The same error range is set at 5 % and the same standard deviation ($\sigma = 1.24$) as the previous calculation. The result with Equation 4.4 is a minimal sample of 95. Both calculations have a larger sample size than the minimum of 30 (Dekking et al., 2005). Moreover, the sample size is close to size of Du et al. (2009), which had 100 samples for the transfer walking time. Therefore, the minimum walking time sample size for each segment is 95.

$$n = \frac{\sigma^2 Z_{\alpha/2}^2}{d^2} = \frac{1.24^2 * 1.96^2}{0.25^2} = 95 \tag{4.4}$$

For the discrete categories, the minimal sample sizes are also determined with a different method of Richardson et al. (1995). The shares within each category are obtained from the 12 pre-pilot samples and are presented in Table 4.5. The exception is the vertical transport-specific categories sample sizes because of the unknown share of elevator users. So, a minimum of 5 samples is the target per vertical transport mode, but mainly for the elevator. The alighting location was also not part of the pre-pilot, but the assumption is that alighting occurs uniformly along the platform, or in other words, around 33 % per platform part. For each segment, the target is to collect these sample sizes within the calculated sample size based on the walking time in Equation 4.4.

Table 4.5: Minimal sample sizes for each category calculated by Richardson et al. (1995) from pre-pilot collection.

Category	**Samples**	**Sample share (%)**	**(Chosen)** p	**Minimal sample size:** $n_c \geq \frac{Z_{\alpha}^2 p(1-p)}{d^2}$
Gender	Male: 8 Female: 8	50 % 50 %	0.5	16
Luggage	No/small: 11 Large: 1	92 % 8 %	0.92	5
Groupsize	Alone: 8 In group: 4	67 % 33 %	0.67	14
Crowding	Free: 5 Crowded: 7	40 % 60 %	0.40	15
VT Mode	Not part in pre-pilot		Assume of 5 sample per mode	
VT Wait			Assume of 5 sample per waiting condition	
Alighting	No part of pre-pilot	Assume 33 % each	0.33	14

4.3.2. Collection method in each segment

A laptop with a touch screen was used in the collection period. The observer collected the data with a prepared Excel VBA form, including a built-in stopwatch. In Appendix A, the excel forms used for the data collection are given. For each segment and moment, a new form was created. Each form had a limit of 30 samples, in order to spread the collection of a segment over different moments, for having a representative population in a segment. The Excel form rounds up the collected time to full seconds. The observer notes the condition of the segment at the top of the collection form before starting the collection. For example, the closure of some vertical transport modes or width restrictions in the corridor or at the platform. The observer starts the timer when a chosen sample passes the start boundary of a segment and stops the time when the sample passes the other boundary. During or after the timing, all categories are filled in by the observer. All boundaries for all segments are in Appendix B.

4.3.3. Moments to collect data

The data collection in Beurs occurred at the moments in Table 4.6. Most of the collection occurred during the afternoon and on weekdays because a more considerable sample variation was expected. On moment was explicitly chosen to include the weekly evening shopping in the city. The pilots helped to estimate the required collection time. Fixed intervals were set to collect samples in each segment to keep the observer focused. At platforms, it was around 30-45 minutes and for other segments around

15-30 minutes. Significantly, the platforms of A/B/C required more time because of reduced metro service during the collection period (RET, 2022a).

Table 4.6: Data collection moments at Beurs station.

Segments	Day	Time	Remark
Pre pilot A/B/C EB	7 October, Friday	14:00-15:00	Evening peak
Pilot 1: A/B/C EB, OV1, OV2, OV3, OV4, VW4	14 October, Friday	14:00-16:30	Afternoon
Pilot 2: A/B/C EB, OV1, OV2, OV3, OV4	18 October, Tuesday	07:00-10:00	Morning peak
A/B/C EB, both D/E , OV1, OV2, OV3, OV4	21 October, Friday	15:00-18:00	Evening peak
A/B/C WB, OV1, WV1, WV2	22 October, Saturday	10:00-12:30	Late morning
D/E NB, OVI2, VO1, VO2, VO3, VO4, VWL	25 October, Tuesday	15:30-18:30	Metro disruption
D/E SB, VWR1, VWR2, VW3, VW4	26 October, Wednesday	10:00-13:00	Late morning
A/B/C EB, OV1, OV2, OV3, OV4, OVI2, OVI3	27 October, Thursday	13:00-16:00	Afternoon
D/E NB, A/B/C EB, VO3, VO4, VWR2, VW4, WV1, WV3	28 October, Friday	18:00-21:00	Evening shopping
D/E SB, OVI3, OV4, VO1, VO2, VO3, WV1, VWL, VWR2	29 October, Saturday	15:00-18:00	Evening peak and Weekend
A/B/C WB, OV1, OV2, OVI3, OV3, VO1, VWL	30 October, Sunday	13:00-16:00	Weekend
OV2, OVI2, WV1, WV2, VW3, VW4, VWR2	1 November, Monday	07:00-09:00	Morning peak
OV2, OVI3, VO3, VW3, VWL, WV2, WV3	2 November, Tuesday	15:30-18:30	Evening peak
OVI3, VO1, VO2, VO4, VW3, WV2	3 November, Wednesday	07:00-10:00	Morning peak
VO1 , VWL	3 November, Wednesday	16:00-17:00	Evening peak

4.3.4. Limitations and assumptions on data collection method

In the data collection method, certain assumptions and limitations are considered. As a transfer path is split up into multiple segments, not all segments are used by transfer passengers only. However, in the segments, all passengers are treated as transfer passengers because of the assumption that there is no difference between in behaviour of transfer passengers and entering/exiting passengers.

At escalators, the samples ”walking time” is the travel time to diverse the escalator when standing still. The assumption is that all passengers stand still on the escalator once boarded. Furthermore, the travel time of an escalator or lift is assumed to be constant throughout the collection period. Therefore, the observer will stop the timing when a sample boards the escalator or the lift doors close. However, the observer checks whether this statement is true for all segment escalators. If a significant group of passengers walks on the escalator during a pilot or collection period, then the assumption is dropped for that escalator.

At platform segments, the time starts for all samples to alight when the doors open of the metro. The result is that some samples' walking time is the waiting time in the vehicle to alight due to crowding or capacity restriction of alighting the vehicle (door width). Not all samples can directly alight a vehicle when the doors open. The reason to use this assumption is to determine the difference between vehicle arrival time and the time for a passenger to leave the platform. To represent an actual alighting distribution of passengers at a platform. If only the walking time were used, from when a sample exits a vehicle to when it leaves the platform, it is not the complete alighting population alighting time. This might lead to a false interpretation of the transfer walking time for many passengers and potentially missed transfers.

One of the main limitations is the transfer passengers population. The considered samples are only at an adult age because of the recommended standard study conditions on walking behaviour from Bosina and Weidmann (2017). Therefore, the elderly, children, and teens are not chosen as samples because those are expected to have a lower share of the transfer population. Moreover, the collected sample could be biased because the observer chooses a sample each time. The observer tries to choose various samples based on the variables, but there could be a predefined bias of the observer only picking certain samples to follow.

Only (official) signed transfer routes in the stations are part of the segments, while some passengers may take a different route to transfer for convenience. Furthermore, the observer only collects samples which walk directly through a segment, only (short) stopping due to crowding, waiting for a lift, orientation or doing a small activity (waiting for group members) is allowed. Otherwise, the sample is discarded when its break in a segment is long.

In the crowding variable, the exact crowding level is out of scope in this collection because of the practicality for the observer to collect data. The observer should determine the crowding easily in a qualitative manner based on the definition of Table 4.4. This leads to the case if only one person is walking in front of the chosen samples, then the condition is noted as crowding. At the same time, the rest of a spatial area could be empty. Different crowding levels could be used in future research on walking times in stations. However, a general remark on the crowding levels is given after the collection period.

4.4. Pre-pilot and pilot data collection

Before the data collection started, certain aspects had to be checked and tested in the pre-pilot and the pilots. All objectives are given in Table 4.7. The objectives are mainly preparation steps for the collection. The observer should get familiar with the station's layout, segments and standing locations. One of the positive results, some segments could be combined because of the excellent visibility for the observer. Only two objectives were essential to start the pilot, the transfer population and whether passengers walked on the escalator. The transfer population is needed to determine the minimum sample sizes for each category outside the walking time.

4.4.1. Pre-pilot

The pre-pilot objective was mainly to know the station layout and to determine the number of segments in each transfer path in Table 4.1. Besides the preparation for the complete collection, a small pre-pilot observation was done at the A/B/C EB platform (see Table 4.6) for around 45 minutes. The pre-pilot was performed as planned in the actual collection set-up, but only the passenger and alighting characteristics were collected. These were used to calculate the sample size per category in Table 4.5.

A general observation was done in the pre-pilot on all vertical transport segments. The share of passengers between the stairs and the escalator was roughly 50 % each, and very few samples took the lift. Therefore, the current same sample size value is used from the luggage variable, which is 5. Furthermore, on the escalators, all samples stood still. Therefore, based on the positive results from that objective, the assumption that all samples remain standing on the escalator is true. Thus, the travel time for each escalator from Appendix C can be used in the corresponding segments with escalators.

Furthermore, the general observations at the vertical transport segment led to the decision to split up some stairs and escalators halfway into different segments because of a safe standing position for the observer and poor overview. The split led to the following assumptions and decisions for the remaining collection. The waiting condition will not be collected in the second part of a stair mode because a sample already uses the mode. Moreover, the vertical transport mode is not collected in the second part because of the assumed similar travel time on the remainder of the stairs and escalator. Lastly, if a lift is also present, then the whole lift travel time is part of the first segment because the lift is easier to include completely.

4.4.2. Pilot data collection

The pilot study at the stations is for testing the data collection setup and collecting the first samples. The main objective is determining the time it takes to collect all required information per sample using the

Excel form. Furthermore, the collection effort is tested to determine if all planned segment boundaries are usable. During the pilot, the observer also detects possible risks that could later hinder the data collection or representation. For example, closed-off sections or broken vertical transport modes.

4.4.3. Pilot objective

The pilot is to study the ease of collecting the required data and determine an improved variation of the samples from the collection in order to adjust the sample size calculations in section 4.3. The pilot's objective is in Table 4.7. To make possible adjustments, the pilot is done twice in each station. For Beurs, the pilot was done on Friday, 14th (13:30-16:00) and Tuesday, 18th of October (07:00-10:00, morning peak). The observer will stand at the platform for 1 hour and at least 15 minutes in each segment. As Beurs has multiple transfer paths, only one was part of the pilot because of the limited time for the pilot. The transfer path A/B/C EB to D/E NB was chosen for the pilot.

Table 4.7: Pre-pilot and Pilot collection objectives.

Part	Objective	Result
	Pre-pilot objectives	
	Observer gets familiar with station lay-out	Neutral
	Determine the number of segments for each transfer route.	Positive
	Determine standing positions for the observer in each segment	Neutral
	Determine travel times of escalators and lifts.	Neutral
	Investigate the transfer passenger population regarding the attributes.	Neutral
	Check whether passengers walk or stand on the escalator.	Positive
	Pilot objectives	
Collection method	Check if collection is possible with 1 observer.	Positive
	Check if the Excel form is usable for collection.	Positive
	Get an indication of collection time for each segment	
Station segments	Check if set segment boundaries are feasible.	Adjustment
Sample size	Collect 5-10 % of the walking time sample size in each segment from Equation 4.4, ($n = 95$).	Positive
	Check if following multiple samples in a segment is possible.	Positive
	Collect at least 5 samples per platform section (front, middle, back).	Positive
Attribute	Check variation of each variable after pilot.	Neutral

4.4.4. Pilot results

The pilot was quite successful according to Table 4.7. The collected data was valid and therefore will be part of the data set. The Excel form had a slight adjustment to collect more samples when a metro arrives at the platform. The segments were usable to do the collection, some could be combined to reduce the number of segments. In each segment, around 30 samples were collected, Appendix C contains all pilot results on the segments. The collection is doable with one person, but it requires significant collection time because samples pass in bundles. In other words, the collection time is mainly waiting for samples to arrive. However, following multiple samples simultaneously is possible, but it depends on the level of crowding.

The shares of samples for all categories were in line with the shares from Table 4.5. Thus, the minimal sample sizes for each category remain the same. On the walking time itself, the type of distribution is unclear. Some show a normal distribution, but the sample size is too low to conclude that. All in all, the minimum sample size of 95 for the walking time is kept.

4.5. Reflection on data collection at Beurs

This section reflects on the main data collection. The reflection includes possible problems or limitations during the collection, which might affect the data analysis. The largest one is a potential bias on sample walking in groups. These were mostly collected in low or moderate levels of crowding because then samples walking in a group were easier to detect. However, in more crowded situations, they were still relatively easy to detect because samples therein walked closer to each other than to passengers around them.

At the D/E Southbound platform, the front metro door stopped (partly) outside the platform segment (and into segment VO1). Therefore, the first alighting samples were not collected because walking time would already be zero. Only samples that alighted a bit later were part of the collection because the timer had already started. Moreover, the phenomenon changed the boundary of stopping the time for segments VO4 and WV3 slightly backwards because a metro part did stop there. Extending the platform segment was not chosen because the escalator is an additional exit point, making the collection harder. On the D/E Northbound platform, all doors did stop on the platform segment. All platforms at Beurs can accommodate metros with 8 cars. However, in the collection period, only 6-car metros run. Therefore, the alighting sections (front/middle/back) are based on these 6-car metros, each section is 2 cars long. The rough location of the alighting sections is given on the Appendix B. In other words, the data analysis for the category alighting is only valid for 6-car metro lengths.

Lastly, the observer unintentionally discarded samples which walked through the complete segment. The samples usually stopped (too long) to orientate (look at the signs and their phones) and find their correct route. These are expected to be unfamiliar with the station. Around half of the time, these samples asked the observer or another staff member if they were on the correct route and continued later on. Therefore, familiarity could play a role in the transfer walking time and should be studied in future studies.

In segment VO3, the vertical transport towards the A/B/C Eastbound platform, only one sample used the lift. Therefore, in the data analysis, the lift will be left out because of its very limited use. In a different vertical transport segment, VWL the lift use was also limited. The adjacent stairs to that lift were never used. Therefore the stairs are left out as a vertical transport mode and a lower sample size was used for the lift because of the low share of passengers using it. Thus only the escalator and lift remain in that segment.

Some passengers were sometimes eager to use the escalator when a bi-directional escalator was present. Especially in segments OV1 and OVI2 because the escalator mainly travelled in the opposite direction. They sometimes waited a significant time before the escalator turned in their preferred direction. These passengers are assumed to transfer indirectly because of the long waiting period. Therefore, these samples could be left out of the data analysis. However, the waiting behaviour for the bidirectional escalators might be the subject of further investigation.

When a sample passed with large luggage, the exact items were noted during the collection. In Beurs, the main large luggage items are strollers and bicycles. In other words, the variation in items is limited. In segment VO3, the observer did notice that samples with these items mainly used the escalator. The reason could be to avoid waiting for the lift, and the items are relatively easy to carry on the escalator. Furthermore, the observer confirmed that most samples prefer the escalator over stairs, as stated in the study of Daamen et al. (2015).

The next metro departure displays in some segments had some influence on the passenger walking behaviour. When the display showed a departure time of zero (indicating that a metro is arriving)or one minute, more passengers tend to walk faster or even run. The reason could be that passengers require a specific metro line to catch their destination, as shown in Figure 4.2, or they are unwilling to wait for the next metro. However, the effect of departure times displayed in metro transfer corridors could be studied in future studies on walking behaviour in stations.

5 Data analysis

The objective of this chapter is to analyse the collected walking times and determine the significant variables for each segment in Beurs station. Followed by combing the results per segment into one transfer walking time distribution for each transfer path and significant variable, which should answer the third research question. The structure of the chapter is as follows: First, in section 5.1 with descriptive statistics, the walking time in each segment is analysed. The type of distribution of the walking time is also part of the section. Next is the statistical test for significance for each variable to the walking time in section 5.2. To generalise the walking time results, the transition to passing speeds is done in section 5.3. After that, in section 5.4, all segments from the transfer route are combined to form the transfer walking time for that specific transfer through a Monte Carlo simulation. The chapter ends with the main data analysis conclusion and the answer to the third sub-question in section 5.5.

5.1. Results from the collected data

First, an overview of the collected sample sizes and the consequences on the data interpretation and results is given. After that, the basic statistics of each segments walking time are shown. The last step is to check if the data is normally distributed.

5.1.1. Number of samples per segment and variable

Except for the waiting condition, each variable had a calculated minimum sample size from section 4.3 to collect. Most minimum sizes were fulfilled in each segment. The exact number of samples for each segment is in Appendix D. However, in half of the segments, the number of samples with large luggage is below the minimum. In section 4.5, the large luggage items were limited to only two: strollers and bicycles. Therefore, the low sample size is acceptable for this variable. The impact of too few samples has consequences for concluding the significance of that variable. The low sample size cannot conclude the significance of that variable in that segment in the next analysis phase. A different approach is to conclude that the metro transfer population in Rotterdam consists mainly of passengers without large luggage.

On the platform segments to the two different exits of A/B/C Eastbound, two alighting variables had too few samples. An explanation for the shortage of samples is that the other exit is nearby for that specific alighting spot. Most passengers used the closest exit on this platform, as expected from the study of Daamen et al. (2015). The few samples that used the further exit are assumed to do that intentionally because that exit has the direct transfer route for their specific D/E platform. Otherwise, the indirect transfer route from Figure 4.3a has to be taken, which the sample might assume to have a longer walking time.

In segments in opposite directions: OV1/VO4 and WV1/VW4, the occurrence of crowding differs. For example, in segment WV1, crowding was present in more than 50 %, while in the opposite direction in segment VW4, crowding only occurred in less than 30 % of the samples. An explanation for the difference is the location of the segments near platform A/B/C Westbound. When a metro arrives, many passengers pass through segment WV1 simultaneously because the platform is directly linked to that segment. Thus, crowding is present. In the other direction, passengers pass at a lower intensity but for a longer period. In other words, there are fewer other passengers around to encounter.

Furthermore, the stairs or escalator use depends on the direction of the transfer path. Most samples took the stairs downwards from the platform (OV1), while upwards (VO3), the escalator, is preferred. The reason is the direction of the bi-direction escalator. In the collection period, most of the time, the escalator was going upwards towards the platform. Therefore descending passengers could only use the stairs (or the lift). A decision could be in further analysis that passengers leaving the A/B/C Eastbound platform only can use the stairs or lift.

At last, the determined shares from the pre-pilot in Table 4.5, to calculate the minimum sample size in each category, are reflected on the actually collected samples. As each segment has around 100 samples, this is an easy comparison to the shares from the pre-pilot. For the category ”gender”, the shares match with around 50% for each variable. The evidence is less strong for the ”luggage” because some segments have up to 20 % of the samples with large luggage, while in most segments, it is between 5-10 %. In the category group size, the share of samples walking in a group was larger in the pilot than in the main collection. In the pre-pilot, 33 % of the samples walked in groups. While from the complete sample size, the share is around 20-25 %. With 20 %, a low minimum sample size would be required. Nevertheless, there are too few samples in a group in only three segments. This will be mentioned in the data analysis result per segment.

5.1.2. Descriptive statistics

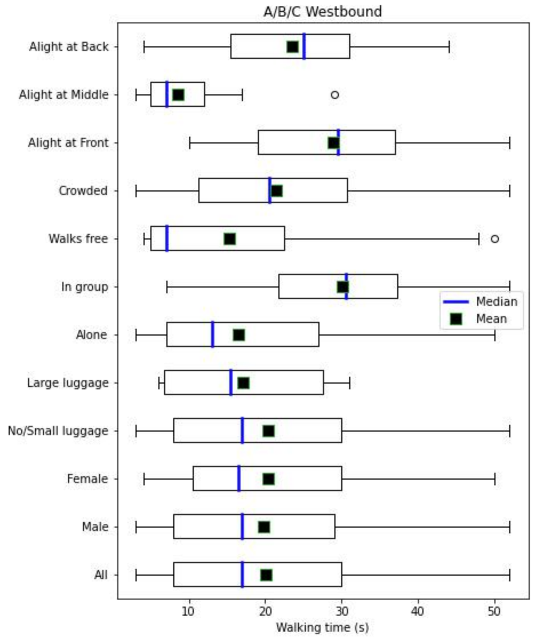
Next are the descriptive statistics on the walking time for all samples of each segment. These values determine the range of walking time. The main values are the sample mean, sample standard deviation, minimal, median and maximal of the walking time. In Table 5.1, all descriptive statistics are given. For almost all segments, the mean value is larger than the median, indicating that some samples had a significantly longer walking time than the rest of the samples in the segment. This is proven by the median difference between minimal and maximal walking time. The maximal walking time is significantly further from the median for all segments, an indication that the collected data is not symmetrical distributed for the segments and thus, the median is more important to analyse than the mean walking time. Furthermore, not all statistics are similar for segments with the same boundaries for both directions of the transfer route (OV1/VO4, OV2/VO3, OV3/VO2 and WV2/VW3). For example, segment OV2 has a standard deviation of 11.83 seconds and a median of 18 seconds, while segment VO3 has values of 5.94 seconds and 22 seconds for the same statistics, respectively. Therefore, walking away from platform A/B/C Eastbound has more variation in the walking time than walking towards it. The explanation could be the level of crowding because roughly 50% of the samples walked in crowded conditions in segment OV2, while only 20 % had crowding around in segment VO3.

Table 5.1: Descriptive statistics of each segment's walking time in seconds.

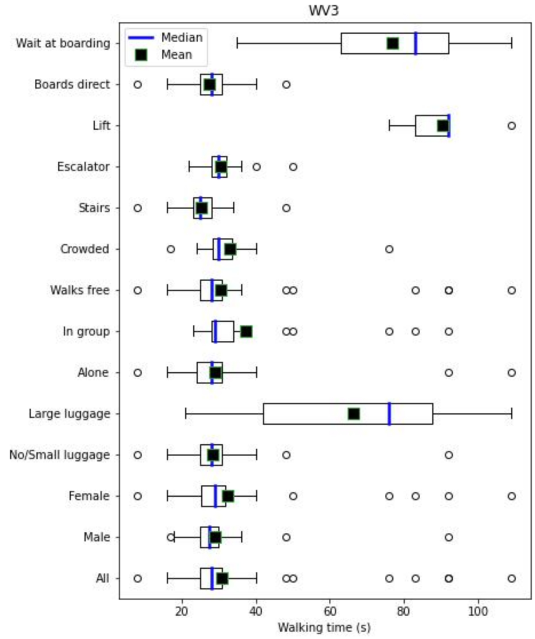
Segment	Mean ($\bar{x}$)	Std (s)	Min	Median	Max
WV1	15.4 s	5.17 s	6 s	15 s	39 s
WV2	12.6 s	5.36 s	6 s	11 s	45 s
WV3	31.0 s	15.01 s	8 s	28 s	109 s
VWR1	24.8 s	2.05 s	20 s	25 s	32 s
VWR2	15.0 s	5.41 s	8 s	13 s	40 s
VWL	75.6 s	26.44 s	39 s	66 s	142 s
VW3	10.1 s	2.86 s	4 s	10 s	26 s
VW4	16.0 s	6.05 s	6 s	15 s	39 s
OV1	23.5 s	27.94 s	4 s	12 s	144 s
OV2	21.5 s	11.83 s	6 s	18 s	86 s
OVI2	29.8 s	13.58 s	10 s	28 s	91 s
OVI3	19.0 s	4.09 s	9 s	19 s	32 s
OV3	29.3 s	6.48 s	11 s	29 s	50 s
OV4	9.50 s	2.62 s	3 s	10 s	22 s
VO1	15.6 s	4.87 s	8 s	15 s	44 s
VO2	29.6 s	6.25 s	8 s	30 s	44 s
VO3	22.0 s	5.94 s	10 s	22 s	40 s
VO4	13.2 s	4.19 s	4 s	13 s	26 s
D/E NB	33.8 s	21.68 s	3 s	33 s	78 s
D/E SB	30.8 s	20.13 s	3 s	28 s	79 s
A/B/C WB	20.1 s	13.00 s	3 s	17 s	52 s
A/B/C EB to eastern exit	18.0 s	11.18 s	3 s	15 s	50 s
A/B/C EB to western exit	25.7 s	17.11 s	3 s	22 s	73 s

To gain the first insight into the walking time per variable, the boxplot is a valuable tool to show possible differences per variable. Here, the boxplot gives the median, mean and range between the 25th % percentile and 75th % percentile of the walking time in that segment for each variable. As mentioned in the previous paragraph, the variable luggage should be analysed carefully due to the few samples available in most segments.

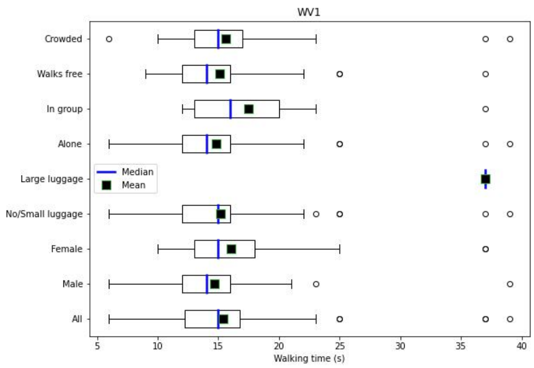
In Figure 5.1, the boxplots of three random segment types are given, one of the platform, one in the corridor and one with vertical transport. In the boxplot from the platform segment, the alighting location shows a large variation in the walking time because the alighting location determines the distance a sample has to walk to leave the platform. Little variation between the variables is within the corridor segment. Only walking in a group has a different boxplot. In the vertical transport segment, the mode choice, waiting condition and large luggage show different walking time intervals compared to the other variables. The boxplot of all other segments are in Appendix F and show similar results.



(a) Boxplot, platform A/B/C Westbound.



(b) Boxplot, segment WV3 with vertical transport.



(c) Boxplot, segment WV1 in a corridor.

Figure 5.1: Walking time boxplots of different segment types.

5.1.3. Test for distribution type of the walking time

The first indication of the type of distribution on the walking time is using the skewness (γ) statistic. The skewness estimates if the data is symmetrically distributed, which a normal distribution has. The previous section's boxplots revealed that most segments do not have a symmetrical distribution. The skewness parameter is calculated in two ways, as shown in Appendix G. One through the method of Field (2013) and the other through the standard Python method Scipy (2022). The conclusion of the expected distribution is given in Table 5.2. Most segments have symmetrical distribution, while for six segments, the skewness results contradict.

Table 5.2: Skewness and expected distribution of walking time.

Segments	Expected distribution	Number of segments
All platforms, OVI3, OV3, OV4, VO2, VO3, VO4, VWR1	Symmetrical	12
OV1, OV2, VWR2, VWL, WV2	Right-tailed	5
OVI2, VO1, VW3, VW4, WV1, WV3	Unsure	6

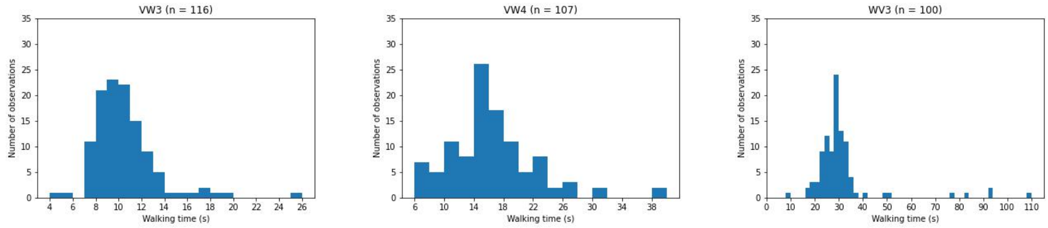
A further aspect to investigate is the platform segments' walking time distribution because the exits are located at different spots for each platform. However, all segments have similar skewness values and thus symmetrical walking time distribution. The histograms of the walking times at the platforms in Appendix E show a similar uniform or slightly triangular distribution. Concluding that exit location does not influence the distribution type for the walking time.

Two tests are done to check if a distribution is normally distributed (Field, 2013) or lognormal distributed. As stated in the chapter 3, the performed test is the Kolmogorov-Shrinov test (KS-test). The null hypothesis (H_0) is that walking time in a segment follows a normal distribution with the mean and standard deviation. The alternative hypothesis (H_1) is that the walking time does not follow a normal distribution. For the lognormal test, the same H_0 & H_1 apply but for the lognormal distribution. The test results per segment are in Appendix G, but the concluding distribution type is presented in Table 5.3. For three segments (VW3, VW4 and WV3), the KS-tests and skewness do not provide clear evidence on the distribution type. Comparing the walking time histograms for each of these segments in Figure 5.2 show that the walking time has normal and lognormal characteristics. The main conclusion is that most segments have a lognormal distributed walking time, as in line with the observations from Du et al. (2009). Therefore, only non-parametric tests can check significant differences in walking time between categories.

Table 5.3: Conclusions from normal and lognormal distribution KS-tests per segment.

Segments	Walking time distribution	Number of segments
D/E NB, OV3, VO3	Normal	3
A/B/C EB Eastern exit, OV2, OVI2, OVI3, WV1, WV2, VO1, VO4, VWL	Lognormal	10
A/B/C EB Western exit, A/B/C WB, D/E SB, OVI3	Both	4
OV1, OV4, WV3, VO2, VWR1, VWR2, VW3, VW4	None	8

The final check is on the difference between off-peak and peak walking times. A two-sample KS test was used to determine this. However, for four segments, intentionally, the walking times were collected during only the peak or off-peak times in Beurs station. The results of that test are in Appendix G, with the conclusion that the walking times are the same for off-peak and peak conditions in the station. Therefore, the statement of the large variance in transfer walking time between train-metro Fujiyama and Cao (2016) does not apply to Beurs station. However, extensive crowding was absent in Beurs station. Therefore, there could be a significant difference in transfer walking time in different metro stations between non-crowded and crowded situations.



(a) Segment VW3. (b) Segment VW4. (c) Segment WV3.

Figure 5.2: Walking time histograms of the unsure expected distribution from Table 5.2 and Table 5.3.

5.2. Significant variables in each transfer segment

This section describes the method to determine the significance of each walking time variable. Firstly, the chosen tests are described, followed by the results and the interpretation of the walking time per segment.

5.2.1. Choice of statistical test: Kruskal-Willis test and Mann-Whitney U test

Based on the result in Appendix G and the skewness from Table G.1, a non-parametric test will be used to test the significance of each variable of the transfer walking time. The main interest is to see differences between all variables and within a category on the walking time distribution. Therefore, the choice is to use the Mann-Whitney U test for categories with two variables and the Kruskal-Willis test with three or more variables (Field, 2013). Both tests require that each group has independent samples, which each variable within a category has.

5.2.2. Significance test for all and per category

The test is performed for each category, with its specific variables. The null hypothesis (H_0) is that all variables have the same walking time distribution, while the alternative hypothesis (H_1) says that the distributions differ between variables. The Mann-Whitney U test is done for all main categories, gender, luggage, group size and crowding. The test results are in Table 5.4, with the corresponding test statistic value and p-value. A significant level of 95 % ($\alpha = 0.05$) is used in all coming tests. When p-value $\leq \alpha$, the H_0 is rejected and the H_1 is accepted. Thus that walking time differs in that category.

The results from the test in Table 5.4 show that gender is insignificant for the walking time for all segments, except in OVI2. However, the decision is to ignore the gender significance in segment OVI2 because it is not part of the direct transfer route from Table 4.1 and as it is the only segment with the significance. The luggage category shows significance between the walking time for each variable in some segments. Nevertheless, segments OVI3, VWR1 and VW4 have an insufficient sample size, making the result unreliable. The remaining (OV1, OV2, WV3) significant segments for luggage include vertical transport. Therefore the vertical transport choice of a sample might correlate with the significance of having no or large luggage. The significance of OV2 might depend on the layout of the segment, as there are a short flight of stairs or a ramp. Samples with large luggage might favour the ramp over the stairs and must walk a longer distance than those taking the stairs.

The most important result is the significant difference between walking alone or in a group for most segments. Only segments with one vertical transport mode (VWR1 and VWL) and a segment between the fare gates (VW3) show little evidence of a difference. As most segments have a significant result, the variables alone and in a group will also be split up in further data analysis on the transfer walking time. Lastly, the level of crowding is only important in five segments, which is explainable by the lack of large crowding in most segments observed in the data collection.

The result of Table 5.4 shows that group size is the main significant category influencing the walking time. Therefore, the exact difference in walking time between walking alone and in a group for each segment is presented in Table 5.5 based on the median of each category. Both in time and as a percentage, the platforms have the biggest walking time difference. In contrast, in the station corridors and vertical transport segments, the difference is significantly lower, up to 41 %. The large variance might result from relatively short segments, as the length varies between 7 m and 42 m. A second difference between the two groups could already present a large percentage difference when the segment is short. However, the most important difference is found in the boxplots in Appendix F for this category. The walking time 25th percentile bound for samples walking in a group is usually the same or larger than the median of the samples walking alone. Therefore, passengers walking in groups require significantly more time in Beurs station. This conclusion and results might not yield for all metro stations because the case focused on only one metro station.

Table 5.4: Significance tests all variables and the main categories for 95% significance ($\alpha = 0.05$). (Green marked cells have p-value ≤ α, * marked have less than 5 samples or the minimal sample size for each category.)

Mann-Whitney U test		Category			
Segment	**Test values**	**Gender**	**Luggage**	**Groupsize**	**Crowding**
A/B/C EB to eastern exit	test statistic	1326	101.5	415.5	1247
	p-value	0.46	0.08*	0.06*	0.95
A/B/C EB to western exit	test statistic	1157.5	452.5	297	1181
	p-value	0.34	0.37	0.01	0.96
A/B/C WB	test statistic	1315.5	427	431.5	601
	p-value	0.7	0.64	0.01	0.01
D/E SB	test statistic	1446.5	551	622	1676
	p-value	0.26	0.85	0.01	0.41
D/E NB	test statistic	1519.5	430	587	1325
	p-value	0.83	0.14	0.01	0.047
OV1	test statistic	1114	85.5	475.5	1268.5
	p-value	0.36	0.01	0.01	0.98
OV2	test statistic	1081.5	373	458.5	1071.5
	p-value	0.28	0.01	0.01	0.22
OVI2	test statistic	766	106.5	578	888
	p-value	0.01	0.01	0.02	0.42
OVI3	test statistic	1082.5	32	472.5	871
	p-value	0.06	0.02*	0.01	0.01
OV3	test statistic	1147	192	711.5	1234
	p-value	0.48	0.35*	0.048	0.99
OV4	test statistic	1051	436	625.5	597.5
	p-value	0.13	0.48	0.01	0.86*
WV1	test statistic	1078.5	1.5	547	1135.5
	p-value	0.14	0.1*	0.01	0.32
WV2	test statistic	1155	98.5	283	787.5
	p-value	0.08	0.89*	0.03*	0.36
WV3	test statistic	1055.5	95.5	596.5	440
	p-value	0.26	0.01	0.01	0.11*
VO1	test statistic	1016.5	158	633	1075.5
	p-value	0.11	0.21	0.04	0.26
VO2	test statistic	1183	148.5	344	1635.5
	p-value	0.07	0.32*	0.01	0.35
VO3	test statistic	1114	32	452.5	566
	p-value	0.19	0.1*	0.01	0.01
VO4	test statistic	1164.5	162.5	423	1160.5
	p-value	0.27	0.84	0.01	0.23
VWR1	test statistic	1093.5	45	693	649
	p-value	0.52	0.04*	0.31	0.01
VWR2	test statistic	1352.5	39.5	611	835
	p-value	0.64	0.14*	0.01	0.24
VWL	test statistic	70	37.5	50.5	16
	p-value	0.93	0.88	0.26	0.56
VW3	test statistic	1626.5	7	566	1275
	p-value	0.86	0.13*	0.21	0.18
VW4	test statistic	1139.5	13	520.5	914.5
	p-value	0.13	0.01*	0.01	0.10

Table 5.5: Median walking time (s) difference between walking alone and in a group per segment.

Segment	Alone	In a group	Difference in time	Difference % alone-in group
ABC EB East exit	14 s	24 s	10 s	71 %
ABC EB West exit	19 s	41 s	22 s	116 %
ABC WB	13 s	31 s	18 s	135 %
DE SB	29 s	47 s	18 s	62 %
DE NB	22 s	40 s	18 s	82 %
OV1	11 s	16 s	4 s	41 %
OV2	17 s	21 s	4 s	24 %
OVI2	27 s	34 s	7 s	26 %
OVI3	18 s	20 s	2 s	11 %
OV3	28 s	31 s	3 s	9 %
OV4	9 s	10 s	1 s	11 %
WV1	14 s	16 s	2 s	14 %
WV2	11 s	14 s	4 s	32 %
WV3	28 s	29 s	1 s	4 %
VO1	14 s	17 s	3 s	21 %
VO2	30 s	34 s	4 s	15 %
VO3	21 s	26 s	4 s	21 %
VO4	12 s	15 s	3 s	25 %
VWR1	25 s	25 s	0 s	0 %
VWR2	13 s	16 s	2 s	19 %
VWL	65 s	68 s	3 s	5 %
VW3	9 s	10 s	2 s	17 %
VW4	14 s	19 s	5 s	36 %

A Krukal-Walllis test is done for the category alighting location for the platform segments because of the three variables present in this category. The test results are in Table 5.6, where all segments significantly differ in the alighting location on the walking time. The result is expected because the alighting and exit location determines the distance to walk for a passenger and, thus, the required walking time. When the exit is further away, the sample has a longer distance to cover and requires a longer walking time.

Table 5.6: Significance test for category alight location for each platform segment for 95% significance ($\alpha = 0.05$). (Green marked cells have p-value ≤ α)

Platform segment	Krukalis-Wallis test	Category: Alight
A/B/C EB to eastern exit	test statistic	24.47
	p-value	0.01
A/B/C EB to western exit	test statistic	49.87
	p-value	0.01
A/B/C WB	test statistic	52.28
	p-value	0.01
D/E SB	test statistic	83.61
	p-value	0.01
D/E NB	test statistic	86.12
	p-value	0.01

Three platforms require an additional Mann-Whitney test because they might have similar characteristics regarding the exit position towards the transfer infrastructure. Especially the D/E platforms (North- and Southbound) have identical layouts. Only the stopping positions of the metro are slightly different. The test compares the complete walking time distribution for each platform. The significance test result in Table 5.7 verifies that the samples from both D/E platforms are equal. Therefore, only one D/E platform will be included in further transfer analysis from the D/E lines. As the metros on the Northbound platform stop slightly further from the desired exit, this platform is used in further data analysis. The test is also performed on the A/B/C Eastbound platform because of the different locations of the exits. For this platform, the walking time is significantly different for each exit. Therefore, each exit is treated separately regarding the walking time for the transfer from this platform.

Table 5.7: Significance test for platforms for 95% significance ($\alpha = 0.05$). (Green marked cells have p-value ≤ α)

Segment	Mann-Whitney U test	All samples
A/B/C EB eastern exit & western exit	test statistic	4177
	p-value	0.01
D/E SB & D/E NB	test statistic	6717
	p-value	0.35

The segments with vertical transport had two additional categories, the vertical transport mode choice and the waiting condition of the sample. The waiting condition is tested through a Krukalis-Wallis test (KW), while the test for the mode choice depends on the number of variables within the category. For segments OV1 and WV3, the KW test is also done because three modes are available. In segments OVI2 and VO3, only two vertical modes are present, resulting in a Mann-Whitney U test (MU). Segment VO3 has 3 modes, but only one sample took the lift, so the mode lift is left out. The resulting values from the test are in Table 5.8 and show that the waiting condition is significant for all these segments. The mode is also significant in the segments with the three vertical transport modes. That result is expected because of the lift, where the time lost to waiting is more likely than for the stairs or escalator, increasing the walking time in that segment. As the KW test does not show which variable (stairs, escalator or lift) is significant, additional post hoc tests are done on the corresponding segments. The vertical transport mode test result is insignificant for segments OVI2 and VO3, which have only stairs and escalators. Herefore, the walking time does not differ between stairs and escalators when only these modes are available and only traverse half a vertical level because of the set segment boundaries. Segments with a complete vertical level might present a different result between stairs and the escalator.

Table 5.8: Significance test for categories vertical transport choice and waiting for 95% significance ($\alpha = 0.05$). (Green marked cells have p-value ≤ α)

Segment	Category	VT mode	Wait by MU Test
OV1 (VT mode is by KW test)	test statistic	59.69	432.5
	p-value	0.01	0.01
OVI2(VT mode is by MU test)	test statistic	0.15	166
	p-value	0.70	0.01
WV3 (VT mode is by KW test)	test statistic	41.24	3
	p-value	0.01	0.01
VO3 (VT mode is by MU test)	test statistic	1.63	21.5
	p-value	0.20	0.01
VWR1 (Only one VT mode present)	test statistic		208.5
	p-value		0.01
VWL (Only one VT mode present)	test statistic		16.5
	p-value		0.01

The significant difference between direct boarding and waiting to board a vertical transport mode in Table 5.8 invites to show the exact differences. The median walking time within this waiting category is presented in Table 5.9. The results support the reflection statement in section 4.5 that Beurs station does not have capacity issues regarding their vertical transport because most medians walking time with waiting is up to 12 seconds longer. Nevertheless, the relatively short segments penalise the waiting to board when only comparing the percentage difference. The only outlier is segment WV3, with an additional 55 seconds. The respective boxplots of the segment walking time in Appendix F and sample size of those waiting in Appendix D show that the waiting time presumably correlates with samples taking the lift and only 7 samples encountered a wait to board. Furthermore, the lift-only segment VWL shows a similar difference between direct and waiting to board. Therefore, waiting to use the lift increases the walking time of those passengers by almost a minute.

Table 5.9: Median walking time (s) difference on waiting condition for boarding vertical transport.

Segment	Direct boarding	Wait to board	Difference in time	Difference in %
OV1	8 s	15 s	7 s	76 %
OVI2	27 s	38 s	12 s	43 %
WV3	28 s	83 s	55 s	196 %
VO3	22 s	32 s	10 s	48 %
VWR1	25 s	28 s	3 s	12 %
VWL	62 s	110 s	48 s	77 %

5.2.3. Post-hoc test for significance per variable

If the Kruskal-Willis rejects the null hypothesis, it remains to be seen which exact variables are significant. Therefore, post-hocs tests are required to see which groups do differ. A popular post hoc test for the Kruskal-Willis test is Dunn's test with Bonferroni correction (Dinno, 2015; Field, 2013) and thus used in the analysis. Again, the null hypothesis (H_0) is that there is no difference between the tested variables and the alternative hypothesis (H_1) is that there is a significant difference between the variables. Dunn's post hoc test results are in Table 5.10. Almost all platforms alighting locations significantly differ in walking time because all p-values are below the significance value. The insignificant result between the front and back at the A/B/C West platform was expected because the exit was in the middle of the platform.

Table 5.10: Dunn's test result on alighting location for 95 % significance ($\alpha = 0.05$).(Green marked values have p-value $\leq \alpha$)

Dunns test p-values	ABC East to eastern exit	ABC East to western exit	ABC West	DE South	DE North
Front - Middle	0.01	0.20	0.001	0.001	0.001
Front - Back	0.001	0.001	0.42	0.001	0.001
Middle - Back	0.003	0.001	0.001	0.001	0.001

The exact differences in alighting location on the walking time depend on the platform layout and exit location, according to Table 5.11. For example, the A/B/C Westbound platform has a wide exit in the middle. Within 7 seconds, half of the passenger share can leave the platform, which alights in the middle. The other alighting locations have a median walking time of 25 seconds, three times longer than the middle alight section. The other platforms have their exit near on end of the platform. The walking time of the respective alighting locations also shows that. Alighting at the front or back has the lowest or highest walking time.

For segments OV1 and WV3, the post hoc test is done to see if specific vertical transport modes are significant compared to the others. At first glance, the medians of each walking time in Table 5.12 show a large difference between the lift and the stair modes. Between the stairs and escalators, the walking time has a smaller difference. However, in segment OV1, the difference between stairs and the escalator in walking time could depend on the availability of the escalator, according to the reflection in section 4.5.

According to the post hoc test results in Table 5.13, all three modes significantly differ in walking time to pass through the segments. Even the between the stairs and the escalator, the walking time is different.

Table 5.11: Median walking time (s) difference between platform alighting locations.

	Alight at			Absolute difference middle-back		Absolute difference middle-front	
Platform segment	Front	Middle	Back	In time (s)	In %	In time (s)	In %
ABC EB East exit	11 s	20 s	43 s	24 s	121 %	8 s	44 %
ABC EB West exit	55 s	33 s	11 s	22 s	67 %	22 s	67 %
ABC WB	30 s	7 s	25 s	18 s	257 %	23 s	321 %
DE SB	57 s	32 s	10 s	22 s	70 %	25 s	78 %
DE NB	10 s	30 s	55 s	25 s	83 %	20 s	67 %

Table 5.12: Median walking time (s) difference between vertical transport modes.

	Vertical transport mode			Difference between stairs-escalator		Difference between stairs-lift	
Segment	Stairs	Escalator	Lift	In time	In %	In time (s)	In %
OV1	10 s	14 s	58 s	4 s	40 %	48 s	480 %
WV3	25 s	30 s	92 s	5 s	20 %	67 s	268 %

These results contradict the result from segments OVI2 and VO3 with only stairs and escalators, where both modes had insignificant results. Only segment WV3 covers the complete stairs and escalator, while segments OV1, OVI2 and VO3 only cover half of their length. But the strong evidence of an insignificant difference in walking time between stairs and escalators cannot be neglected. However, the expectation is that the insignificance might relate to the escalator operational speed, which at Beurs station is around 0.5-0.6 m/s.

Table 5.13: Dunn's test result on vertical transport mode for 95 % significance ($\alpha = 0.05$).(Green marked values have p-value $\leq \alpha$)

Dunns test p-values	Segment OV1	Segment WV3
Stairs - Escalator	0.017	0.001
Stairs - Lift	0.001	0.001
Escalator - Lift	0.018	0.04

5.2.4. Conclusion of significant walking time per segment

The test result shows that certain variables have different walking times in the segments of Beurs. The main categories influencing the walking time are group size, alighting location, vertical transport choice (only for three options) and the waiting condition for vertical transport. The level of crowding and presence of large luggage was in less than 25 % of all segments significant. Luggage was most significant in segments with vertical transport.

5.2.5. Relate results with literature

Within the category of gender, there was an insignificant result, contradicting the findings of (Z. Chen et al., 2016;Bosina and Weidmann, 2017). However, those studies focused on walking speed rather than walking time, which are related to each other. Supported by the majority of insignificance in segments, the conclusion is that gender is unrelated to transfers in metro stations. Nevertheless, the insignificance of crowding in walking time in this study can be related to the observed crowding during the data collection. Large or dense crowding was unobserved, therefore the walking speed roughly remained the same and thus the walking time. If large crowding would be present (Daamen and Hoogendoorn, 2007), then the walking speed would reduce significantly.

The data analysis did confirm the difference in walking time between individuals and passengers in groups from (Bosina and Weidmann, 2017; Z. Chen et al., 2016). The difference in group size related (Bosina and Weidmann, 2017) to the walking time cannot be concluded because most samples walking in groups were duos. for the luggage category, the sample sizes were too small for passengers with large luggage. However, most segments did show an insignificant result between no and large luggage. Nevertheless, most luggage items were only bicycles and strollers. Having different larger items might be a significant difference between no and large luggage. Investigating which large items could influence the walking time could be done in future studies.

5.3. Normalized results (passing speeds in segments)

The walking time per segment must be normalised in order to be used for further studies or advanced data analysis on (transfer) walking behaviour. The general approach would be using the walking speed rather than the walking time when modelling passenger movements in stations. However, in most segments, the observed speeds are not the actual walking speed of passengers because some waiting time could be included in the observed walking times. Especially, segments with vertical transport, platforms or fare gates have a high probability of hidden waiting times inside the walking time. Therefore, the term passing speed is more appropriate to present the normalised results of the walking time. The passing speed is expected to be slightly lower than the actual walking speed.

Therefore, all walking times are translated into passing speeds by dividing the segment's walking distance by the walking time of the samples. However, the walking distance generates issues in segments with multiple routes and their distances and the slight curvature in the metro station. Therefore, the limitation is that the average walking distance is used based on six measurements of possible walking distances in each segment. Significantly, the segments with vertical transport and the fare gates have a larger variation in the covered distance of the samples. In order to compare the walking speed from the literature review and the passing speeds, only the mean and median are checked from the passing speed per segment. The segments are divided along the set segment types from Zhou et al. (2016): corridors, vertical transport points and the platform.

5.3.1. Passing speeds in corridor segments

The segments are categorised as platform, corridor and vertical transport segments because of similar characteristics of the segments. At first, the corridors are discussed, where the passing speeds are expected to be almost identical to the walking speed when no physical boundaries such as fare gates are present. These segment passing speeds can be compared with the set walking speeds of (Bosina and Weidmann, 2017). The boxplot in Figure 5.3 presents the range of observed mean and median for all corridor segments based on the categories from Bosina and Weidmann (2017). The observed passing speeds are usually lower than the set walking speed from Bosina and Weidmann (2017). The most noticeable result is that the 50 percentile of the mean and median are identical within the categories of male, alone and in a group. The 50th percentile average passing speed is higher for females, indicating that a significant group does walk a bit faster. Furthermore, the male and alone have mostly similar boxplot sizes for the mean and median of the corridor segments. Therefore, the difference between males and walking alone on an average of 50th percentile is neglectable.

The setting might explain the difference between the observed and literature walking speed. In a station environment, people might walk slower than outside on the street. The study of (Bosina and Weidmann, 2017) did set a standard walking speed of 1.22 m/s in transport terminals. Compared to the 50th percentile of all medians in Figure 5.3, the same value is obtained. However, the range of passing speed is more critical, which lies between 1.12 and 1.37 m/s, because this can be related to the segment's layout. For all samples and the male and female groups, the mean passing speed is significantly higher than the median passing speed, which is the result that more samples tend to walk faster to catch their desired next metro. Lastly, the boxplot of males and females shows differences in mean and median passing speeds, while gender was insignificant as a category from section 5.2 for the walking time. Therefore, these results contradict each other, and the recommendation is to investigate this further in future studies.

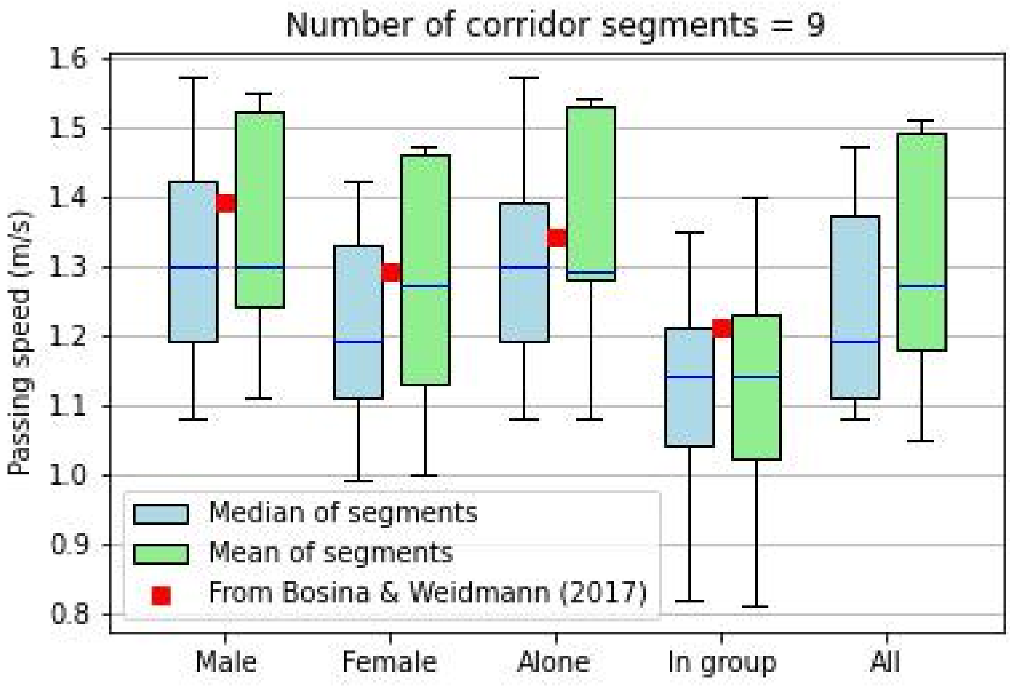


Figure 5.3: Boxplots of observed mean and median passing speeds in corridors.

One variable is introduced which might help to explain the variance in mean and median passing speeds from Figure 5.3, which is the presence of fare gates. From the nine corridor segments, four had fare gates at their boundaries, which can be seen in their layout in Appendix B. The hypothesis is that passengers encounter a slight wait to tap their ticket and the opening of the fare gates, which results in a lower passing speed. The effect of fare gate presence is clearly visible in Figure 5.4, where those with gates have a lower passing speed for all categories. One of the non-fare gates segments also has a relatively low passing speed, but this segment is the last one before the platform ans was used as a waiting zone to board the next metro.

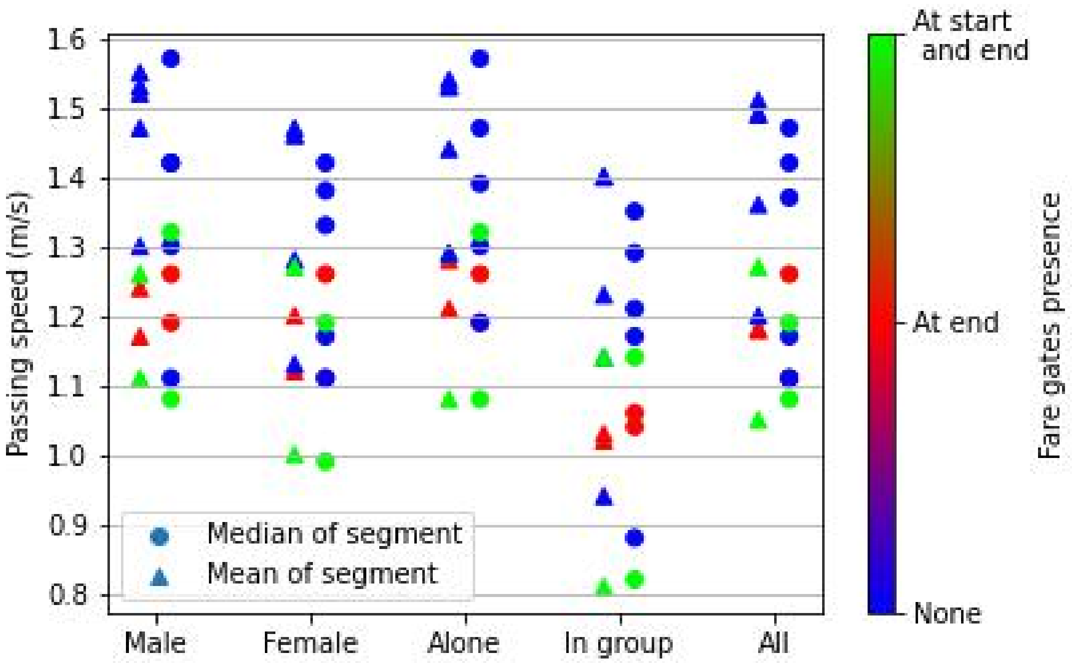


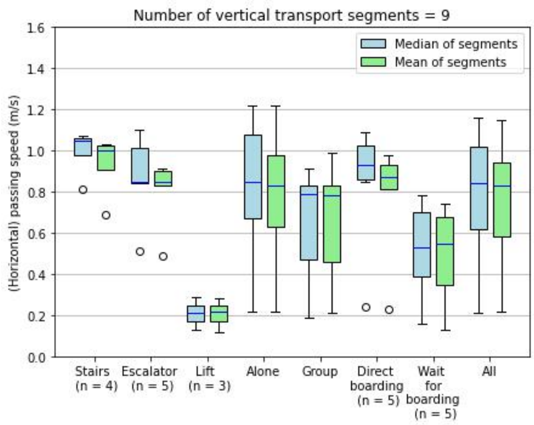
Figure 5.4: Influence of fare gate presence on passing speeds.

5.3.2. Passing speed in segments with vertical transport

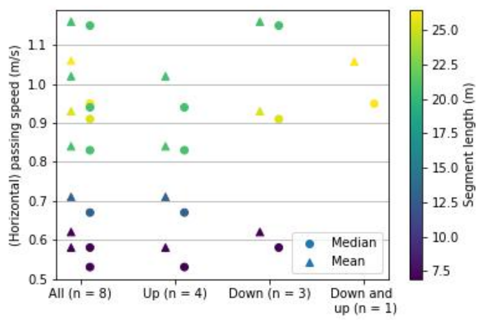
For this category, the limitation is that the passing speed on vertical transport can only be presented as the horizontal passing speed. Moreover, some corridor parts are included due to the layout of these segments. In Figure 5.5a the mean and median variation of the segment with vertical transport are given of the passing speed for the significant categories form section 5.2. Nine segments have vertical transport, but not all include segregation on the exact vertical mode choice because the mode choice was irrelevant to note in the second part of stairs/escalator as explained in section 4.3. In the vertical transport mode, the stairs and escalator have similar passing speeds, where the stairs users are slightly quicker than on the escalator. While for the elevator samples, the passing speed is significantly lower. The 50th percentiles of alone, in group and all sample passing speeds are the same. The only difference is the size of the lower quantile of the in-group median and mean passing speeds.

To determine the passing speed on the stair modes, the length of each segment with stairs and/or escalators is given in Figure 5.5b. The stairs/escalator lengths in the segment are between 3-5 meters. As can be seen, in the few short segments, the median passing speed is between 0.50-0.60 m/s, which could be close to the actual walking speed on stair modes or the operational escalator speed in Beurs station of 0.50 m/s. The passing speeds are higher in the longer segments because of the large share of corridor length. Furthermore, the direction of vertical movement shows a slight difference, where ascending is slower than descending. However, that conclusion lacks strong evidence because of the very few segments.

Lastly, the effect of waiting to board a vertical transport mode has a significant effect on the passing speed in Figure 5.5. All observed passing speed median and means of direct boarding are above the boxplot range of those waiting to board. On average, the passing speed is 1/3 lower when someone has to wait. However, this result is expected to be correlated with the vertical transport mode because, for the lift, longer waiting times were observed compared to the stairs or escalator. The conclusion from the waiting effect is that direct boarding of a stairs mode has a passing speed of 0.90 m/s. When minor queues occur, the passing speed is around 0.60 m/s. For the lift, the passing speed of 0.20 m/s could be used. All these conclusions are valid only when a corridor part is included in the vertical transport segment.



(a) Boxplots of observed mean and median horizontal passing speeds on segments with vertical transport.



(b) Influence of segment length with stairs/escalator on horizontal passing speed.

Figure 5.5: Horizontal passing speed in segments with vertical transport.

5.3.3. Passing speed at the platforms

The passing speed on the platforms will have the largest variance because the exact covered distance of each sample has not been collected. The transition from walking time to passing speed is done through the average length of each platform part. Each platform was divided into three parts for the collection. A platform part corresponds to 1/3 length of the metro rolling stock. So, the assumed length for each sample in a platform part is the distance from the middle of the part to the exit of the platform. Furthermore, the complete range of samples is observed rather than only the mean and median passing speeds because this gives better information on the passing speed variance. The platform part names are changed from the front, middle, and back to the closest, middle and furthest parts to the exit. Moreover, the distance effect of each part will be given to explain the variance in passing speed.

The resulting boxplots with the passing speed at the platforms are in Figure 5.6, where it can be seen that the passing speed variance is the largest in the closest parts and the smallest in the furthest parts. The highest passing speeds in the closest parts are due to the largest error in the actual covered length of the sample and the assumed length for the passing speed. Therefore, the passing speeds in the furthest part are assumed to be the most representable at the platform, which are between 1.10 and 1.20 m/s. These speeds are close to the set walking speed in stations from Bosina and Weidmann (2017). The relation between average length and variance in passing speed is also visible in the boxplots because at the A/B/C westbound platform, the lengths are the smallest, and the variance is the largest. The low passing speeds for the A/B/C Eastbound to the east exit are due to the capacity-restricted segment directly behind it and the short length to the exit. For the other exit with the same segment type behind it, the passing speed are similar to the other platforms.

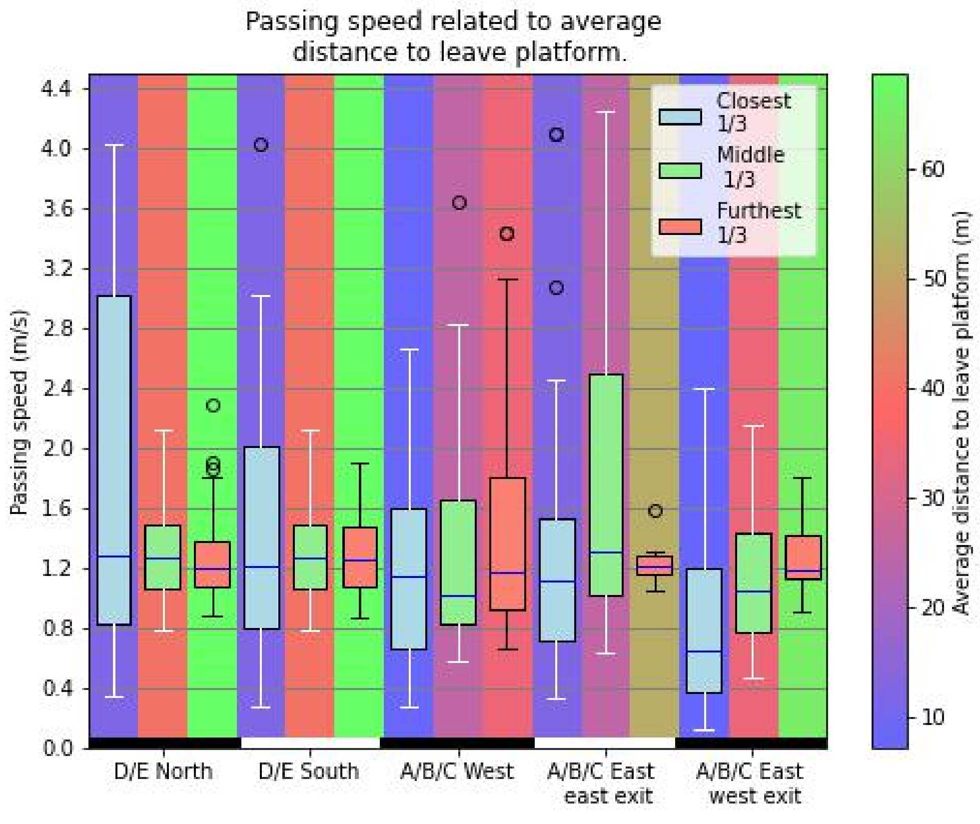


Figure 5.6: Boxplots of observed passing speeds at the platforms, related to the average length of each platform part.

5.3.4. Conclusions on the passing speed

The passing speed is close to or below the actual walking speed of passengers because of the transition from walking time to speed and the hidden waiting time inside it. In the corridors, the passing speeds are the closest to the actual walking speed. Overall, the median passing speed range is between 1.12 and 1.37 m/s. The presence of fare gates reduces the passing speed in corridors, and the difference between males and females is significant, while the walking time analysis presented an insignificant result. The passing speed in the vertical transport segments includes parts with corridor elements. When a passenger can directly board a stair mode, the passing speed is around 0.90 m/s. If a queue is present and the passenger has to wait, the passing speed is around 1/3 lower. The passing speed is around 0.20 m/s for passengers using the lift. Lastly, the passing speed showed the largest variance at the platform because of the error of the exact covered distance per sample, as these were uncollected and averaged. The samples for the furthest alighting are favoured to investigate the passing speed. The passing speed at the platform varies between 1.10 and 1.20 m/s, which is lower than the passing speed in the corridors. The main conclusion is to use different passing speeds for each specific segment type when modelling a metro station.

5.4. Combine segments walking time to a transfer walking time

The previous analysis results are on a segment level. However, research question three requires the most significant factor in the complete transfer walking time. Therefore, the translation from walking time in each segment to the transfer walking time must be done. The Monte Carlo simulation is a tool to do that because of the various walking time distribution on a segment and category level. Firstly, the simulation procedure is discussed, followed by the results from the simulation.

The first step in the Monte Carlo process (Raychaudhuri, 2008) is to generate continuous walking time distributions from the empirical distributions for each segment and significant variable from section 5.2. The advantage of the continuous distribution is that more walking time values can be picked because the empirical distributions might not have values along the complete walking time interval. For the D/E platforms, only the northbound (NB) is considered because the platforms are identical and have no differences in the walking time from section 5.2. The data analysis results from subsection 5.1.3 indicate that most segments have a log-normal walking time distribution. Therefore, in this study, the Monte Carlo simulation assumes that all segments and variables have a log-normal distribution of walking times. In future studies, one might study assuming a different distribution than a lognormal one.

In the second step of the simulation, a random walking time is picked from each distribution at a segment level (Raychaudhuri, 2008). Not only on the overall walking time distribution but also from each significant variable-specific walking time distribution. The significant categories are group size for all segments, vertical transport mode and waiting condition in the vertical transport segment and alighting at the platform. There is no distribution for the lift for one vertical transport segment (OV3) because only one sample was collected for that mode. This random generation of walking times is repeated 1,000 times by the researcher's choice. The main motivation is that each segment had around 100 samples. Performing the simulation 1,000 times ensures that all values on the complete range are picked at least once.

After that, per random pick iteration, the segment walking times are summed up from one of the transfer paths. Each transfer path with the corresponding segments is given in Table 4.1. This gives the simulated transfer walking time per iteration, on an overall level and per significant variable. However, the vertical transport segment and platform-specific significant variables are only present in those segments. Therefore, the overall walking time value is added from the remaining segments without those specific variables in a transfer path. The final simulation output is a distribution with the 1,000 estimated transfer walking times.

5.4.1. General results of Monte Carlo simulations

Firstly, the overall transfer walking times are discussed for all transfer paths in Beurs station. In Figure 5.7, the simulation's cumulative transfer walking time distributions are given. The choice of the indirect or direct route from A/B/C EB to D/E has a difference of around 30 seconds. Furthermore, the difference in the exit choice at the A/B/C EB platforms seems to be neglectable. Around 80 % of the cases with the direct route completed the transfer walk in 120 seconds. In the opposite direction (D/E to A/B/C East), at least 140 seconds is required for the same share. The quickest transfer walk is from A/B/C West to one of the D/E platforms, ranging from 45 seconds as a minimum walking time and up to 80 % of the passengers are assumed to require 90 seconds or less. However, the longest transfer walking times are required for the other direction (D/E platform to A/B/C WB) with the use of the lift, where at least a full minute (60 seconds) is required and three minutes (180 seconds) for most simulated transfers. All in all, opposite transfer paths have different transfer walking time distributions.

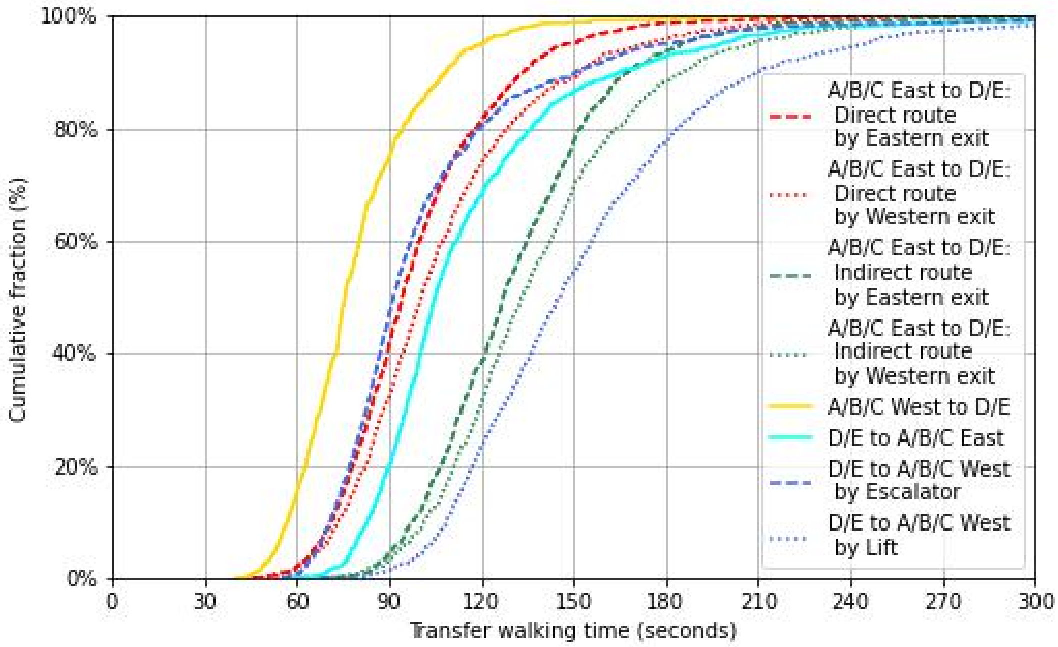
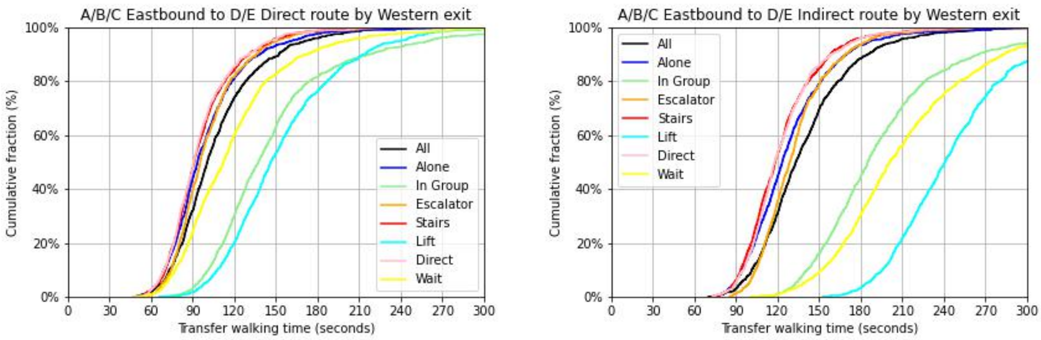


Figure 5.7: Simulated cumulative overall transfer walking time distributions for Beurs station.

5.4.2. Impact variables on transfer walking time

Nonetheless, the most important to discuss is the results for the significant variables on the transfer walking time. Firstly, an overview of the most considerable difference is given, resulting from choosing the direct or indirect route for the transfer between A/B/C East and D/E. In Figure 5.8, the number of vertical transport points in a path significantly impacts the transfer walking time. Especially with the lift on the direct route, the median lies around 140 seconds. While for the indirect route, with two lifts, the median is 240 seconds. A lift also influences the waiting variable because, on the direct route, the transfer walking time with waiting is slightly longer than the overall transfer walking time. Nevertheless, on the indirect route, the walking time distribution with the lift is at least 60 seconds behind the overall distribution.



(a) Direct route (1 vertical transport point)

(b) Indirect route (3 vertical transport points)

Figure 5.8: Monte Carlo simulation transfer walking time A/B/C East to D/E by the Western exit.

Furthermore, the distributions in Figure 5.8 indicate that the difference between direct boarding, using the stairs or escalator and walking alone is minimal. All are roughly within 10 seconds of each other. The use of the escalator is only slightly longer. Similar results are found in the different transfer paths in Appendix H. Walking alone or direct boarding has the lowest transfer walking time values in most transfer paths. Waiting for boarding vertical transport or walking in a group has the most considerable transfer walking time.

The 80th percentile values are compared to better see the differences between each variable and the overall transfer walking time. The choice to check the 80th percentile is because of the lognormal distribution shape of the walking times. The remaining 20 per cent of values are assumed to be extreme outliers. Table 5.14 confirms the findings from Appendix H that walking alone, using the stairs or escalator, or having direct boarding has the quickest transfer walking times. Walking alone has around 6 % faster transfers and using the stairs around 10 %. Nevertheless, passengers walking in groups require at least 25 % more time to complete the transfer. Using the lift has the most negative effect on the transfer walking time; at least 46 % is the walking time longer. However, the impact of multiples lift shows contradicting results because, for the transfer routes with two lifts in the transfer path, the simulated median is over 80 % longer. While one of the transfer routes with one lift also has an increase of around 80 %.

Table 5.14: Transfer walking time median per variable compared to the overall median per transfer path.

Transfer path	Alone	In Group	Escalator	Stairs	Lift	Direct boarding	Wait for boarding
A/B/C EB to D/E: Direct route by Eastern exit	-8%	+32%	-11%	-14%	+47%	-15%	+14%
A/B/C EB to D/E: Direct route by Western exit	-6%	+38%	-9%	-13%	+46%	-13%	+13%
A/B/C EB to D/E: Indirect route by Eastern exit	-7%	+39%	-7%	-13%	+78%	-14%	+60%
A/B/C EB to D/E: Indirect route by Western exit	-8%	+38%	-8%	-13%	+71%	-13%	+53%
A/B/C WB to D/E	-10%	+31%	-3%	-9%	+64%	-6%	+63%
D/E (NB) to A/B/C EB	-3%	+18%	+1%	-1%	-	0%	+11%
D/E (NB) to A/B/C WB by Escalator	-4%	+20%	-	-	-	0%	+3%
D/E (NB) to A/B/C WB by Lift	-5%	+15%	-	-	-	-12%	+14%

Nevertheless, the variable alighting location has been left out to compare with the other variables because this variable is mainly related to the additional walking distance. Implicitly, it says how the platform layout regarding the exits affects the transfer walking time. Therefore, the alighting variable is studied separately. The overall result is a transfer walking time difference of up to 30 %, according to Table 5.15, mainly for the D/E Northbound, where the exit is located at the far back. When the exit is more in the middle, the effect of the platform alighting location becomes smaller. Furthermore, the length of the transfer path plays also affects the difference. The differences between the direct and indirect routes from A/B/C East are slightly less for the longer indirect routes.

Table 5.15: Transfer walking time 80th percentiles compared between alighting location and overall.

Platform / Alighting location	Front	Middle	Back
A/B/C EB by Eastern exit: Direct route	-8% *(Closest)*	-1%	+18% *(Furthest)*
A/B/C EB by Western exit: Direct route	+19% *(Furthest)*	+5%	-12% *(Closest)*
A/B/C EB by Eastern exit: Indirect route	-4%	+1%	+14%
A/B/C EB by Western exit: Indirect route	+15%	-4%	-9%
A/B/C WB to D/E	+10% *(Furthest)*	-17% *(Closest)*	+4%
D/E (NB) to A/B/C East	+12% *(Furthest)*	-6%	-24% *(Closest)*
D/E (NB) to A/B/C West by Escalator	+14%	-7%	-28%
D/E (NB) to A/B/C West by Lift	+9%	-5%	-17%

The last comparison between the variables in the Monte Carlo simulation is the difference within each category. For this purpose, the median and 80th percentile values are given in Table 5.16. Between walking alone and in a group, the difference is at least 20 % and could be up to 50 %. In the vertical transport modes, the most significant fluctuations occur. With the lift, most passengers might require at least 50 % more time to transfer in metro stations. When a transfer already has two lifts in the route, the transfer walking time is expected to be double as long. The reason is the lost time waiting for the lift to arrive. The difference could be neglectable for the waiting condition before a vertical transport mode, or the walking time is at least 70 % longer when multiple vertical transport segments are encountered. For the alighting location of passengers, the effect is the largest when the transfer is relatively short. Such as D/E to A/B/C WB by escalator, between the closest and furthest part, the difference is up to 58%. If the transfer is longer, the alighting location becomes smaller, as seen for the indirect A/B/C EB to D/E path, to around 30%.

These results can be compared with the study of Bosina and Weidmann (2017) on walking speed. The walking speed difference is only 10%-20% between walking alone and in a group. While the walking times in Table 5.16 are at least 20% different. Therefore, one could conclude that groups are even slower in station environments.

Table 5.16: Difference in transfer walking time with 50th and 80th percentile within categories.

/ Variables Transfer path		Group -Alone	Escalator -Stairs	Lift -Stairs	Lift -Escalator	Wait -Direct	Front -Middle	Back -Middle	Front -Back
Base value		Alone	Stairs	Stairs	Escalator	Direct	Middle	Middle	Back
A/B/C EB to D/E: Direct route by Eastern Exit	$q_n(0.50)$	+34%	+5%	+63%	+56%	+22%	-7%	+25%	-25%
	$q_n(0.80)$	+43%	+4%	+72%	+66%	+34%	-7%	+19%	-22%
A/B/C EB to D/E: Direct route by Western Exit	$q_n(0.50)$	+46%	+6%	+63%	+55%	+23%	+15%	-19%	+44%
	$q_n(0.80)$	+47%	+4%	+67%	+61%	+30%	+11%	-16%	+35%
A/B/C EB to D/E: Indirect route by Eastern Exit	$q_n(0.50)$	+41%	+9%	+109%	+92%	+70%	-5%	+18%	-19%
	$q_n(0.80)$	+49%	+7%	+105%	+91%	+87%	-5%	+14%	-16%
A/B/C EB to D/E: Indirect route by Western Exit	$q_n(0.50)$	+50%	+9%	+104%	+87%	+70%	+13%	-14%	+32%
	$q_n(0.80)$	+50%	+6%	+97%	+85%	+77%	+11%	-12%	-26%
A/B/C WB to D/E	$q_n(0.80)$	+50%	+7%	+94%	+81%	+67%	+31%	+23%	+6%
	$q_n(0.80)$	+45%	+6%	+79%	+69%	+74%	+32%	+26%	+5%
D/E to A/B/C EB	$q_n(0.50)$	+29%	+1%	-	-	+11%	+21%	-20%	+51%
	$q_n(0.80)$	+22%	+1%			+7%	+19%	-20%	+48%
D/E to A/B/C WB by Escalator	$q_n(0.50)$	+33%	-	-	-	+2%	+26%	-22%	+61%
	$q_n(0.80)$	+25%				+3%	+23%	-23%	+58%
D/E to A/B/C WB by Lift	$q_n(0.50)$	+26%	-	-	-	+30%	+16%	-15%	+37%
	$q_n(0.80)$	+23%				+30%	+14%	-13%	+31%

5.5. Conclusions from data analysis on the transfer walking time

At a segment level of a transfer path, the attributes within the categories of group size, vertical transport mode choice, the waiting condition to board a vertical transport mode, and the alighting location at the platform have significantly different walking times. However, the vertical transport mode is insignificant in the walking time when only stairs and escalators are present. Passengers walking in a group require 10-40 % more time to pass through a segment than passengers walking alone. At the platform, the difference is larger, ranging between 60-140 %. The vertical mode choice affects the walking time, especially between the stair and lift modes. Passengers who depend on the lift require, on average, 45-70 seconds more time to clear a segment with vertical transport. The statement that crowding or long waiting times to board a vertical transport is absent in Beurs station is reflected by the average waiting time. Except for the lift, the time difference between direct boarding and waiting to board was around 10 seconds, which is almost neglectable. The last significant category, the platform alighting location, showed that for most platform samples, the difference between leaving the metro in the middle and front/back is around 20 seconds. However, that statement is only valid for platforms with one exit or considering only one exit, as was done A/B/C Eastbound, while there are two transfer-specific exits. The exit location has a minor influence on these differences.

Furthermore, most segments had lognormal distributed walking times, as expected from the research conclusions of Du et al. (2009). Translating the walking times into passing speed showed these are mostly lower than the actual walking speed. The presence of fare gates has shown an important reduction in the passing speed and should be included as a variable in the walking time model. The vertical transport segments include some horizontal elements. Thus the actual passing speed on a vertical transport is even lower than the presented passing speeds. The passing speed is significantly different at the platform because of the average distance used per platform segment. The passing speed of the furthest parts to the exit is the most representative for all passengers walking on the platform.

The Monte Carlo simulation of the complete transfer walking time highlighted the significant differences in the walking time. Between walking alone and in a group, those walking in a group require 20-50 % more time to complete the transfer. The waiting condition for vertical transport showed even a large range of the transfer walking time, between 2-87% more time for the passenger who always encounters waiting to board the vertical transport mode. The alighting location of the passenger affects the transfer walking time between 10-60 %, depending on the exit location at the platform. However, the most significant difference is between the stairs/escalator and the lift. With only one lift present, the stair mode users are already 55% quicker than the lift users. The difference ranges between 55% and 94%. If a transfer path has two lifts, the transfer walking time is around twice as long for the lift users. These conclusions are the answer to the third sub-question.

Based on these findings, the suggestion is to determine the transfer walking time for the following categories besides the general transfer walking time: passengers depending on the lift, walking in a group and whether waiting to board a vertical transport is present in the busier periods. Especially, the passengers relying on the life might take a metro later after the transfer compared to the other passengers as their transfer walking time could be double as long. The effect of the alighting location on the transfer walking time is smaller, and the recommendation is only to consider when the platform exit is at the front or back of the platform. Then, two transfer walking times could be assumed, one for the passengers alighting close to the exit and one for those alighting far from the exit.

6

Modelling the transfer walking time

This chapter presents the models for the transfer walking time and walking speed in a (metro) station with attributes from the data analysis to answer the fourth sub-question. The model is based on the collected data from station Beurs. The objective is to generate a walking time model which captures the significant walking time attributes. The chapter describes the model setup and assumptions in section 6.1. Followed by the generation of the model from the data. Lastly, the results are tested and validated. This process is separated for each segment type. Firstly, for corridors in section 6.2, followed by vertical transport segments section 6.3 and finally, the platforms in section 6.4. The chapter ends with the conclusion of the models and the answer to the fourth sub-question in section 6.5.

6.1. Model setup motivation

The model setup is based on previous literature regarding walking time and walking speed modelling and the results from the data analysis of station Beurs. Previous studies of Z. Chen et al. (2016) and Du et al. (2009) used a linear estimation of the walking time or speed in metro stations. Primarily, the model of Z. Chen et al. (2016) uses variables which were insignificant in the data analysis of section 5.2 in this study. The walking time can be modelled because of the successful implementation in the model of Zhou et al. (2016). In combination with the various nominal categories collected per sample in section 4.2, a multiple linear regression model is chosen in this study. The underlying equation to fit the walking time WT is given in Equation 6.1, and the βs are the constants to estimate in the multiple linear regression.

$$WT = \beta_{const} + \beta_1 X_1 + \beta_2 X_2 + ... \tag{6.1}$$

However, walking speed is more relevant in studies researching walking behaviour because of the normalisation and practicality when modelling passenger behaviour in stations. Nevertheless, the walking speed cannot be modelled with the collected walking time because multiple waiting aspects were part of the observed walking times as discussed in section 5.3. Especially at the platforms and the vertical transport segments, the probability of waiting to continue walking is high. Therefore, most observed speeds in the segments are passengers' passing speeds rather than actual walking speeds. The passing speed ps can be derived based on the observed walking time WT_o and the segment-specific length (L) as shown in Equation 6.2.

$$ps = \frac{L}{WT_o} \tag{6.2}$$

In section 5.3, the passing speed significantly differs between the three segment types: corridor, vertical transport and platform segment. Therefore, the chosen approach is to model the walking time separately for each segment type. The same method was applied in the study of Zhou et al. (2016), modelling certain segment types apart. However, that model only included escalators as a vertical transport mode. While in this study, multiple vertical transport modes can be modelled.

The studied literature and data analysis results derive the model setup. A multiple linear estimation model with an ordinal least squares (Field, 2013) fit will model the walking time in two ways. One is by making a walking time model, and the other is the generalised one through the passing speed. From the data analysis results on passing speeds, the modelling is separated into three distinctive segment layouts: the corridor, the vertical transport and the platform segments. With the separation per segment type, the estimated walking time is only representable per segment and not for the complete transfer walking time. Therefore, comparing the Monte Carlo simulation results from section 5.4 is impossible.

The modelling approach is the cross-validation method from Picard and Cook (1984), where the segments are split into a training and validated segments. The training segments estimate the multiple linear regression fit parameters on the walking time and passing speed. The validation segments will test the accurate predictive power of each model. Most of the validation segments are identical to one of the training segments and, thus, fulfil the cross-validation method.

6.1.1. Model development setup

The model generation, which is the same for each segment type, is done in five steps and presented in Figure 6.1. Firstly, the walking time model is derived from the collected categories from the data collection in step 1. Both significant and insignificant categories from the data analysis in section 5.2 are included because the model might indicate different significant categories compared to the ones from the data analysis. The length is also added as a variable because the walking time mainly depends on it and is different for each segment. Moreover, the outlier variable is part of the first model step to include the outlier from the walking time distribution. Step 2 in Figure 6.1 comprises mainly layout-specific variables, which follow from the passing speed data analysis or generic segment layout, and describe their effect on walking time.

With the choice of ordinal least square model, the risks of overfitting are present (Field, 2013). Therefore, the correlation matrix is checked for a relationship between variables. The third step consists of the backward elimination of highly correlated or insignificant variables in the model. The backwards elimination removes each insignificant or correlated variable one by one until a walking time model remains with only significant attributes (Tranmer et al., 2020). The passing speed model is estimated with the significant attributes from the walking time model in the fourth step, except the length variable. If insignificant variables exist, backward elimination is also applied as the fifth step in the passing speed model. The result is a walking time and passing speed model per segment type.

6.1.2. Model validation setup

The validation of the models per segment type is split into a quantitative and qualitative test with the validation segments. The quantitative or goodness of fit test is done by computing the root mean square error (RMSE) and mean absolute error (MAE) from the estimated walking time from both models. Both values are calculated because each has strengths and weaknesses (Hodson, 2022). The RMSE is a commonly used error to report. At the same time, the MAE is presented because the errors are expected to be small because of the minor variance observed in walking times per segment in subsection 5.1.3 and has a smaller penalty on outliers. The outliers in walking time are essential to cover in the model because the outliers could be specific passenger characteristics. A model is valid when the difference between the MAE and RMSE is small. Whether the walking time or passing speed model is better depends on the differences between the RMSE and MAE of each model. If the RMSE and MAE are similar, then both models are usable.

The qualitative check compares the cumulative distributions between the observed and estimated walking times. The model is assumed valid when the predicted walking time distribution for all combinations of variables is within or close to the interval from the observed walking times. If the qualitative and quantitative validation gives contradicting conclusions, then the verification is done on the training segments.

Based on the validation results, the conclusion is drawn if one or both of the models is used to predict walking times in metro transfer stations. A model is valid when both the RMSE and MAE are small, and the cumulative plot from the model estimated model lies within the observed range of walking times per segment. The remaining models are compared to those found in the literature and present their usage limitations, which are the basis for the answer to the last sub-question.

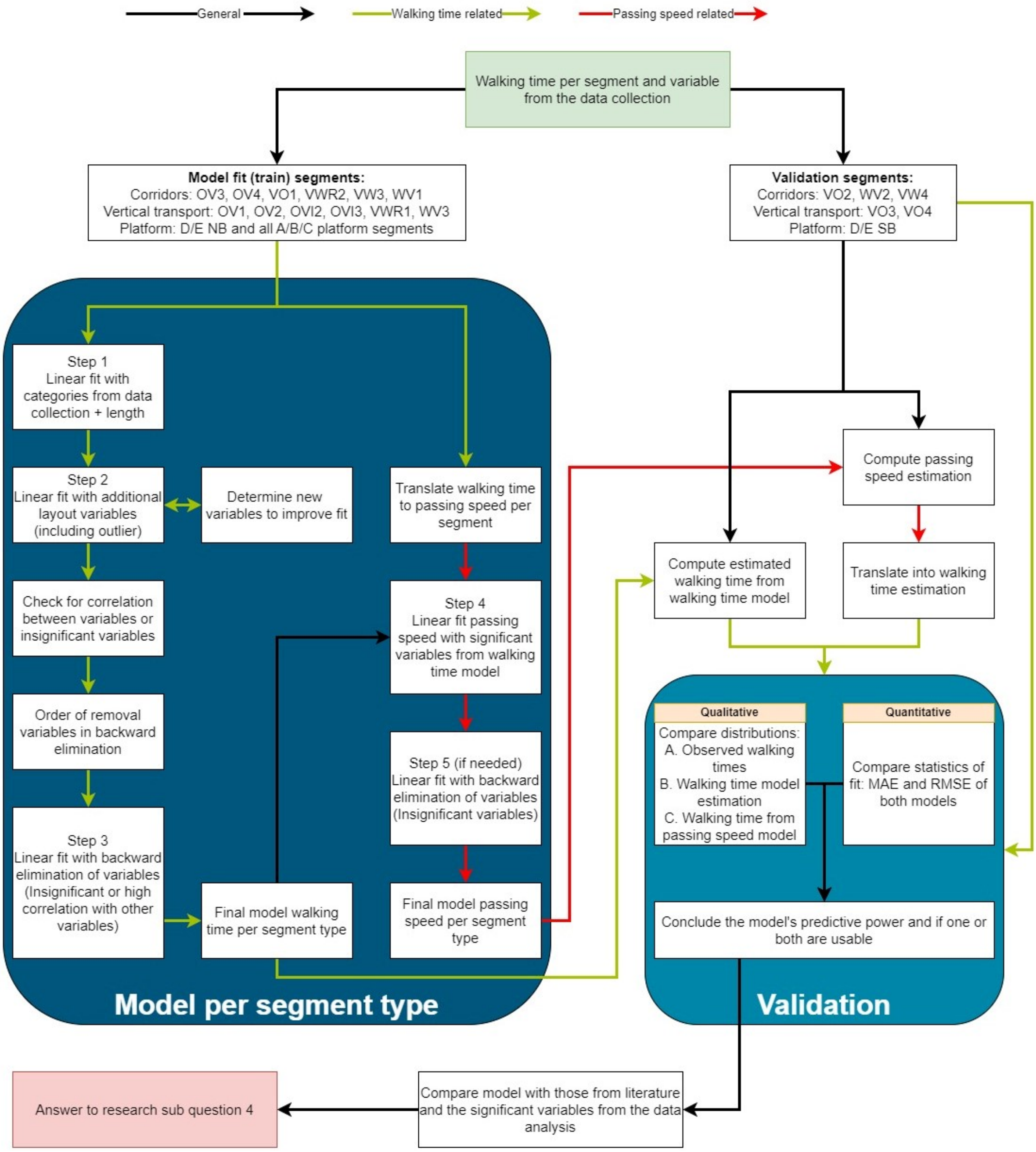


Figure 6.1: Modelling and validation framework of the walking time.

6.1.3. Model variables

In Table 6.1, the main categories and the related variable for each segment type model are given. Besides the length (L) and outlier (X_O), all variables were part of the data collection per segment. The platform and vertical transport-specific variables are presented in their model generation section. Each model should predict the walking time or passing speed for following the base case as in line with the model of Z. Chen et al. (2016); a male walking alone without luggage and without any crowding. The dummy variable for each data collection category should predict the walking for the different states of a sample. One new variable is introduced: the outlier (X_O). This variable should provide a minimum and maximum besides the estimated mean of the walking time because the boxplots in Appendix F of each segment and variable show a distribution of walking times. Therefore, the model can generate a lower, upper and mean walking time value per combination of variables. The base state of this variable is the mean walking time. Quicker walkers will have a negative state because their walking time will be shorter, while slower walkers will have a positive state as their walking time is longer. The boundaries of the outlier are set at 20 per cent of the quicker or slower samples by the researcher's choice. For the passing speed model, only 20 per cent of the quick walkers are used for the outlier state. Otherwise, if the 20 per cent slower walker were included, the passing speed model would only have the outlier as a significant variable. The model aims to include any related passenger or segment layout influence on the passing speed. Thus, the outlier can only predict the upper bound of the passing speed in the station segments. Furthermore, the passing speed has a significantly smaller variance than the walking time. Therefore, the fitted betas values are smaller and assumed to have a higher correlation between them.

The model significant variables will be compared in section 6.5 with those from the data analysis section 5.2. Dummy variables represent a particular category state of a data sample. Variable L represents the average length of the segment, mainly the distance when a sample walks in the middle of a segment. However, the observed segment lengths are relatively short, up to 50 m for the corridor and vertical transport segments and up to 90 m for the platform segments. Therefore, the walking time model is assumed to be only accurate for modelling short segments. This is due to the choice of linear regression model, where each category besides the length adds a fixed number of seconds, independent of the length. From these main categories, length and group size are the starting variables in each walking time model, besides the other significant variables from each specific segment type.

Table 6.1: Standard fitting variables in each walking time model from the data collection besides the length in step 1 and the outliers in Step 2.

Data collection category	**Description**	**State (Value per state)**	**Variable Name**	**Type Variable**	**Constant to linear fit**
Gender	Defines gender	Male (0) Female (1)	X_G	Dummy	β_G
Luggage	Luggage size	No/Small (0) Large (1)	X_{LU}	Dummy	β_{LU}
Group size	Defines group size	Alone (0) In Group (1)	X_{GR}	Dummy	β_{GR}
Crowding	Defines crowding condition	Free (0) Crowded (1)	X_C	Dummy	β_C
Layout variables					
Length	Horizontal walking distance in segment	-	L	Interval	β_L
Linear constant					
Constant	Intersect walking time constant		None		β_{const}
Walking distribution related variable (in step 2)					
Outlier	Quick or slower passenger	< 20th percentile walking time: Fast (-1) Between 20-80th percentile: Average (0) > 80th percentile walking time: Slow (1)	X_O	Categorical	β_O

The models (the walking time model and the passing speed model) are multiple linear regression models and were generated with Python, as stated in section 3.4. The remainder of the chapter explains the model generation along the way of Figure 6.1 for each segment type. The order of segment types is along the complexity of each segment type. The first segment type is the corridor in section 6.2, which are the easiest to model because of the similarity in segment layouts. The next segment type is the vertical transport ones in section 6.3, which have the most complex layout variables to include. Finally, the platforms are modelled in section 6.4, which has the difficulty of having a large variance in walking distances, and thus the passing speeds as explained in section 5.3.

6.2. Walking time model corridors

Nine segments in Beurs are categorised as corridor segments. The model is trained with six segments and validated with three segments. These tree segments have the exact layout as the ones to the model but are in the opposite direction of walking. The limitation is that all segments are relatively on the short side. Therefore, the corridor model is assumed to predict accurate walking times for short corridors.

Three additional variables are introduced, which describe the corridor layout in Table 6.2 for the second model step in Figure 6.1. The fare gates (X_{FG}) showed a lower passing speed in section 5.3 and increased walking time. The first segment variable is introduced because, during the data collection, slower walking behaviour was seen in the segments just before the platform. In contrast, the opposite direction segment showed relatively quicker walking behaviour. The corridor width is used because of the presence in the model of Zhou et al. (2016) and the assumption that narrower corridors increase the probability of congestion and thus longer walking times. The threshold width is that of the A/B/C platforms, which is 3.5 meters.

Table 6.2: Additional layout variables to fit for the walking time corridor model in Step 2.

Data collection category	**Description**	**State**	**Variable Name**	**Type Variable**	**Constant to linear fit**
Fare gates	Presence of fare gates in segment	None Present	X_{FG}	Dummy	β_{FG}
Width	Width of segment	< 3.5 m ≥ 3.5 m	X_{WI}		β_{WI}
Last segment	Segment directly before the platform.	None Before platform or None	X_{LG}	Dummy	β_{LS}

The linear estimation fit results for the corridor segments are given in Table 6.3. The fit with only the categories from the data collection was expected to be relatively low, which was true according to the adjusted R^2. However, the adjusted R^2 values are misleading because it presents the fit to the mean from all walking times in the model. In contrast, the prediction related to a segment's walking time distribution is information to obtain. Therefore, the fit of cross-validation with the remainder segment better indicates the model performance. The correlation test between steps 2 and 3 only showed a negative correlation between ”lastsegment” and the length. The reason is that only two corridor segments are directly before the platform, and one is used to train the model and is relatively short. The other segment is part of the validation process. Now, the backward elimination will remove insignificant attributes step by step.

Table 6.3: Walking time model development for the corridor segments.

Walking time corridor segments	$F(df_{res}, df_m)$	F-statistic	p	Adjusted R^2
Step 1: Data collection variables only + length	(616, 5)	235.0	< 0.001	0.65
Step 2: Additional layout variables + outlier	(612, 9)	388.3	< 0.001	0.85
Step 3: Additional layout variables & backward elimination.	(615, 6)	582.9	< 0.001	0.85

The combined linear estimation of all variables gave the following constants in Table 6.4. As with the data analysis, the group size is significant, but the effect in the model is almost negligible with less than a second. The positive sign from the standard variables in Table I.2 was expected because the model base case is always someone walking alone and without luggage. In most segments, the presence of luggage and walking in a group had a longer walking time as shown in Appendix F. However, the beta value for X_{GR} was expected to be at least double because most corridor segments in Table 5.5 had a difference of 2 seconds between walking alone and in a group. The sign for beta X_{WI} indicates that a wider corridor increases the walking time. However, a negative beta value would be more logical because passengers have more difficulty overtaking others in narrower corridors, and therefore, the walking time would increase compared to wider corridors. The variable first segment X_{LS} shows a minor influence when a segment is directly behind a platform; passengers leaving the platform are estimated to walk slightly quicker, and passengers walking towards the platform slower. The last case was indeed observed in some corridor segments in section 4.5, while the first case has not. A possible explanation for the first case, where passengers walk slightly quicker after the platform, cannot be given.

The negative constant indicates that the model only works from a certain length; the shortest corridor segment is around 10 meters. Comparing the Student t values, the length and outlier impact the model the most and has the largest beta value. This result is expected because the value points out a certain upper and lower limit of the walking time per combination of categories. Lastly, the largest influence besides the outlier variable is the fare gates' presence X_{FG}. The fare gates constant of around 3.8 seconds is assumed to be accurate to the actual time lost when passing a fare gate. However, the exact time loss at the fare gates could be analysed in further studies. The final walking time model for corridor segments is given in Equation 6.3 based on the coefficients from Table 6.4.

Table 6.4: Final estimation β values in walking time model of corridor segments.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	-2.1194	0.735	-2.882	0.004	-3.564	-0.675
length	0.7710	0.022	34.672	0.000	0.727	0.815
groupsize	0.6096	0.303	2.011	0.045	0.014	1.205
luggage	1.9808	0.677	2.927	0.004	0.652	3.310
faregates	3.6733	0.466	7.881	0.000	2.758	4.589
lastsegment	2.7124	0.585	4.636	0.000	1.563	3.861
outlier	6.1188	0.226	27.027	0.000	5.674	6.563

$$WT_C(s) = -2.119 + 0.771L + 0.61X_{GR} + 1.981X_{LU} + 3.673X_{FG} + 2.712X_{LS} + 6.119X_O \tag{6.3}$$

6.2.1. Corridor passing speed model

The passing model is firstly fitted with the significant variables from Table 6.4. The passing speed linear estimation of the parameters is in Appendix I. The passing speed model showed that all attributes were significant. Therefore, the final passing speed (ps_C) model is given in Equation 6.4 with an adjusted $R^2 = 0.51$. The lower R^2 was expected because of the relatively large variance in the passing speed compared to the walking time. Furthermore, this model shows that the main passing speed (someone walking alone, without luggage etc.) is 1.42 m/s, which is close to the observed walking speed of 1.40 m/s in the Netherlands (Daamen and Hoogendoorn, 2007) and the determined walking speed for men from Bosina and Weidmann (2017). The signs for each beta correspond inversely proportional to those in the walking time model. For example, walking with luggage or in a group decreases the passing speed and thus increases the walking time. When comparing the two models, the influence of group size differs between both models. In the walking time model, the effect of walking in a group is smaller than in the passing speed model because the difference in the walking time model is almost neglectable. Still, in the passing speed model, the difference is around 7 %.

The largest beta value is for the outlier available X_O, representing this model's 20th per cent quicker walkers. However, the accurate representation of the variable is in doubt because when the base case is used, the passing speed becomes over 2 m/s or 7.2 km/h. That passing speed is at the boundary between walking and running for young men (Rotstein et al., 2005); therefore, this speed is already considered running for most passengers. The suggestion is to exempt this variable when modelling passengers' (normal) walking behaviour because only very few passengers will run in a metro station.

$$ps_C(m/s) = 1.42 - 0.102X_{GR} - 0.181X_{LU} - 0.286X_{FG} - 0.158X_{LS} + 0.693X_O \tag{6.4}$$

6.2.2. Validation of corridor models

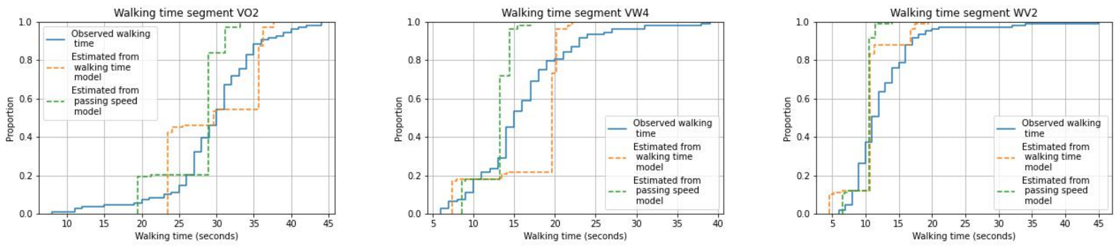
The validation results are given in Table 6.5, with the MAE and RMSE from both the walking time and passing speed model. Both models estimate small errors, indicating that models can predict the walking time well. Based on the RMSE value, the walking time model has a slightly smaller error. According to the MAE results, both models have similar error sizes. Therefore, the qualitative check should conclude whether one or both models are usable.

The qualitative check is given in Figure 6.2 for the three segments. In all three segments, both model minimum and maximum estimated walking times are within the range of the observed walking times.

Table 6.5: Validation walking time estimation from the walking time and passing speed model.

	RMSE		MAE	
Validation segment	WT_c	ps_c	WT_c	ps_c
VO2	4.39	4.86	3.52	3.69
VW4	4.78	5.79	3.83	3.81
WV2	4.14	5.40	2.45	3.05

The passing speed model predicts slightly shorter walking times than the walking time model. However, the difference is with 5 seconds, which is neglectable. In all three segments, the longest predicted walking time is close to the 80th percentile of the observed walking time, while the longest from the walking time model is around the 90th percentile. This is unsurprising because samples in the walking time model in the 80th percentile or higher were separated in the variable outlier (X_O) to represent an upper bound. The conclusion from Figure 6.2 and Table 6.5 is that the walking time model slightly performs better. Still, the passing speed is also usable to predict the walking time in corridors. However, on a generalised level, the passing speed model favours.



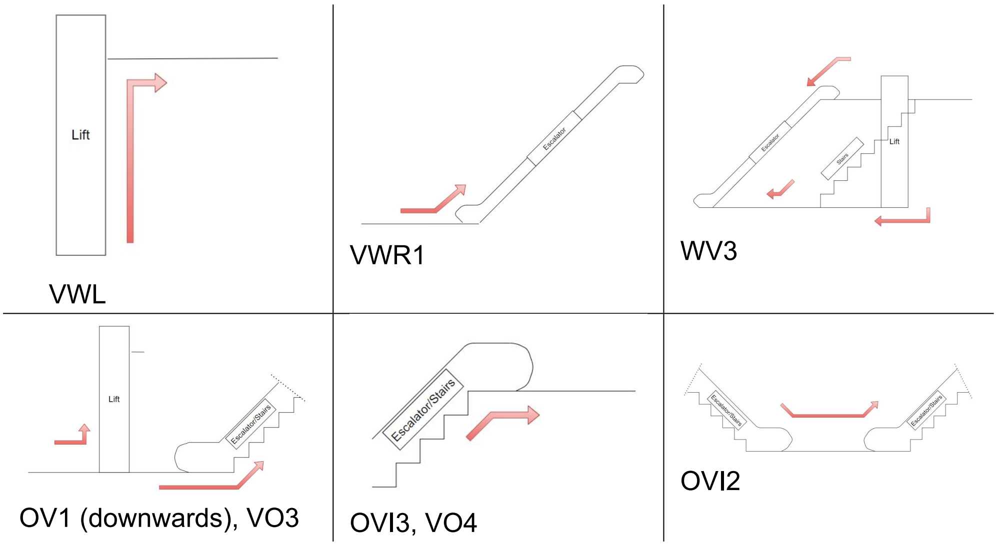
(a) Segment VO2 (b) Segment VW4 (c) Segment WV2

Figure 6.2: Cumulative observed and estimated walking times from the validation corridor segments.

6.3. Walking time model vertical transport segments

The segment layout is significantly different for the vertical transport modelling because of segment boundaries set in the data collection section 4.4 and the station layout in Appendix B. The segment layout can be categorised into six groups, which are presented in Table 6.6. The main difference is the location of the flat corridor or part in the segment, which could be in the front, back or between the vertical transport mode. Therefore, the location of the flat part could play a role in the walking time. Furthermore, the layout could partly cover the stairs/escalator of the segment. These layout elements led to the decision to split the length variable into one describing the flat corridor length and one the vertical transport mode length.

Table 6.6: The various layout of the vertical transport segments used during the data collection.



Furthermore, the decision is to only differentiate between the lift and a stair mode (stairs or escalator) as vertical transport modes because the data analysis in section 5.2 showed an insignificant difference. However, this assumption is only valid when the escalator has the same travel speed as in Beurs station.

Additional (layout) variables are used to improve the model for the second model step and better represent different vertical transport layouts. The Table 6.7 presents all extra variables to test besides the standard ones. Most of the extra layout variables explain the different layouts from Table 6.6, such as available vertical transport mode covering the complete stairs. The only exception is the variable (X_B) bottleneck because only one of the vertical transport was a capacity restriction segment. The slope (X_S) variable is included because of the significant difference in passing speed when ascending or descending from section 5.3. The base case in the model regarding the layout attributes, is a complete vertical transport segment without fare gates and has no bottleneck effect.

Table 6.7: Additional layout variables to fit for the walking vertical transport time models.

Data collection category (Step 1)	Description	State	Variable Name	Type Variable	Constant to linear fit
VT	Vertical transport mode	Stair mode (escalator or stairs) Lift	X_{VT}	Dummy	β_{VT}
Wait	Waiting condition to board VT.	Direct Wait	X_{WA}	Dummy	β_{WA}
V length	Horizontal length of vertical transport.	-	L_v	Interval	β_{L_v}
H length	Horizontal length before/after vertical transport	-	L_h	Interval	β_{L_h}
Layout variables (Step 2)					
Width	Width of platform	< 3.5 m ≥ 3.5 m	X_{WI}	Dummy	β_{WI}
Fare gates	Presence of fare gates in segment	None Present	X_{FG}	Dummy	β_{FG}
First segment	Segment directly after the platform.	After platform, before platform or None	X_{FG}	Categorical	β_{FS}
Complete	Segment contains complete vertical transport length	Partly Complete	X_{CO}	Dummy	β_{CO}
Stairs only	Segment contains only stair modes	No Yes	X_{SS}	Dummy	β_{SS}
Part1	Location horizontal part	Before After Unclear/Middle	X_P	Categorical	β_P
Slope	Direction of vertical movement	Up Down Unclear	X_S	Categorical	β_S
Bottleneck	Segment develops queues	Partly Complete	X_B	Dummy	β_B

There was a significant correlation between certain layout variables, mainly the fare gates and slope had some strong relationships with different variables. The complete correlation matrix is in Appendix I. Therefore, besides the already insignificant ones, these are variables to remove in the backward elimination. One correlation relation is kept in the third model step, which is between vertical transport mode and luggage because passengers with large luggage tend to use the lift. However, when the lift is absent, these passengers might require more walking time to clear a stair mode. In the third fit, only the variable describing the location of the horizontal part was relevant from the layout perspective. Only a correlation is present between the vertical transport mode and luggage for the data collection variables. This is obvious because passengers with large items mostly cannot take the stairs or escalator. The model development in the three steps is shown in Table 6.8. Removing the insignificant and correlated variables increases the F-statistic in step three, indicating a better model fit. The adjusted R^2 gives the same false information in this model as the corridor walking time model because the adjusted R^2 compares to the mean of all included samples, while each segment has its specific mean walking time as shown in Table 5.1.

Table 6.8: Walking time model development for the vertical transport segments.

Walking time vertical transport segments	$F(df_{res}, df_m)$	F-statistic	p	Adjusted R^2
Step 1: Data collection variables only + length	(616, 8)	184.4	< 0.001	0.70
Step 2: Additional layout variables + outlier	(611, 13)	184.0	< 0.001	0.79
Step 3: Additional layout variables & backward elimination.	(616, 8)	293.6	< 0.001	0.79

The final estimated parameters are in Table 6.9. The signs of all beta values highlight that the walking time adds time when a certain characteristics are met other than the base case of someone walking alone, using the stairs and having small luggage. The vt mode and group size betas values align with the differences observed in the walking times per segment in Table 5.5 and Table 5.9. However, the beta value of 45 seconds is double as long as the actual travel time of the observed lift. Therefore, the parameter already includes a certain average waiting time for lift passengers. The parameter for the waiting condition is on the low side compared to the observed waiting to board and direct boarding difference in Table 5.9. However, the parameter does highlight that this walking time model is only applicable when vertical transport is not a bottleneck or encounters intense crowding to board a vertical transport mode. The second largest influence is the model outlier, where the slower passengers are penalised with almost 10 seconds, while the quick walkers significantly reduce their walking time by 10 seconds. Especially for the quicker passengers, the beta value might underestimate the walking time because the difference between the minimum and median walking time for some vertical transport segments in Table 5.1 is smaller than 10 seconds. For the large luggage value adding 4.6 seconds when someone has large luggage is logical because, especially in segments with only stairs or escalators, a significantly different walking time was observed in Table 5.4. Lastly, the location of the flat part in the vertical transport segment is important, deducting or adding 2 seconds when in front or behind the vertical transport.

Table 6.9: Final estimation β values in walking time model of vertical transport segments.

	coef	std err	t	P> |t|	[0.025	0.975]
const	3.2515	0.847	3.837	0.000	1.589	4.914
luggage	4.5633	1.040	4.387	0.000	2.523	6.604
groupsize	2.4592	0.602	4.082	0.000	1.277	3.641
vt	43.7111	1.323	33.040	0.000	41.116	46.307
wait	4.6326	0.908	5.104	0.000	2.852	6.413
v_length	1.8151	0.125	14.477	0.000	1.569	2.061
h_length	0.4811	0.063	7.645	0.000	0.358	0.605
part1	2.4093	0.403	5.983	0.000	1.619	3.199
outlier	9.9441	0.459	21.655	0.000	9.043	10.845

The final walking time estimation model for vertical transport segments is shown in Equation 6.5. Using the lift is the largest significant factor. On the passenger side, large luggage affects the walking time the most. The same limitation as with the corridors applies; the model only works for relatively short segments and one transporting between one level.

$$WT_V(s) = 0.481L_h + 1.815L_v + 2.459X_{GR} + 43.711X_{VT} + 4.633X_{WA} + 4.563X_{LU} + 9.944X_O + 2.409X_P \quad (6.5)$$

6.3.1. Vertical transport passing speed model

To translate the walking time to the passing speed, the total length of the segment is used rather than the horizontal and vertical transport length as in the walking time model, for modelling convenience. The passing speed model had to be adjusted because variables part1 (X_p) and luggage (X_{LU}) were insignificant, as shown in Appendix I. The last variable removal can be explained by the correlation between samples with large luggage preferring the lift. Why X_p is insignificant in the passing speed model is unknown. Therefore, the passing speed (ps_V) model is given in Equation 6.6 with only significant variables. The adjusted ($R^2 = 0.55$) determines that the fit is similar to that of the corridor passing speed model. The base horizontal passing speed in the vertical transport is then 0.86 m/s, higher than the modelled walking speeds in Z. Chen et al. (2016) but close to the observed passing speed median of all vertical transport segments in Figure 5.5. However, these segments include a horizontal part, where passengers walk faster than one the vertical transport mode. All beta values, except the beta outlier, are negative, indicating a lower passing speed and, thus, a higher walking time when certain attributes are present. The reduction in the passing speed when using the lift (X_{VT}) has the largest impact, in line with the walking time model and the observed difference. Similar to the walking time model, the waiting for boarding penalty is twice as high as walking in a group. The only positive value is the outlier, but this value might present conditions when passengers are running on the stairs. Combined with the vt mode, X_{VT} attribute, the passing speed remains the constant value. This case would present a case where a passenger using the lift can directly board the lift.

$$ps_V(m/s) = 0.860 - 0.081X_{GR} - 0.567X_{VT} - 0.162X_{WA} + 0.537X_O \tag{6.6}$$

6.3.2. Validation of vertical transport models

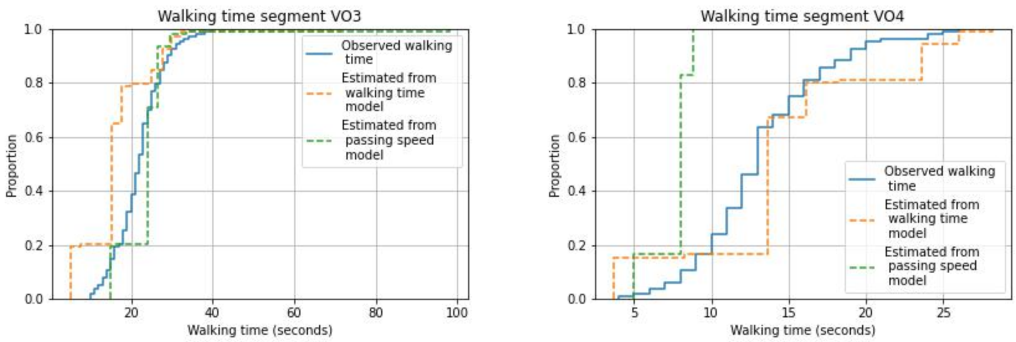
The validation of both models is done with segments VO3 and VO4. In Table 6.10, the quantitative goodness of fit parameters are given from the validation segments. These have contradictory results because, in segment VO3, the fit of the passing speed is better, while in segment VO4, the opposite is true. Based on these differences, the decision is to calculate both parameters for the remaining segments, which estimates the model's fit. Furthermore, the qualitative validation of the models might also explain the different results in these parameters.

Table 6.10: Validation walking time estimation from vertical transport segments' walking time and passing speed model.

	RMSE		MAE	
Validation segment	WT_V	ps_V	WT_V	ps_V
VO3	7.41	7.02	6.23	3.35
VO4	3.60	6.60	2.89	5.57

The visual of both models in Figure 6.3 already gives more motivation for the differences seen in Table 6.10. In segment VO3, the first estimated walking time from the walking time model is below the observed times. Furthermore, most of the estimated walking times are shorter than the observed ones, which could be the reason for the higher RMSE and MAE. The passing speed model estimates walking time within the observed time range but with one larger outlier around 100 seconds. This outlier is from the only sample in the segment taking the elevator that has all the remainder characteristics from the passing speed model (with the exemption of the outlier).

In segment VO4, the passing speed model predicts all walking times below the 20th percentile of the observed ones, while the walking time model predicts the time within the complete range. Therefore, the walking time model can predict the walking time better in both segments. The provisional conclusion is that the walking time model is better for the vertical transport segments.



(a) Segment VO3

(b) Segment VO4

Figure 6.3: Cumulative observed and estimated walking times from the validation vertical transport segments.

To support the findings from the previous paragraph, Table 6.11 shows the RMSE and MAE from the model calibration segments. The largest errors are all in segments with elevators, especially the RMSE from the passing speed in segments WV3 and VWL are significantly larger. The reason could be in line with the outlier seen in Figure 6.3 in segment VO3. Moreover, except for segment OVI3, the walking time model has smaller errors. Therefore, the walking time model is the preferred choice for vertical transport segments for estimating walking times.

Table 6.11: Goodness of fit parameters of segments used to fit the models.

	RMSE		MAE	
Calibration segment	WT_V	ps_V	WT_V	ps_V
OV1	13.15	18.84	7.92	8.33
OV2	7.79	10.91	4.83	7.49
OVI2	8.59	10.92	5.32	6.96
OVI3	3.69	5.66	2.81	4.77
WV3	4.35	76.87	3.17	17.96
VWR1	4.90	3.10	3.40	2.20
VWL	19.93	89.69	15.47	55.61

6.4. Walking time model platform segments

For the last model, besides the standard variables from the data collection, one platform-specific variable is the alighting position. However, this variable requires an adjustment from the collected ones. The sample alighting variable is transformed from the relative position in the metro (front, middle and back carriages) to the relative position to the platform exit (closest $\frac{1}{3}$ to the exit, middle $\frac{1}{3}$ and furthest $\frac{1}{3}$ to the exit). This procedure is the same as for the analysis of the passing speed on the platform in section 5.3. The base case for this value is a passenger alighting in the middle. When someone alights closer to the exit, a negative value is given because of the expected shorter walking time for those samples. A positive value is issued when a sample alighted further from the exit than the middle because of the longer walking time to leave the platform. The exact values for alighting are presented in Table 6.12.

Furthermore, each sample's actual walking distance has not been collected. Therefore, the average length from each alighting part is used as the length in the model. The maximum difference between the estimated and actual walked distance is around 15 meters, which is about 16 % of the total considered platform length of 90 meters. Therefore, this will include bias to the walking time model, and the output might not be re-presentable. Furthermore, for both models, the data input included samples which could not directly leave the metro and, thus, encounter a wait. The walking time model will predict the alighting time and not an exact walking time. For the passing speed model, the estimated speeds are the alighting speed of passengers and, by far, do not represent the walking speed. In Table 6.12, one platform-specific layout variable is the number of exits (X_E) because the A/B/C Eastbound platform has two exits, while the other platforms only have one (transfer) exit. Therefore, in further studies, the effect of exits could be studied.

Table 6.12: Additional layout variables to fit for the walking time platform models.

Data collection category (Step 1)	Description	State	Variable Name	Type Variable	Constant to linear fit
Alight	Relative average distance to (transfer) exit from a metro part.	Closest $\frac{1}{3}$ Middle $\frac{1}{3}$ Furthest $\frac{1}{3}$	X_A	Categorical	β_A
Layout variables (Step 2)					
Exit	Number of exits at platform	1 Exit 2 Exit	X_E	Dummy	β_E
Width	Width of platform	< 3.5 m ≥ 3.5 m	X_{WI}	Dummy	β_{WI}

The base case of the platform models is someone walking alone, without luggage and crowding, which alights in the middle relative to the exit's location, and there is a platform with only one exit. The first walking time model fir with only the data collection variables, and the length already has a decent fit of a R^2 of 0.64. However, again the R^2 is misleading because of the various platform layouts and walking time distributions per platform. In the second calibration step, the fit improved according to Table 6.13, with an R^2 of 0.84, when including the outlier and exit variables. After the second fit, the correlation matrix in Appendix I showed a strong relation between outlier, alight and length. This was expected because the closer a sample alighted to the exit, the shorter the walking distance becomes. Therefore, this correlation is accepted in the model. The backward elimination removed the width and the exit variables because of the insignificant contribution to the model.

Table 6.13: Walking time model development for the platform segments.

Walking time vertical transport segments	$F(df_{res}, df_m)$	F-statistic	p	Adjusted R^2
Step 1: Data collection variables only + length	(417, 6)	126.3	< 0.001	0.64
Step 2: Additional layout variables + outlier	(414, 9)	240.2	< 0.001	0.84
Step 3: Additional layout variables & backward elimination.	(419, 4)	537.6	< 0.001	0.84

The remaining variables are given in Table 6.14. This final fit determines that a time constant is insignificant. Besides the outlier and length, the same variables are significant as in the data analysis for the platform segments. The largest attribute of the model is X_O, whether a sample is part of the 20 per cent quickest or slowest walkers. The most noticeable result is the sign of the alight (X_A) variable, which is negative. So, according to the model, someone alighting closest to the exit will receive an 8 seconds penalty, while some alighting far away has an 8-second reduction. The possible cause is the correlation between variables alighting and outlier because the samples alighting closest to the exit will also have the shortest walking time. The variable group size X_{GR} was expected to have a larger value because the overall difference was already 10 seconds between walking alone and in a group in the median values in Table 5.5. The length contribution in the model is similar to the one from the corridor model, indicating that the relationship between the covered distance and the walking time is the same.

Table 6.14: Final estimation β values in walking time model of platform segments.

	coef	std err	t	P> |t|	[0.025	0.975]
const	1.7893	1.332	1.343	0.180	-0.829	4.408
groupsize	2.5804	0.922	2.800	0.005	0.769	4.392
alight	-8.2165	1.035	-7.941	0.000	-10.250	-6.183
length	0.7629	0.041	18.620	0.000	0.682	0.843
outlier	15.0751	0.669	22.544	0.000	13.761	16.389

The walking time model equation becomes as in Equation 6.7. This one does not have a constant compared to the other walking time models because of the insignificant attribute to the model in Table 6.14.

$$WT_P(s) = 0.763L + 2.58X_{GR} - 8.217X_A + 15.075X_O \tag{6.7}$$

6.4.1. Platform passing speed model

The passing speed (ps_P) model in Equation 6.8 had the adjusted $R^2 = 0.55$, and all variables are still significant. However, the beta value of the outlier parameter X_O is large and, therefore, questionable. As discussed in section 5.3, the walking distance was assumed to be the same for all samples per platform segment, the average distance to the exit. While in reality, each sample had a different walking distance. Therefore, the passing speeds show a significantly larger variation, especially at the closest part to the exit, and the outlier beta value represents that variation. Compared to the walking time model, the alighting parameter X_A is now positive. Samples alighting further than the average walking distance will have a high estimated passing speed. In line with the other segment-types passing speed model, walking in a group negatively affects the passing speed. However, the parameter for X_{GR} is double as large as the other models. The constant in the model is the same as the observed lower bound of the median passing speed.

$$ps_P(m/s) = 1.107 - 0.177X_{GR} + 0.177X_A + 1.7791X_O \tag{6.8}$$

6.4.2. Validation of platform models

As only five platform segments are modelled, the validation will directly present the RSME and MAE for the fitted and validation segment in Table 6.15. Not only are the errors smaller in the validation platform for the walking time model, but also for the other platforms. Furthermore, the errors are the same magnitude for the walking time model for both parameters. The walking time model is preferred based on this goodness of fit test.

Table 6.15: Validation walking time estimation with the models.

	RMSE		MAE	
Platform segment	WT_P	ps_P	WT_P	ps_P
A/B/C Eastbound to eastern exit	5.67	7.09	4.80	5.26
A/B/C Eastbound to western exit	8.94	13.17	6.99	8.95
A/B/C Westbound	6.77	7.13	5.61	5.20
D/E Northbound	7.27	20.45	5.65	17.36
D/E Southbound (validation segment)	7.37	12.58	5.70	9.92

The visual validation of both models is shown in Figure 6.4. Both estimations show a similar cumulative distribution of the walking times, where their maximum estimated walking times are the same. Only the minimum walking time from the walking time model is slightly out of range of the observed times. The qualitative validation concludes that both models accurately approximate the walking time. However, combined with the result from Table 6.15, the walking time model favours the passing model because of the smaller errors. But, as mentioned at the beginning of the section, the model is biased because of the average walking distance used per alighting location to fit the model.

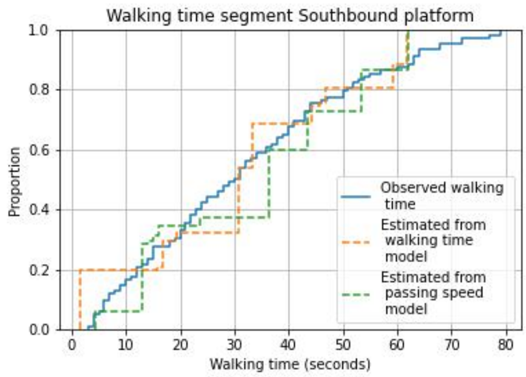


Figure 6.4: Cumulative observed and estimated walking times for the validation platform segment.

6.5. Main conclusion from the walking time modelling

The validation results of all models indicate that the walking time model estimated the walking time more accurately than the passing speed model for all segment types. The walking time model validation errors are smaller for all three segment types, and all estimated walking times are within the range of the observed walking times per segment. Lastly, all walking time models have at least the same significant variables as seen in the data analysis in section 5.2. The multiple linear estimation models for each segment's type on the walking time, with the relevant significant variables, explain the variance in the walking time with the outlier X_O variable. For each combination of attributes, the models can predict an estimated mean and lower bound of the walking time. Additionally, the walking time model estimates an upper bound of the walking time.

Only for the vertical transport segment type, the walking time model has more attributes than the passing speed model. The reason is the choice of backward elimination as the model calibrating approach. Furthermore, the passing speed has a smaller absolute variance between the observed passing speeds compared to the observed walking times. Therefore, a specific combination of attributes has a higher probability of having the same passing speed, while in the walking time, the difference can vary by several seconds. The walking time and passing speed models have the same attributes for their specific segment types in the corridor and platform segments.

Therefore, to answer the research question, how can the significantly related passenger and station attributes be modelled by splitting up the transfer path into three segment types: platform, corridors and vertical transport points. Using the walking time models for each segment type can estimate the walking time per combination of attributes. Each model can predict a mean, an lower and an upper bound of the walking time per combination of attributes per segment

6.6. Discussion on the generated models

One essential remark is given on the corridor segment models. The attribute of group size has a significant effect through the passing speed model, while in the walking time model, the effect is neglectable. However, both models have the same attributes, therefore the passing speed could be more useful to predict different walking times in a corridor. Furthermore, the walking time models only add or subtract a certain amount of time per attribute combination independently of the length because of the chosen multiple linear estimation approach. However, the difference could become smaller on longer segments and, thus, insignificant. Therefore, a different non-linear approach could be tested for the walking time model, where the contribution of attributes also depends on the segment length. The passing speed models lack this problem because the length is unrelated to the passing speed.

Moreover, the effect of large luggage has contradicting results between the walking time analysis and the models. The category luggage was insignificant in the data analysis in section 5.2. However, the corridor models showed that large luggage is a significant attribute, while the platform models showed an insignificant influence on the walking time prediction. For the vertical transport segment model, large luggage is only present in the walking time model. As explained in section 5.1, too few samples with large luggage were collected in multiple segments. When more samples had been collected, the data analysis might also indicate significant walking time differences between non-luggage and large luggage passengers.

Three large limitations apply to these models. Firstly, the corridor and vertical transport segments are tested and validated for short segments (under 50 meters). Therefore, further studies could test or improve the predictive power of walking times for longer segments. Secondly, the vertical transport walking time model is only valid for segments where capacity restrictions or queues are mostly absent. In addition, the model is only calibrated and validated with segments where significant waiting time was absent from boarding a vertical transport mode. Lastly, the models can only estimate the walking time at a segment level. The estimated walking time per segment should be summed to derive a transfer walking time per combination of attributes.

7 Conclusions

This chapter provides the main conclusions from this study. The setup is as follows: section 7.1 answers the research sub-questions and the main question. Besides the answer, the main limitations on the conclusion of this study are in section 7.3. The study contribution on a scientific and societal level are in section 7.2. Lastly, the recommendations in section 7.4 discuss practical recommendations and future study topics from the answers and limitations of this study.

7.1. Answer to the research questions

This section presents the answer to the main research question and the supporting sub-research questions. The conclusions from each sub-question contribute to answering the main research question. Therefore, these are answered first before the main research question.

1. *What passenger's characteristics or decisions, station layout or other transfer-related elements play a role in the transfer walking time?*
   A conceptual model has been devised in Figure 2.1 to present the possible influential elements on the transfer walking time, which were split into passenger and layout-specific attributes. The walking speed of passengers depends on various aspects, and the main influential ones are the gender, group size and trip purpose of a person, whereas the last one is recommended to study further for metro transfers. A different significant one in the walking speed when using the stairs is the age of the passenger, especially between young adults and middle-aged passengers. Lastly, people carrying large luggage tend to walk slower as well. Furthermore, the transfer walking time is affected by the passenger route choice in the station, whether multiple transfer paths are present, and by the vertical mode choice. Passenger base their choice of vertical transport between stairs and the escalator on crowding conditions. Moreover, the lift is often excluded as a vertical transport mode because of the expected low usage. Additionally, the transfer walking time depends on the length of the transfer path or the travel time of a vertical transport mode.

   From the station layout and traffic elements, capacity-restricted queues or crowding affects the walking speed and, thus, the walking time the most. Especially vertical transport boarding has a high likelihood of queueing and increases the passing time. Longer queues might influence the choice between stairs and escalators, but they prefer to take the escalator in a station. The lift is an unconsidered vertical transport mode in previous literature, while some passenger groups, such as those with walkability issues or older adults, depend on using them. Extensive crowding in combination with bidirectional flows significantly decreases the walking speed and thus increases the walking time. Lastly, the moment of making a transfer might significantly impact the transfer walking time, as this is revealed for transfers between the train and the metro. However, the conclusion cannot be drawn if this statement also applies to metro transfers. In Beurs metro station, the hypothesis was false as the walking time between peak and off-peak conditions showed an insignificant difference.

2. *How are metro transfer walking times currently estimated or collected?*
   Two methods are frequently used to collect transfer walking times for metros transfers. The first one uses covert observations, where observers or data collectors follow passengers making the transfer in the station. Other possible tools are Bluetooth tracking or videotapes, but these methods have been used for passenger movements in a station rather than collecting the transfer walking time. These covert observations discovered that the transfer walking time followed a lognormal distribution.

   Nowadays, most transfer walking times are estimated using Automatic Fare Collection and Automatic Vehicle Location data sources. The drawback of this method is the underlying assumptions of the generation of the walking time distribution and the lacking information about passenger characteristics, behaviour or the transfer path layout. Moreover, these models mainly predict the transfer time, including the waiting time of passengers before the arrival of the following metro.

   A different estimation approach to predict the transfer walking time uses only the station capacity elements, such as platform exit locations, corridor widths and escalator capacity. A walking speed function then computes the transfer walking variance to distinguish a certain level of crowding. This model only had the escalator as a vertical transport mode and lacked passenger characteristics.

3. *Which attributes have the most significant effect on the transfer walking time?*
   This research can only answer on an observed transfer path segment level or for the complete estimated transfer. Answering this question on the whole observed transfer walking time is impossible in this research because of the data collection setup in the station.

   From the observations in metro station Beurs, three categories were observed to have significantly different walking times on a segment level. Firstly, almost all segments showed a difference between walking alone and walking in a group. The walking time was found to be between 10-40 % longer when passengers were walking in a group. Secondly, at the platform segment, the passenger's alighting location from the metro was observed to be significant because of the walking distance towards the platform exit and, thus, the walking time of a passenger. The further an alighting part was from the exit, the longer the walking time became. The average difference was recorded to be around 20 seconds between the alighting parts, but the statement is only valid when considering one exit. The last and most significant category relates to the vertical transport mode choice. Passengers using the lift were found to take around 45-60 seconds longer than those using the stairs or escalator. The magnitude of a longer walking time depends on the waiting time and queue to board the lift. Compared to the median of the vertical transport modes, the walking time is at least double as long for those using the lift. Lastly, the waiting condition to board a vertical transport mode had a walking time difference of around 10 seconds.

   The Monte Carlo simulation's estimated total transfer walking time shows similar results as the observations. However, the exact difference depends on the transfer path layout. Some significant differences occurred in the walking time on these relatively short transfer path lengths. Compared to the overall estimated transfer walking times, walking in a group increases the time by around 15-40%. Furthermore, the transfer walking time can be quicker, up to 30% for the passengers alighting the platform closest compared to the exit to the overall transfer walking time distribution. While passengers who alight relatively far from the exit, the transfer walking time could be 20% longer compared to the average walking time. The differences are more considerable when comparing each furthest and closest part of the platform to the exit. Between 20% and 60% is the walking time longer for passengers who alight far away from the exit than those who exit the metro close to the exit. Passengers who must wait to board a vertical transport mode require, on average, between 30 % and 50 % more time to transfer than those who can board directly. The exact difference depends on the number of vertical transport to pass. When more vertical transport points are passed, the difference becomes larger.

   The lift user always has longer transfer walking times in the vertical transport mode choice. Between the overall and lift user-specific transfer walking time, the time was observed to be at least 45% longer. When using two lifts in a transfer, the difference between the lift-specific and overall transfer walking time increases to 71%. Between a stair mode and a lift mode, the transfer walking time is at least 55% longer when taking the lift. With two lifts in the transfer path, the transfer walking time becomes at least double as long as compared to passengers who only take stairs or escalators. Therefore, the final answer to this question is that taking the lift as a vertical transport mode significantly affects the transfer walking time.

4. *How can the information of the significant attributes be used to estimate transfer walking times better?*
A transfer path can be divided into three segment types: corridors, vertical transport and platform segment. Following the observation and data analysis, each segment type has additional segment type-specific attributes. Two multiple linear model types were used to estimate the walking time per segment type. One model predicts the walking time directly, whereas the second model estimates the walking time by computing the passing speed of passengers in the segment and translating it to a walking time with the segment length. The passing speed is close to or lower than the walking speed of passengers because of the data collection setup in this study. Furthermore, walking or passing speed models are favour when modelling passenger movements in a station.

   For the corridor model, group size and luggage are significant passenger attributes, besides layout elements such as the width, fare gate presence and being the last segment before the platform. For the walking time, the length of a segment is also a significant contributor. However, the difference between walking alone and in a group is estimated as smaller than observed. In the model's validation process, the walking time model has a slightly smaller error than the passing speed model, but both models can be used to predict the walking time in corridors.

   The attributes length, group size and luggage are significant in the walking time model for vertical transport segments. Moreover, the vertical transport mode (between a lift and a stair mode) and the waiting condition are relevant as vertical transport-specific elements. The corridor's location near the vertical transport also plays a minor role in the model. In the passing speed model, the luggage is insignificant because of the smaller variance in the passing speed and the correlation between large luggage passengers taking the lift. The walking time model is significantly better for this segment type than the passing time model, according to the validation results of both models. However, the models are only usable when queues or extensive crowding are absent.

   Group size and alighting location attributes influence the walking time for the platform models. The walking time model also has the average distance to the platform exit as an attribute. Again, the walking time model estimates the walking time better for the platform segments regarding the error size in the validation.

   The walking time and passing speed models have the same attributes for each segment type, with the exemption of the vertical transport segments. The passing model includes fewer attributes than the walking time models in the vertical transport model. The reason is the smaller variance in passing speed compared to the walking time, and thus certain combinations of attributes could result quicker in the same passing speed rather than the same walking time. In the validation process for each segment type, the walking time model estimates the walking time better than the passing speed model.

   In all models, an additional variable called "outlier" was introduced to present an upper and lower bound for the estimated walking time rather than only a mean value. A sample is part of the outlier category when the walking time is lower than the 20th percentile or higher than the 80th percentile of the segment's walking time distribution. In the passing speed models, the outlier only represents the passing speed higher than the 80th percentile, in other words, the lower bound of the estimated walking time. Otherwise, all remaining variables were insignificant in the model generation process because of the small variance in passing speeds. In all models, this variable "outlier" was significant.

All the conclusions from the sub-questions are the basis for the answer to the main research question: ***What are the significant passenger characteristics and station attributes that affect the transfer walking time in metro transfers and to which extent can this information be used to improve current estimations of transfer walking times?***

From the observations at a segment level from a transfer path, the group size always affects the transfer walking time. Followed by the vertical transport mode choice and the waiting condition to board the mode, especially between the lift and stair modes. The passengers' alighting position is another significant attribute of the metro platform segment. Crowding had an insignificant effect on the transfer walking time from the observations and within the models. However, large crowding was absent in the case study station.

The most significant effect on the complete transfer walking time is found to be the choice between the lift or stairs as a vertical transport mode during the transfer. Using the lift increases the walking time by 55% compared to those taking the stairs or escalator. When more elevators have to be taken, the walking time at least doubles. Walking alone or in a group is always significant during the complete transfer. Passengers walking alone are between 15% to 40% quicker than those who walk in a group. At the platform, the alighting position relative to the exit is also a significant attribute of the transfer walking time. The transfer walking time difference between the closest metro part to the exit and the furthest is around 20 %. Lastly, in the waiting condition to board a vertical transport mode, passengers who must wait to board require, on average, between 30 % and 50 % more time to transfer than those who can board directly. This depends on the number of vertical transport to pass. However, the exact difference depends on the transfer path layout.

Two model methods are used to predict the walking time per segment type, a multiple linear walking time model and a multiple linear passing speed model. With the use of an "outlier" variable, the model can predict, besides a mean, also a lower and upper bound for the walking time per segment. The passing speed model does include fewer attributes because of the smaller variance in the passing speed compared to the walking time. The model can only predict a mean and lower bound of the walking time through the passing speed. Otherwise, none of the passenger characteristics or layout variables could be included. Furthermore, the walking time model estimates the walking time per segment better than the passing speed model.

To model all relevant aspects of the transfer walking time, a metro transfer should be split into three segment types: corridors, vertical transport and platform segments. Modelling the transfer walking time through different segments type has to do with specific attributes for each segment type. The most significant influence is taking the lift in a vertical transport segment. In all segment types, group size and walking distance also influence the walking time. However, luggage size affects the walking time in corridors and vertical transport in the models. At the platform, the relative alighting location of a passenger is significant to the estimated walking time. Furthermore, some layout-specific attributes are also included for the corridor and vertical transport segment types.

7.2. Contribution of this study

The results and conclusions of this study add to the knowledge of transfer behaviour between rail modes, not only on a scientific level but also on a societal level. The main contributions are discussed below.

7.2.1. Scientific contribution

The most considerable contribution is the insight into metro transfer behaviour based on different passenger characteristics, choices or layout elements. The main results of this study are the significant categories influencing the transfer walking time. The effect of group size, alighting location and the vertical mode choice are the most important, not only on a transfer segment level but also on the complete metro transfer. Furthermore, this study included the lift as a vertical transport mode in stations, besides stairs and escalators. Lift users are passenger groups that were excluded in most station movement studies. Adding the fact that using is the most significant factor in the transfer walk and that multiple lifts in transfer always doubles the transfer walking time compared to those using a stair mode. Studying the effect between male and female walking behaviour, as suggested by Bosina and Weidmann (2017), in a metro station environment could be excluded as this study showed an insignificant difference between their walking time.

An addition is an insight into walking time distribution on a transfer path segment level. Not only are lognormal distributions observed, but also normally distributed ones. The alighting distribution of passengers at the platform is uniform, as suggested by Du et al. (2009), but can also be triangular when the exit is close to the middle. Lastly, the passing speed differs at different segment types, especially between the platform and the corridor. This phenomenon has not been described in previous literature. The generated walking time models have the advantage of including both passenger and layout attributes, while in previous models, mostly only station layout attributes were used (Zhou et al., 2016) or assumptions on the walking speed (Zhu et al., 2020).

7.2.2. Societal contribution

Passengers who use the lift are now part of the transfer population when modelling the transfer walking time. This group has been excluded in most previous studies on walking behaviour in (metro) stations, while their transfer walking time was found to be 50% longer. Especially passengers with walking disabilities or having large luggage are part of this group. Moreover, the difference between walking alone and in a group is now discovered to be significant for the complete transfer. Related to the case study, the public transport authority RET has an improved overview of the (transfer) walking times in Beurs station. Furthermore, the collected large luggage items description might help to improve metro stations or metro train designs to facilitate passengers with large luggage better.

7.3. Limitations

The limited time and scope of this research simplified certain steps in the data analysis and modelling. The limitations are already discussed in each specific section. However, the main limitations are summarised in this section, which is split into data collection, data analysis and model-specific limitations.

7.3.1. Data collection specific

The main limitations of this study regarding the data collection are given below.

- The walking time is the time from the start of a segment to the end of the segment. However, possible layout boundaries such as fare gates or queues to board a vertical transport mode require a certain waiting time to clear. These times have not been collected separately per sample or at a general level. Therefore, the magnitude of the relevant attributes in the models is not compared with the observed waiting times per attribute.

- Some vertical stair modes had to be split for collection convenience. However, in the second vertical transport part, the vertical transport mode was not noted per sample because the difference between the stairs and the escalator was assumed to be irrelevant. The data analysis for the first part proved the assumption was correct for these parts. Therefore, during the modelling stage, the stairs and the escalator were treated as the same mode in these split-up segments. While in reality, the walking time could be different between these modes for the complete length.
- The research focused on passengers who walked directly and only had very short stops to navigate. The effect of this limitation is the negligence of samples unfamiliar with the station layout. These were present in the collection period but mostly limited to one per segment. Therefore, the walking time analysis results do not accurately represent the metro transfer population as a whole.
- The presence of departure information might influence the walking speed and thus the walking time of passengers to catch the next metro. In the collection period, multiple samples were observed running to catch the metro when the departure display showed a time of around 1 minute. In contrast, some samples tended to slow down because of the remaining waiting time for the next metro.
- Trip purpose and the age of passengers were left out in the data collection while having significantly different walking speeds. The trip purpose of passengers has been assumed as commuters. Furthermore, the age was left out because of the standardised recommendation of Bosina and Weidmann (2017) for performing research on walking behaviour, which focused on adult samples.

7.3.2. Data analysis specific

The main limitations of this study regarding the data analysis are given below.

- The collection was done at only one station. Besides the fact that there are five transfer paths present, the results of this study do not represent generalised transfer behaviour in metro systems. Furthermore, the variation in vertical transport elements is also limited, such as the travel speed or capacity of the vertical transport modes
- This study determined that crowding is irrelevant. However, this resulted from the observations that extensive crowding was mostly absent in the stations, besides the strict quantitative crowding definition for a sample. Therefore, these results are assumed to be invalid for metro stations with considerable crowding because crowding would increase the walking time.
- To determine a difference in walking times within a category on a segment level, only the significance tests in section 5.2 were performed. The quantitative difference in the transfer walking time is obtained through the Monte Carlo simulation and the walking time models. The data analysis has not derived the exact empirical difference in walking time per segment because a segment is only a part of the complete transfer path.
- All walking time distributions were assumed to be lognormal for the significant test between categories and in the Monte Carlo simulation. However, some segments also had signs of a normal distribution or distribution type was neither of both according to the statistical tests in Table 5.3, or the walking time histogram in Appendix E. Therefore, the theory that each transfer walking time is lognormal distributed from Du et al. (2009), Zheng et al. (2014) and Zhu et al. (2020) could be questioned because if each segment has a normally distributed walking time, then the resulting transfer walking would never become lognormal. The walking time distributions at the platforms have uniform characteristics as in Du et al. (2009) and Zhou et al. (2016). However, from the histograms, one might suggest that a triangular distribution might also work when the exit is in or the near the platform middle. None of these hypotheses has been tested in the study but could be done in further studies.
- In Beurs metro station, passengers do not walk on the escalators according to the data collection reflection, while in other countries, metro travellers could walk on the escalator. Therefore, the analysis results and the modelling of the escalator as a vertical transport mode are only valid when passengers do not walk on the escalator steps.

- The model assumed an insignificant difference between stairs and escalator users on the walking time and passing speed. The assumption is based on the segment walking time results from the data analysis in segments with only stairs and an escalator. However, the statement depends on the escalators running speed and walking behaviour on the escalator.
- The exact covered length per sample on the metro platform was simplified to an average distance per platform part because the exact alighting position has not been collected. This distance was from the middle of that part of the platform to the exit. The actual walking distance of a sample is 15 meters longer or shorter because one platform part is around 30 meters. For the closest segments to the exit, the error in the distance might become the largest, affects the passing speed because of possible short walking time and, combined with a relatively long distance, gives a high passing speed. While in reality, the actual walking speed of the sample was way lower. Therefore, the model might estimate the wrong walking times for the closest passing speed models.

7.3.3. Model development

The main limitations of this study regarding the model development are given below.

- The models only predict the walking time through different attributes on a segment level of a transfer. The models have only been validated at a segment level. The models have not been used yet to predict a total transfer walking time.
- The walking time models used a linear estimation where the related attributes are independent of the length. While in reality, the attributes could be dependent on the length.

7.4. Recommendations

The study has fulfilled the objective of providing insight into influential attributes on transfer walking times for metro transfers. The insight could improve the metro scheduling or the transfer walking time modelling. Therefore, recommendations are split into practical suggestions, station modelling recommendations to RHDHV and advice on future studies.

7.4.1. Practical recommendations

The practical recommendations are split on a general level and specific for the RET. Firstly, general recommendations are given.

- The most important recommendation is to differentiate the transfer walking time for different passenger groups. At least the difference between walking alone or in a group and between escalator/stair users and lift users. In the last category, per lift present in a transfer path, an average of 45 seconds could be added because the walking time model and general data analysis presented that value. The difference between passengers walking alone and in a group, the walking time per segment, could be determined using the models. For a complete transfer path, the results from the Monte Carlo simulation show that passenger walking in a group requires 15 % to 40 % more time.
- Metro operators can use the walking time or passing speed models to estimate the transfer walking time for different passengers or layout attributes for their metro transfer station. Not only for existing stations but also for new metro stations to determine the expected transfer walking time. Furthermore, the models can also be used for non-transfer stations to estimate passengers' entry or exit times.
- With the updated transfer walking times, the metro schedule should be verified, especially at the end of the service, if all transfers are feasible for all passengers. Otherwise, the metro timetable should be changed to ensure achievable transfers for all passengers.
- In passenger route planners, the suggestion is to include passenger characteristics more, at least in metro trips. The trip planner could provide a section for passengers to fill in their characteristics. For example, only using lifts in the station or walking in a group. In that way, the route planner can present a realistic transfer walking time and, most importantly, display a feasible departure time for the next metro without just missing it. All in all, to better estimate the travel time when using the metro.

- The transfer walking time can be generalised for different passenger groups when modelling metro systems. Therefore, the model can better estimate the number of successful transfers or improve estimating passengers' travel times through the system.

The last recommendations are specific to the public transport operator RET. These are mainly based on the observations in Beurs metro station.

- Study the effect of having a fixed travel direction of the bidirectional escalators in Beurs metro station. Especially the escalator to/from the A/B/C Eastbound platform mainly travels upwards towards the platform. When a metro arrives, most passengers only have two (relatively) narrow stairs to leave the platform. For a quicker throughput of passengers leaving the platform, the RET could consider the escalator to run downwards during busier periods.
- Investigate the signage at the Beurs metro station because some passengers needed additional clarification from the station staff about which route to take to make the transfer. The transfer path of A/B/C Westbound to the D/E lines could have more and better signage compared to the other transfer paths. An example could be using the same signage stickers on the floor as found in the other transfer paths.

7.4.2. Recommendations for RHDHV

For RHDHV, the following recommendations are made for station modelling or scheduling metro systems to provide improved consults to their customers.

- The estimations from the passing speed models can be used to compare the currently used walking speeds in the Massmotion for modelling stations. Otherwise, new agents could be introduced with the relevant attributes from the passing speed model. At least differentiate in the simulation between samples walking alone and in a group because that difference is always significant.
- The differentiation between escalator/stairs and lift users can be implemented in the station model by using different agents in MassMotion. An alternative approach is the assumption that lift users require additional walking time.
- When metro timetables are planned, the transfer times could be checked to see if they are feasible regarding the walking time for all potential metro users. The transfer walking time can be estimated using the walking time or the passing speed models.

7.4.3. Recommendations on future studies

The limitations in section 7.3 present opportunities to further study the transfer walking time. The recommendations for further studies are split into two parts. Firstly, for the data collection and secondly for the model improvements. To further generalise the results of this study, the data collection could be performed at more metro stations because the current findings are from only one station. Moreover, the most important recommendation is to collect more samples making the metro transfer during more crowded situations for metro transfers. The observed walking times in this study were mainly in low or slightly crowded conditions, while extensive crowding could occur in metro stations. Then the walking time model is representable for more situations and might include crowding as a separate variable.

Secondly, collecting information about the following categories in the transfer walk may be desirable to check if additional categories are significant in the transfer walking time.

- *Trip purpose or familiarity in the station:* As the literature study proves, the trip purpose influences walking speeds and, therefore, the walking time. However, this research assumes that all passengers are mostly commuters. The data collection was partly performed on moments when the share of commuters was lower to expand the representation of the walking time for the entire passenger population. Nevertheless, the research findings on the walking time might not be valid for different trip purposes. Therefore, the recommendation is to analyse the transfer walking time along the trip purpose categories. Furthermore, passengers unfamiliar with the station had longer walking times but were out of scope in the data collection. These could also be included in future studies on the transfer walking time.

- *Age:* The walking speed reduces as a person becomes older, according to the literature. During the collection period, older people were indeed walking slower. However, the data collection focused on passengers with ages in the range from 18-50 from Bosina and Weidmann (2017). In future studies, the walking time of older passengers could be collected and compared with those from this research.
- *Include disabled passengers:* In this study, almost all samples were without any walking disabilities because of the assumption that these are an insignificantly small share of the total passenger population. Collecting transfer walking times for disabled passengers might help to improve the understanding of their movements in a metro station.
- *The waiting or passing time to clear certain layout elements* As the waiting time to board vertical transport or to pass a fare gate was not collected, the recommendation is to collect these times separately. Then, these waiting times could be compared with the related model attributes or improve the models.

Another item to investigate followed the data collection reflection: the (voluntary) waiting behaviour for bidirectional escalators in (metro) stations. Some observed samples significantly waited a long time for the escalator to turn in their direction, while there was also a lift available or a different route to continue their walk in the station.

Regarding modelling the transfer walking time, the following recommendations are made.

- *Perform Monte Carlo simulation with a different type(s) of distribution* The Monte Carlo simulation assumed all the walking times per segment of having a lognormal distribution. While some segments had normal or uniform walking time distributions. Therefore, the Monte Carlo simulation could be done with the actual walking time distribution type and check whether the resulting transfer walking times are significantly different compared to assuming only lognormal walking time distribution per segment.
- *Estimate a total transfer walking time with the models* A total transfer walking time has not been estimated with the combination of all three models. Therefore, the recommendation is to estimate a complete transfer walking time with the models. The check could be done to determine whether the same significant difference between attributes is seen on a segment level and with the Monte Carlo simulated transfer walking time.
- *Compare Monte Carlo simulated transfer walking time with those from the models* The three walking models have not validated the transfer walking times from the Monte Carlo simulation study. In new research, this could be done and check which type of estimation is preferred.
- *Calibrate and validate the walking time models with walking time from a different station:* The current models are only validated with data from one metro station. The recommendation is to test if the model also predicts the walking time in a different metro station, preferably in a station with similar walking behaviour as in this studied metro station. Otherwise, the models could be calibrated with walking times from other stations to represent more metro stations.
- *Calibrate or validate the vertical transport models with different vertical transport mode elements:* The current models has limited variation in the travel speed or capacity of the vertical transport modes. The recommendation is to check if the model also applies to modes that travel slower, quicker, or have a different capacity.
- *Calibrate or validate the walking time models for other rail or public transport transfers:* The presented model is validated for metro transfers only. Testing whether the models are also valid for other rail or public transport transfers, such as train-metro, metro-bus or train-train, is recommended. If not, try how these models could be calibrated to be valid for other transfers. In that way, the models could be used for more public transport transfers.
- *Use a non-linear approach to model the walking time* in future studies, a non-linear approach could be used to generate a walking time model that includes the effect of the significant attributes depending on the length.

Bibliography

Bosina, E., & Weidmann, U. (2017). Estimating pedestrian speed using aggregated literature data. *Physica A: Statistical Mechanics and its Applications*, *468*, 1–29. https://doi.org/10.1016/J.PHYSA.2016.09.044

Cascajo, R., Garcia-Martinez, A., & Monzon, A. (2017). Stated preference survey for estimating passenger transfer penalties: Design and application to madrid. *European Transport Research Review*, *9*, 1–11. https://doi.org/10.1007/S12544-017-0260-X/TABLES/2

Centraal Bureau voor de Statistiek. (2021). *Hoeveel wordt er met het openbaar vervoer gereisd?* https://www.cbs.nl/nl-nl/visualisaties/verkeer-en-vervoer/personen/openbaar-vervoer

Chen, Y., Mao, B., Bai, Y., Ho, T. K., & Li, Z. (2019). Optimal coordination of last trains for maximum transfer accessibility with heterogeneous walking time. https://doi.org/10.1155/2019/9692024

Chen, Z., Zhao, X., & Shi, R. (2016). Walking speed modeling on transfer passengers in subway passages. *International Conference on Civil, Transportation and Environment (ICCTE 2016)*, 639–643.

Daamen, W., & Hoogendoorn, S. P. (2007). Free speed distributions — based on empirical data in different traffic conditions. *Pedestrian and Evacuation Dynamics 2005*, 13–25. https://doi.org/10.1007/978-3-540-47064-9_2

Daamen, W., Bovy, P. H. L., Hoogendoorn, S., & van de Reijt, A. (2005). Passenger route choice concerning level changes in railway stations. *84th Annual Meeting Transportation Research Board*, 1–18. https://www.researchgate.net/publication/228868259

Daamen, W., van den Heuvel, J., Ton, D., & Hoogendoorn, S. (2015). Using bluetooth and wifi to unravel real-world slow mode activity travel behaviour. *14th International Conference on Travel Behaviour Research*.

de Dios Ortúzar, J., & Willumsen, L. (2011). *Modelling transport* (4th). John Wiley & Sons Ltd.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. (2005). *A modern introduction to probability and statistics*. Springer. https://doi.org/https://doi.org/10.1007/1-84628-168-7_14

Dinno, A. (2015). Nonparametric pairwise multiple comparisons in independent groups using dunn's test. *The Stata Journal*, *15*, 292–300.

Dixit, M., Brands, T., van Oort, N., Cats, O., & Hoogendoorn, S. (2019). Passenger travel time reliability for multimodal public transport journeys. *Transportation Research Record*, *2673*, 149–160. https://doi.org/10.1177/0361198118825459

Du, P., Liu, C., & Liu, Z. L. (2009). Walking time modeling on transfer pedestrians in subway passages. *Journal of Transportation Systems Engineering and Information Technology*, *9*, 103–109. https://doi.org/10.1016/S1570-6672(08)60075-6

Eltved, M., Lemaitre, P., & Petersen, N. C. (2021). Estimation of transfer walking time distribution in multimodal public transport systems based on smart card data. *Transportation Research Part C: Emerging Technologies*, *132*, 103332. https://doi.org/10.1016/J.TRC.2021.103332

Ensing, J. (2022). *The application of a multi-operator smart card dataset to identify transfer locations with a high potential for transfer time loss minimization.* [Master thesis, Delft University of Technology].

European Commission. (2020). *Communication from the commission to the european parliament, the council, the european economic and social committee of the regions: Sustainable and smart mobility strategy – putting european transport on track for the future*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0789

Field, A. (2013). *Discovering statistics using ibm spss statistics* (M. Carmichael, Ed.; 2nd). SAGE Publications Ltd.

Fruin, J. J. (1992). *Designing for pedestrians*. National Transportation Library: Research; Innovation Technology Administration (RITA). http://ntl.bts.gov/DOCS/11877/Chapter_8.html[7/7/20149:48:44AM]

Fujiyama, T., & Cao, B. (2016). Lengths of time passengers spend at railway termini: An analysis using smart card data. *2016 IEEE International Conference on Intelligent Rail Transportation, ICIRT 2016*, 139–144. https://doi.org/10.1109/ICIRT.2016.7588723

Guerrieri, M. (2023). Metro rail systems. *Fundamentals of railway design* (pp. 229–238). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-24030-0_13

Hodson, T. O. (2022). Root-mean-square error (rmse) or mean absolute error (mae): When to use them or not. *Geosci. Model Dev*, *15*, 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

International Organization for Migration. (2015). *World migration report 2015: Migrants and cities - new partnerships to manage mobility*. International Organization for Migration. United Nations. https://doi.org/10.18356/BE2A2BE6-EN

Kouwenberg, M., Kleft, S., & in 't Veld. (2019). Meer zicht op regionaal reisgedrag: Combinatie van ov-chipkaartgegevens bij de vervoerregio amsterdam. *Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk*.

Li, Q., Ji, C., Jia, L., & Qin, Y. (2014). Effect of height on pedestrian route choice between stairs and escalator. https://doi.org/10.1155/2014/965305

Lin, D., Vos, J. D., Maruyama, T., Bobylev, N., & Cui, J. (2022). Metro-related transfers: A review of recent literature. *Journal of Urban Planning and Development*, *148*. https://doi.org/10.1061/(asce)up.1943-5444.0000858

Liu, T., Cats, O., & Gkiotsalitis, K. (2021). A review of public transport transfer coordination at the tactical planning phase. *Transportation Research Part C: Emerging Technologies*, *133*, 103450. https://doi.org/10.1016/J.TRC.2021.103450

Long, S., Meng, L., Miao, J., Hong, X., & Corman, F. (2020). Synchronizing last trains of urban rail transit system to better serve passengers from late night trains of high-speed railway lines. *Networks and Spatial Economics*, *20*, 599–633. https://doi.org/10.1007/s11067-019-09487-0

MRDH. (2023). *Knooppunt Beurs*. https://mrdh.nl/knooppunt-beurs

Nielsen, O. A., Eltved, M., Anderson, M. K., & Prato, C. G. (2021). Relevance of detailed transfer attributes in large-scale multimodal route choice models for metropolitan public transport passengers. *Transportation Research Part A: Policy and Practice*, *147*, 76–92. https://doi.org/10.1016/J.TRA.2021.02.010

Picard, R. R., & Cook, R. D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, *79*, 583. https://doi.org/10.2307/2288403

Raveau, S., Guo, Z., Muñoz, J. C., & Wilson, N. H. (2014). A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. *Transportation Research Part A: Policy and Practice*, *66*. https://doi.org/10.1016/j.tra.2014.05.010

Raychaudhuri, S. (2008). Introduction to monte carlo simulation. *Proceedings - Winter Simulation Conference*, 91–100. https://doi.org/10.1109/WSC.2008.4736059

RET. (2022a). *Aangepaste dienstregeling metro vanaf 26/9*. https://www.ret.nl/home/reizen/dienstregeling/aangepaste-dienstregeling.html

RET. (2022b). *Mag mijn fiets mee in de bus, tram of metro?* https://www.ret.nl/vraag-antwoord/de-reis/reisinformatie/algemeen-reisinformatie/mag-mijn-fiets-mee-in-de-bus-tram-of-metro.html

Richardson, A. J., Ampt, E. S., & Meyburg, A. H. (1995). *Survey methods for transport planning*. Eucalyptus Press.

Rotstein, A., Inbar, O., Berginsky, T., & Meckel, Y. (2005). Preferred transition speed between walking and running: Effects of training status. *Medicine and Science in Sports and Exercise*, *37*, 1864–1870. https://doi.org/10.1249/01.mss.0000177217.12977.2f

Scipy. (2022). *Scipy.stats.skew*. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.skew.html

Tranmer, M., Murphy, J., Elliot, M., & Pampaka, M. (2020). *Multiple linear regression* (2nd ed.). Cathie Marsh Institute Working Paper 2020-01. https://hummedia.manchester.ac.uk/institutes/cmist/a

UITP. (2019). *Metros: The backbone of mobile communities and sustainable cities*. https://cms.uitp.org/wp/wp-content/uploads/2021/04/Knowledge-Brief-Benefits-of-metro_web.pdf

van den Heuvel, J., & Hoogenraad, J. H. (2014). Monitoring the performance of the pedestrian transfer function of train stations using automatic fare collection data. *Transportation Research Procedia*, *2*, 642–650. https://doi.org/10.1016/J.TRPRO.2014.09.107

van den Heuvel, J., Thiellier, E., & van Gerwn, N. (2013). Privacy by design bij reizigersmetingen op stations, 17–21.

van den Heuvel, J., Ton, D., & Hermansen, K. (2016). Advances in measuring pedestrians at dutch train stations using bluetooth, wifi and infrared technology. *Traffic and Granular Flow '15*, 11–18. https://doi.org/10.1007/978-3-319-33482-0_2

van den Heuvel, J., Voskamp, A., Daamen, W., & Hoogendoorn, S. P. (2015). Using bluetooth to estimate the impact of congestion on pedestrian route choice at train stations. *Traffic and Granular Flow '13*, 73–82. https://doi.org/10.1007/978-3-319-10629-8_9

Weidmann, U. (1993). *Transporttechnik der fussgänger*. ETH Zürich. https://doi.org/10.3929/ethz-b-000242008

Yin, J., D'Ariano, A., Wang, Y., Yang, L., & Tang, T. (2021). Timetable coordination in a rail transit network with time-dependent passenger demand. *European Journal of Operational Research*, *295*, 183–202. https://doi.org/10.1016/J.EJOR.2021.02.059

Zheng, J., Ji, C., & Gao, L. (2014). Research on walking time of transfer pedestrians in urban rail transit station. *Applied Mechanics and Materials*, *610*, 1053–1056. https://doi.org/10.4028/www.scientific.net/AMM.610.1053

Zhou, Y., Yao, L., Gong, Y., & Chen, Y. (2016). Time prediction model of subway transfer. *SpringerPlus*, *5*, 1–11. https://doi.org/10.1186/s40064-016-1686-7

Zhu, W., Fan, W., Wei, J., & Fan, W. ". (2020). Complete estimation approach for characterizing passenger travel time distributions at rail transit stations. *Journal of Transportation Engineering, Part A: Systems*, *146*, 04020050. https://doi.org/10.1061/JTEPBS.0000375

Zhuang, Y., Schadschneider, A., Cheng, H., & Yang, L. (2018). Estimating escalator vs stairs choice behavior in the presence of entry railing: A field study. *KSCE Journal of Civil Engineering*, *22*, 5203–5214. https://doi.org/10.1007/s12205-017-1630-6

A

Data collection forms

Form	Segment type
1	Corridors
2	Vertical transport points
3	Platforms

Maak timers | Delete timers | Delete Data

Day / Time	30-10-22 van 14:50 tot 15:05
Segment	VO1
Remark	

Instructions sheet (open on main desktop screen, otherwise buttons won't match with cells
Add numbers into column sample, same as you want to collect
Run "maak timers"to make timers until last row where in column "sample"a value is given. (Takes a
Press 'start' and 'stop 'to use timer, calculate automatically the walking time of the sample

Gender | Luggage | Group | Crowding

Sample	Start timer	Stop timer		Gender M / F			Luggage NS / L			Group size A / IG			Crowding F / C		Time	Remark (group size, walk-ability) etc
1	Start	Stop		M	F		NS	L		A	IG		F	C		
2	Start	Stop		M	F		NS	L		A	IG		F	C		
3	Start	Stop		M	F		NS	L		A	IG		F	C		
4	Start	Stop		M	F		NS	L		A	IG		F	C		
5	Start	Stop		M	F		NS	L		A	IG		F	C		
6	Start	Stop		M	F		NS	L		A	IG		F	C		
7	Start	Stop		M	F		NS	L		A	IG		F	C		
8	Start	Stop		M	F		NS	L		A	IG		F	C		
9	Start	Stop		M	F		NS	L		A	IG		F	C		
10	Start	Stop		M	F		NS	L		A	IG		F	C		

Maak timers	Day / Time	30-10-22 tussen 15:30-16:00	
Delete timers	Segment	Beurs, lift/trap DE naar hal	E 00:00:16
	Remark	lift/trap in geheel timen, mik op ca 10 samples met de lift	L 00:00:25
Delete Data			

Gender · Luggage · Group · Crowding · VT modes · Wait

Instructions sheet (open on main desktop screen, otherwise buttons won't match with cells)

Add numbers into column sample, same as you want to collect
Run "maak timers"to make timers until last row where in column "sample"a value is given. (Takes a while)
Press 'start' and 'stop 'to use timer, calculate automatically the walking time of the sample
Fill in escalator and/or lift travel time in cells M3-M4 if needed

Sample	Start timer	Stop timer		Gender M / F	Luggage NS / L	Group size A / IG	Crowding F / C	VT mode S / E / L	Wait D / W	Time	Remark (group size, walk-ability) etc	Final time
1	Start	Stop		M F	NS L	A IG	F C	S E L	D W			
2	Start	Stop		M F	NS L	A IG	F C	S E L	D W			
3	Start	Stop		M F	NS L	A IG	F C	S E L	D W			
4	Start	Stop		M F	NS L	A IG	F C	S E L	D W			
5	Start	Stop		M F	NS L	A IG	F C	S E L	D W			
6	Start	Stop		M F	NS L	A IG	F C	S E L	D W			
7	Start	Stop		M F	NS L	A IG	F C	S E L	D W			
8	Start	Stop		M F	NS L	A IG	F C	S E L	D W			
9	Start	Stop		M F	NS L	A IG	F C	S E L	D W			
10	Start	Stop		M F	NS L	A IG	F C	S E L	D W			

Instructions sheet (open on main desktop screen, otherwise buttons won't match with cells
Add numbers into column sample, same as you want to collect
Run "maak timers"to make timers until last row where in column "sample"a value is given. (Takes a
Press 'start' and 'stop 'to use timer, calculate automatically the walking time of the sample
Button "Time" in column starts timer for all cells, "Reset" resets all timers

Maak timers | Delete timers | Delete Data

Day / Time	30-10-22 tussen 13:30-13:45
Segment	Beurs, ABC west
Remark	Min 30 samples

Gender | Luggage | Group | Crowding | Alighting

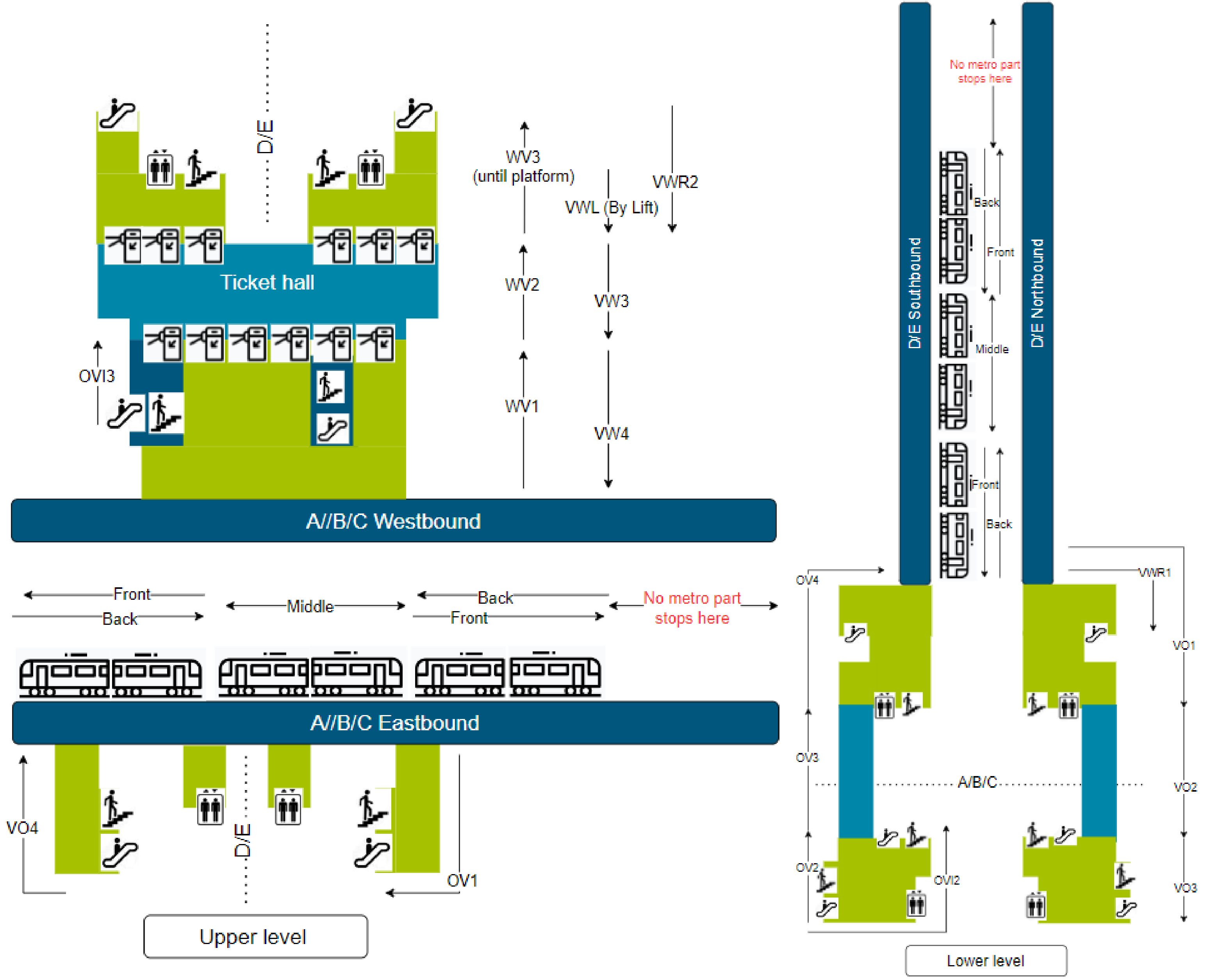
Sample	Start timer	Stop timer		Gender M / F			Luggage NS / L			Group size A / IG			Crowding F / C			Alighting F / M / B			Time	Remark (group size, walk-ability) etc
1	Start	Stop		M	F		NS	L		A	IG		F	C		F	M	B		
2	Start	Stop		M	F		NS	L		A	IG		F	C		F	M	B		
3	Start	Stop		M	F		NS	L		A	IG		F	C		F	M	B		
4	Start	Stop		M	F		NS	L		A	IG		F	C		F	M	B		
Time 5	Start	Stop		M	F		NS	L		A	IG		F	C		F	M	B		
6	Start	Stop		M	F		NS	L		A	IG		F	C		F	M	B		
Reset 7	Start	Stop		M	F		NS	L		A	IG		F	C		F	M	B		
8	Start	Stop		M	F		NS	L		A	IG		F	C		F	M	B		
9	Start	Stop		M	F		NS	L		A	IG		F	C		F	M	B		
Time 10	Start	Stop		M	F		NS	L		A	IG		F	C		F	M	B		

B

Segments in Beurs station

Schematic overview Beurs station

Platform segments for alighting, are based on 6 car metro trains.



Segment: Platform A/B/C Eastbound

Start timing: Metro doors open (for all samples).
End timing: Red Line (Eastern exit) or Orange Line (Western exit)
Standing location: Arrow, Middle at platform.

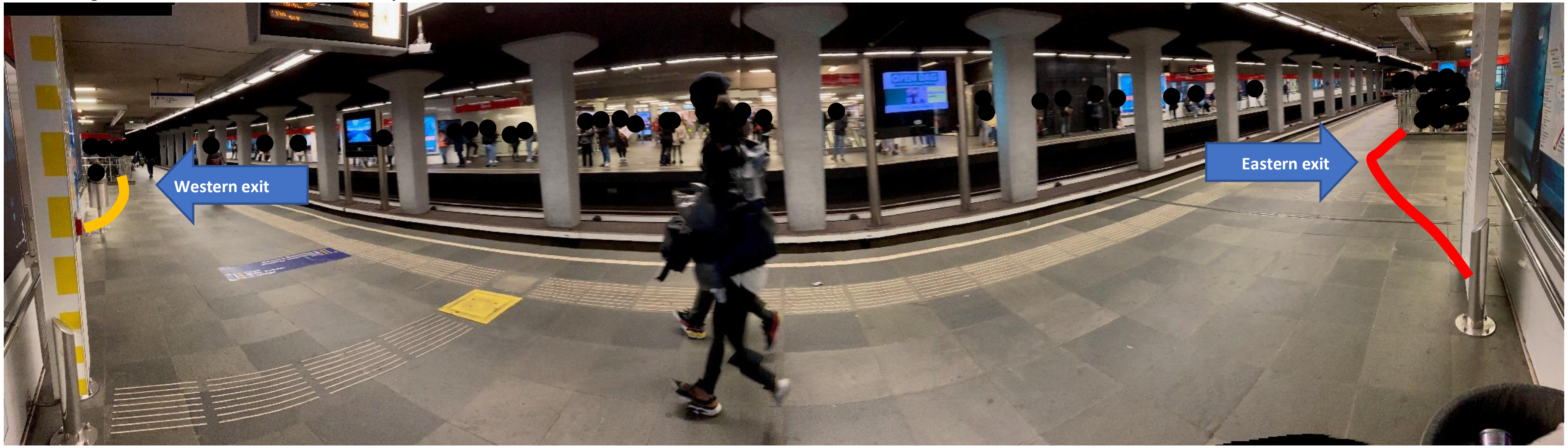


Figure 1-1: Viewpoint from standing location.

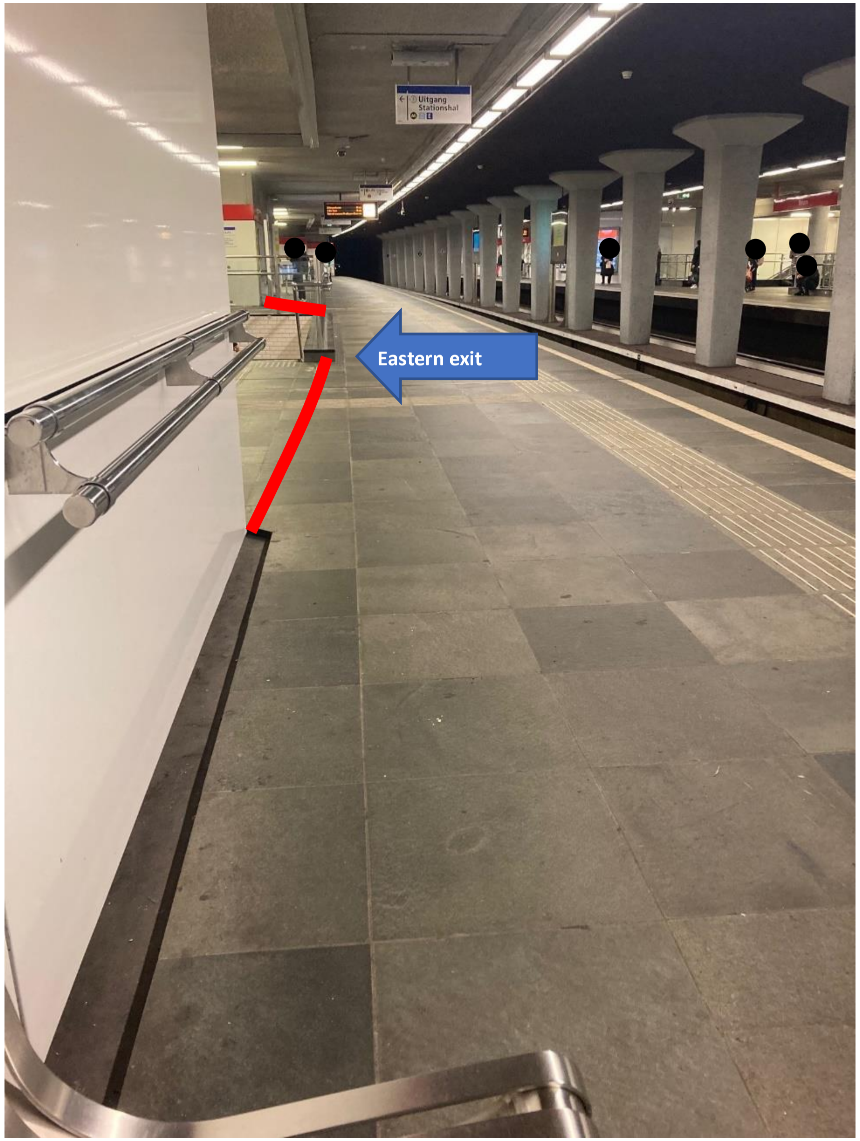


Figure 1-2: Boundary eastern exit to stair/escalator.

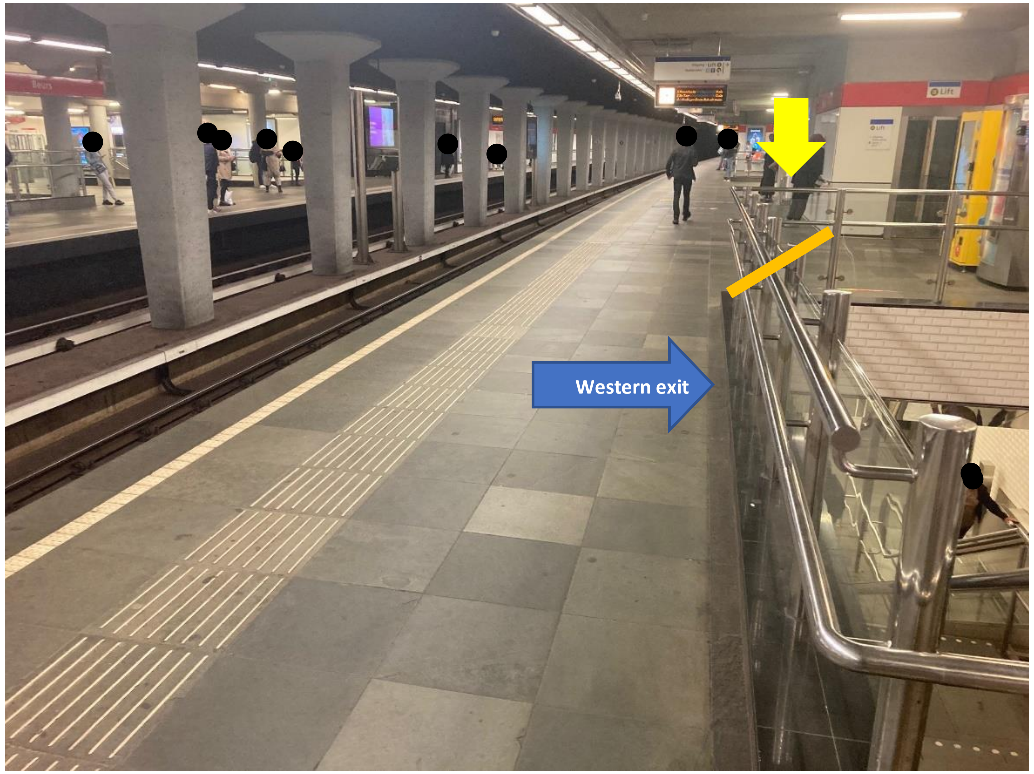


Figure 1-3: Boundary western exit lift.

Figure 1-4/1-5: Boundaries.

Segment: Platform A/B/C Westbound

Start timing: Metro doors open (for all samples).
End timing – Red Line:
Standing location: Arrow

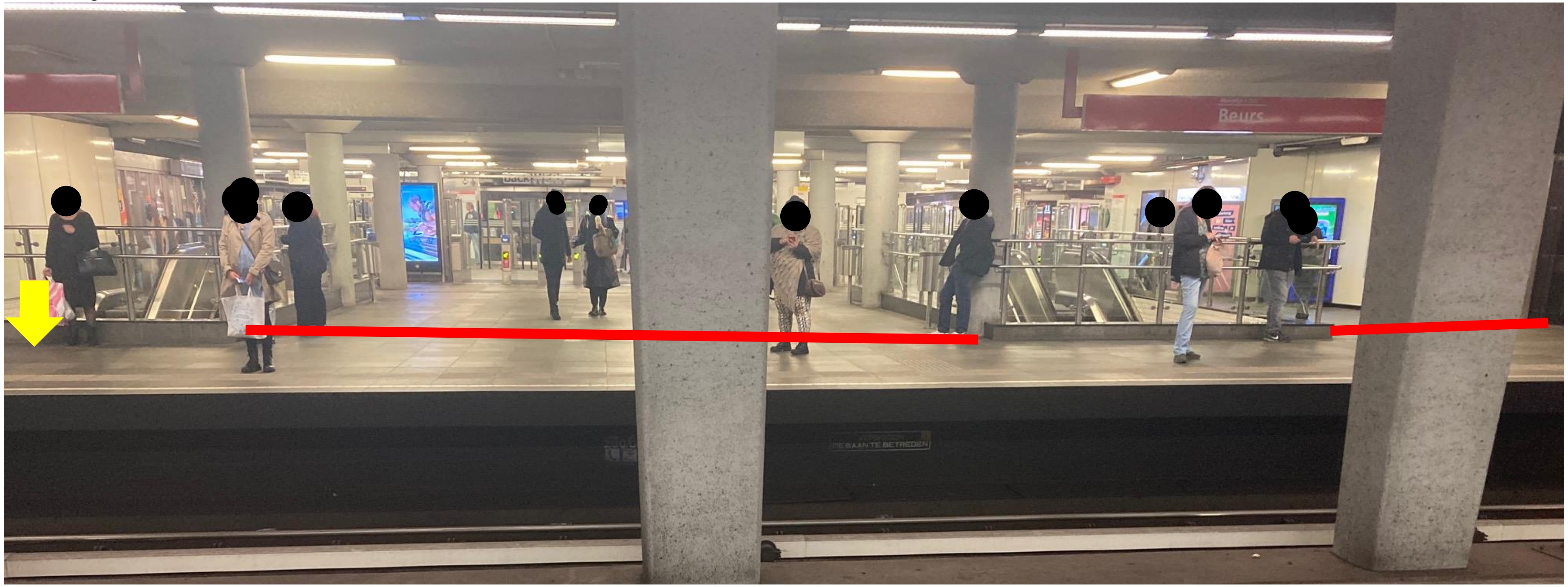


Figure 2: Boundary to stop timing when sample leaves platform and standing location.

Segment: Platform D/E Southbound

Start timing: Metro doors open (for all samples).
End timing: Red line
Standing location: Arrow

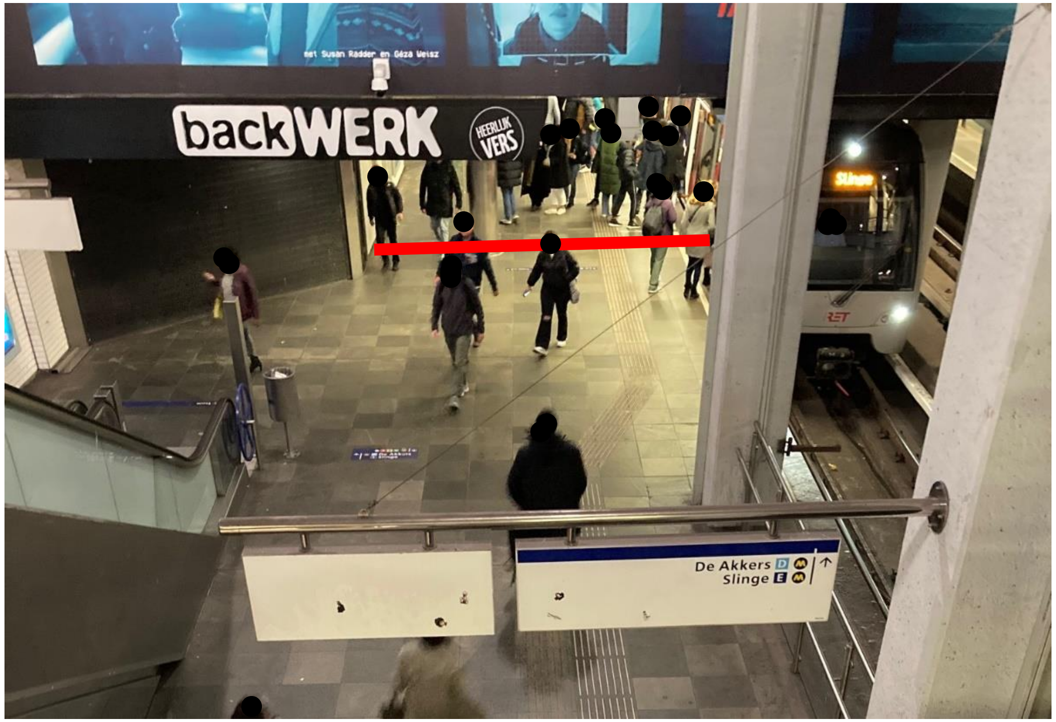


Figure 3-1: Boundary to stop timing.

Figure 3-2: Standing location and boundary.

Segment: Platform D/E Northbound

Start timing: Metro doors open (for all samples).
End timing: Red Line
Standing location: Arrow

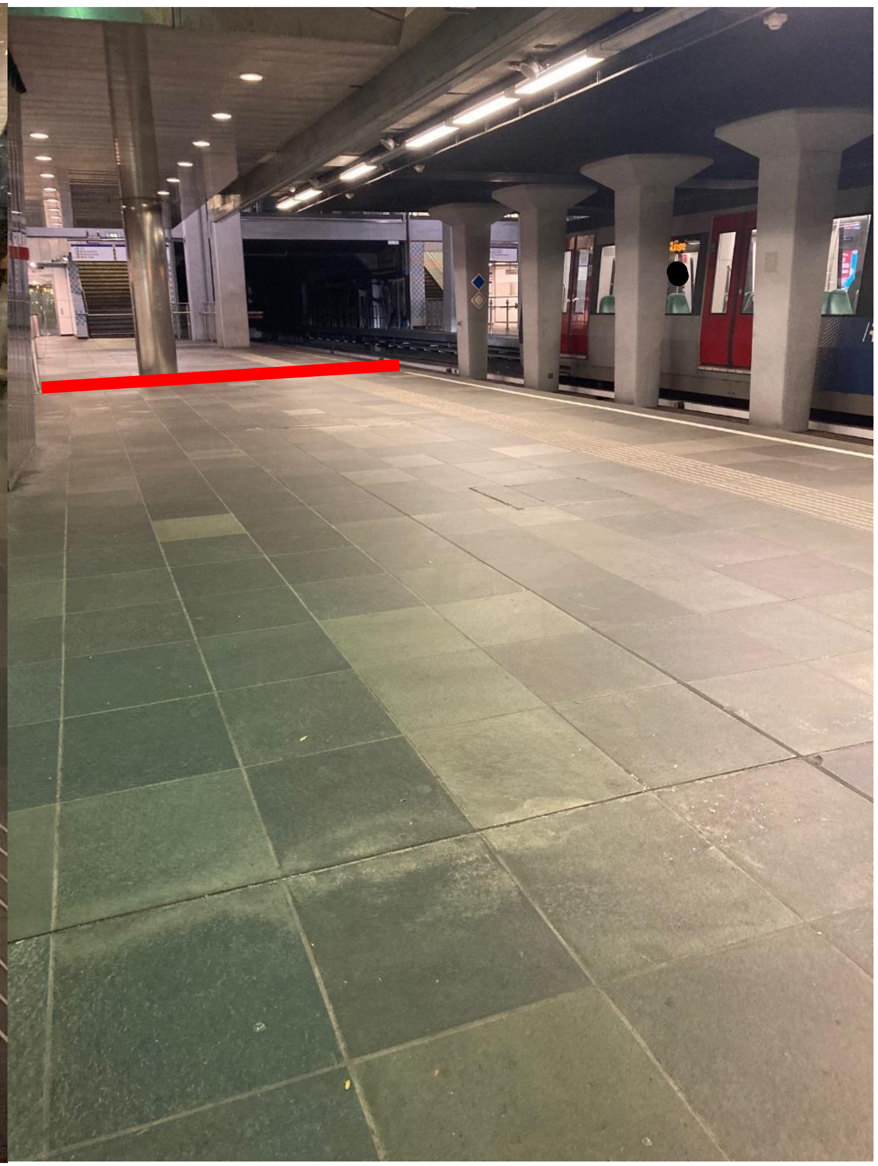


Figure 4-1/4-2: Boundary to stop timing and standing location.

Segment: OV1/VO4

Green line: OV1) Start timing, VO4) Stop timing
Red line (lower end balcony of stairs: OV1) Stop timing (for lift, when doors close) , VO4) End timing
Orange line: OV1) Stop timing when sample takes escalator
Standing location: Arrow

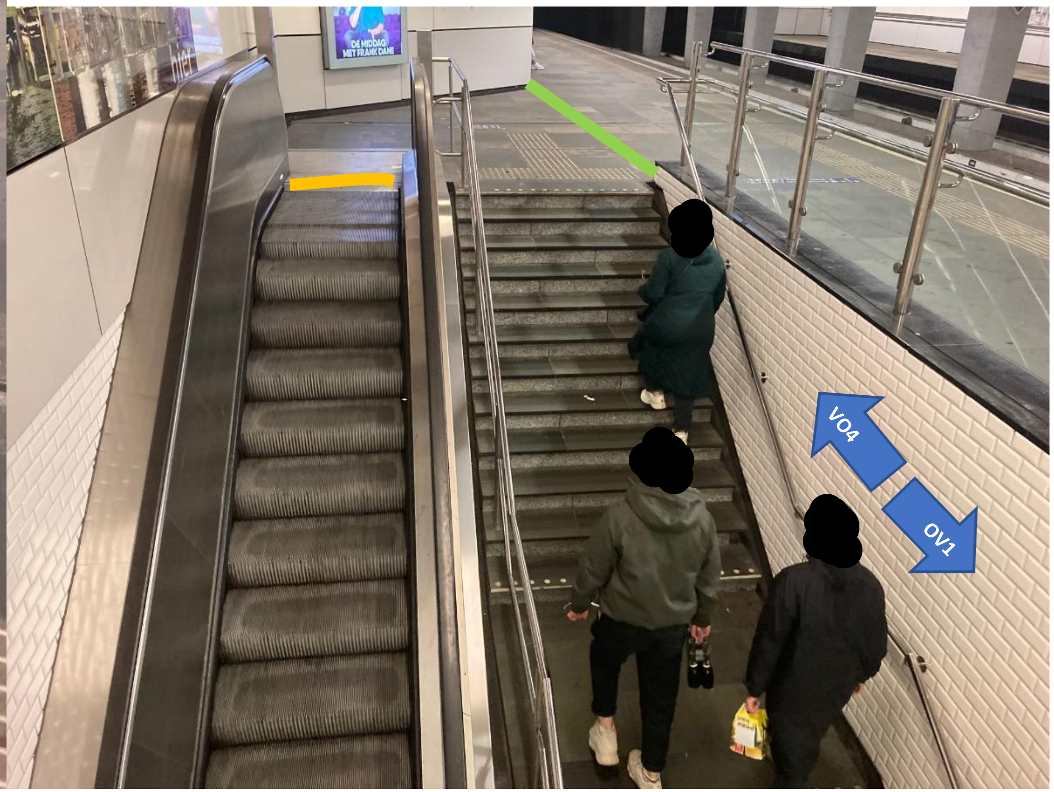


Figure 5-1 (left): Green boundary eastern exit and standing location
Figure 5-2 (right): Green boundary western exit and vie from standing location

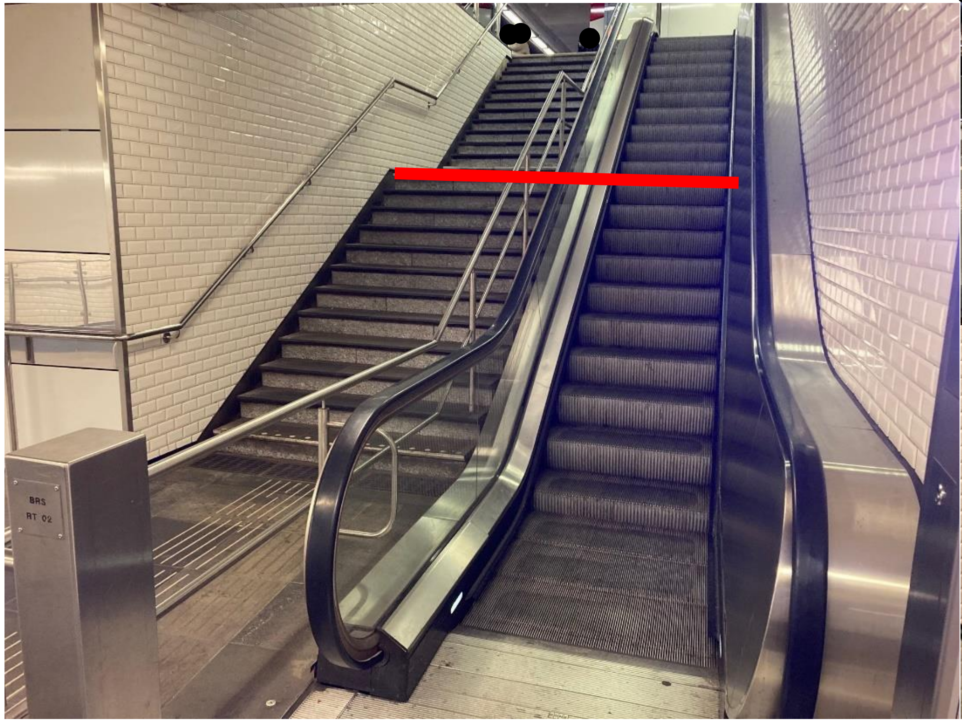



Figure 5-3 (left): Boundaries lower end, passing lower end of stairs balcony.
Figure 5-4 (right): Viewpoint for standing location.

Segment: OV2/OVI2/VO3

Green line: OV2/OVI2) Start timing; VO3) Stop timing (lift and stairs only)
Red line: OV2) Stop timing; VO3) Start timing
Blue line: OVI2) Stop timing
Orange line: VO3) Stop timing

Standing location: Arrow

Figure 6-1(left) & 6-2 (right): Western corridor

Figure 6-3: Eastern corridor from standing location.

Figure 6-4/6-5: More of western corridor.

Segment: OVI3

Green line: Start timing
Red line (center fare gate): Stop timing
Standing location: Arrow

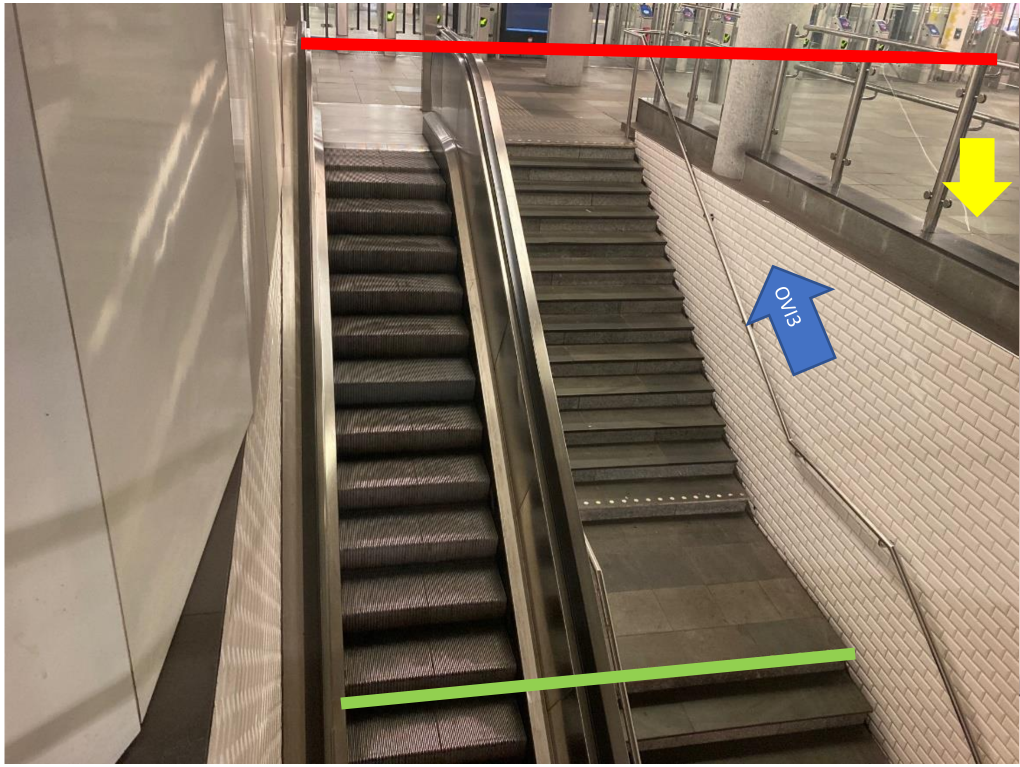


Figure 7-1: Segment boundaries

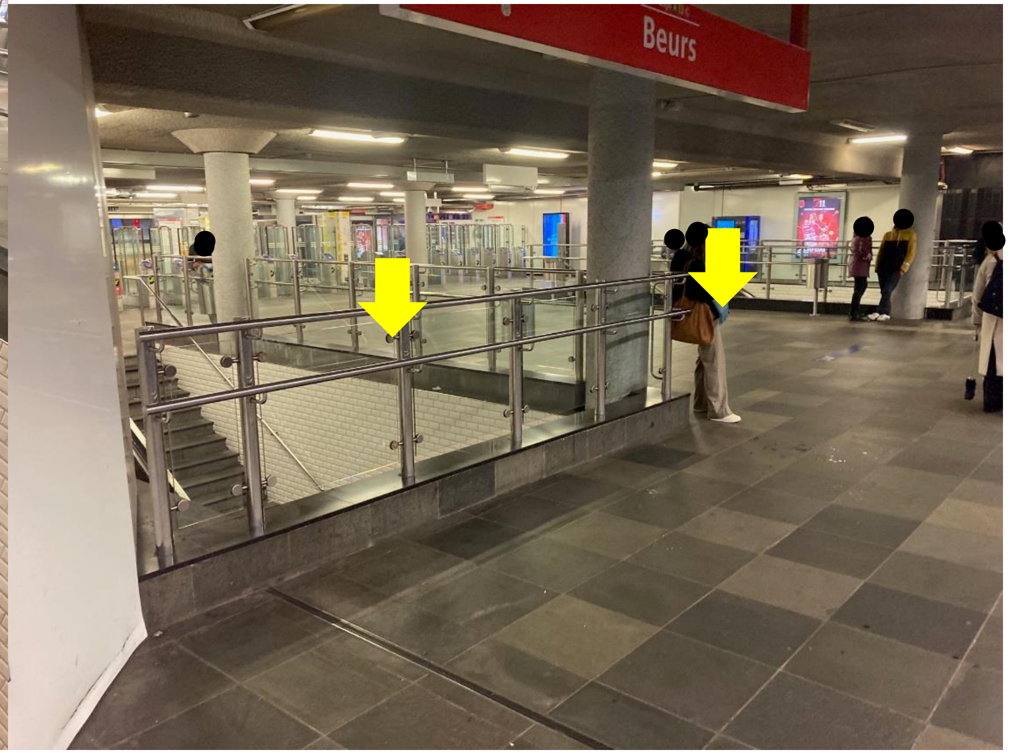


Figure 7-2: Standing locations for both segments.

Segment: OV3/VO2

Green line:	OV3) Start timing	VO2) End timing
Red line:	OV3) Stop timing	VO2) Start timing

Standing location: Arrow

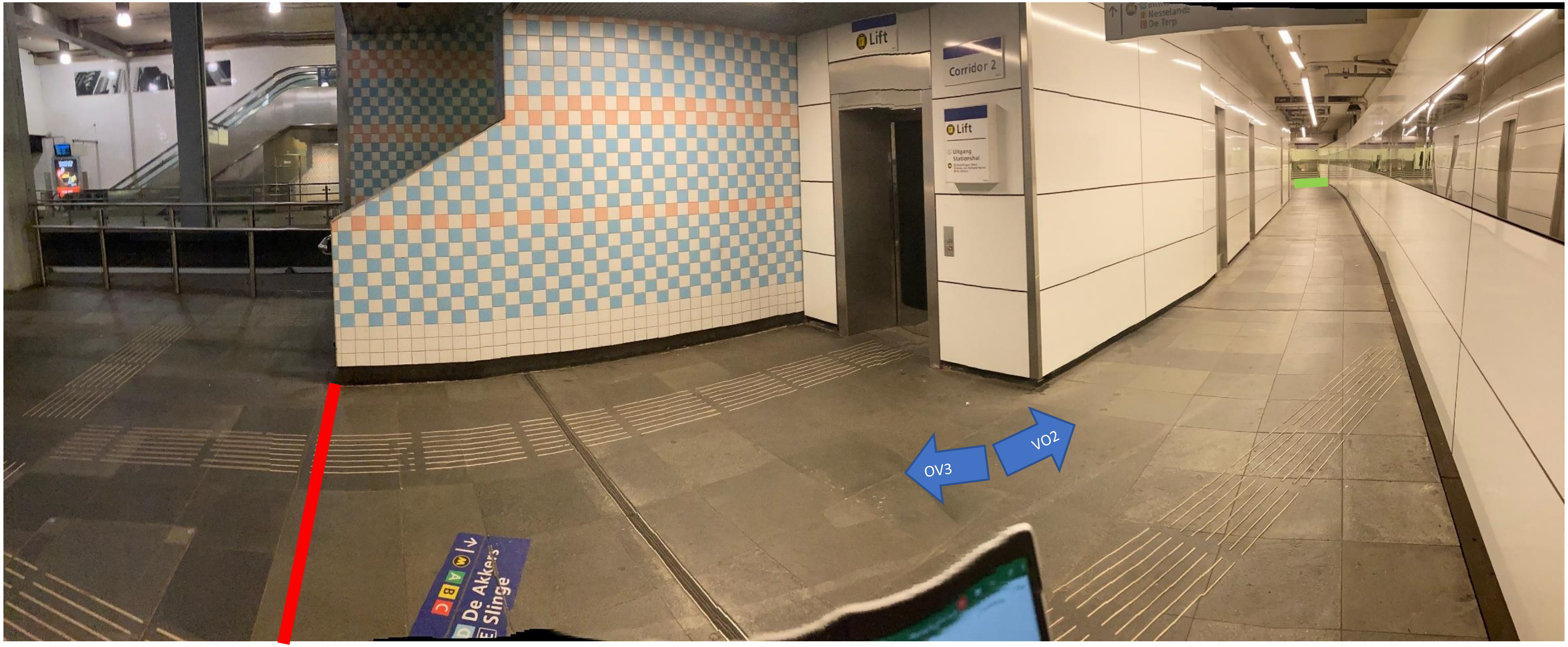


Figure 8-1: Western corridor overview from standing location

Figure 8-2: Western corridor overview and standing position.

Figure 8-3: Zoom in on green boundary.

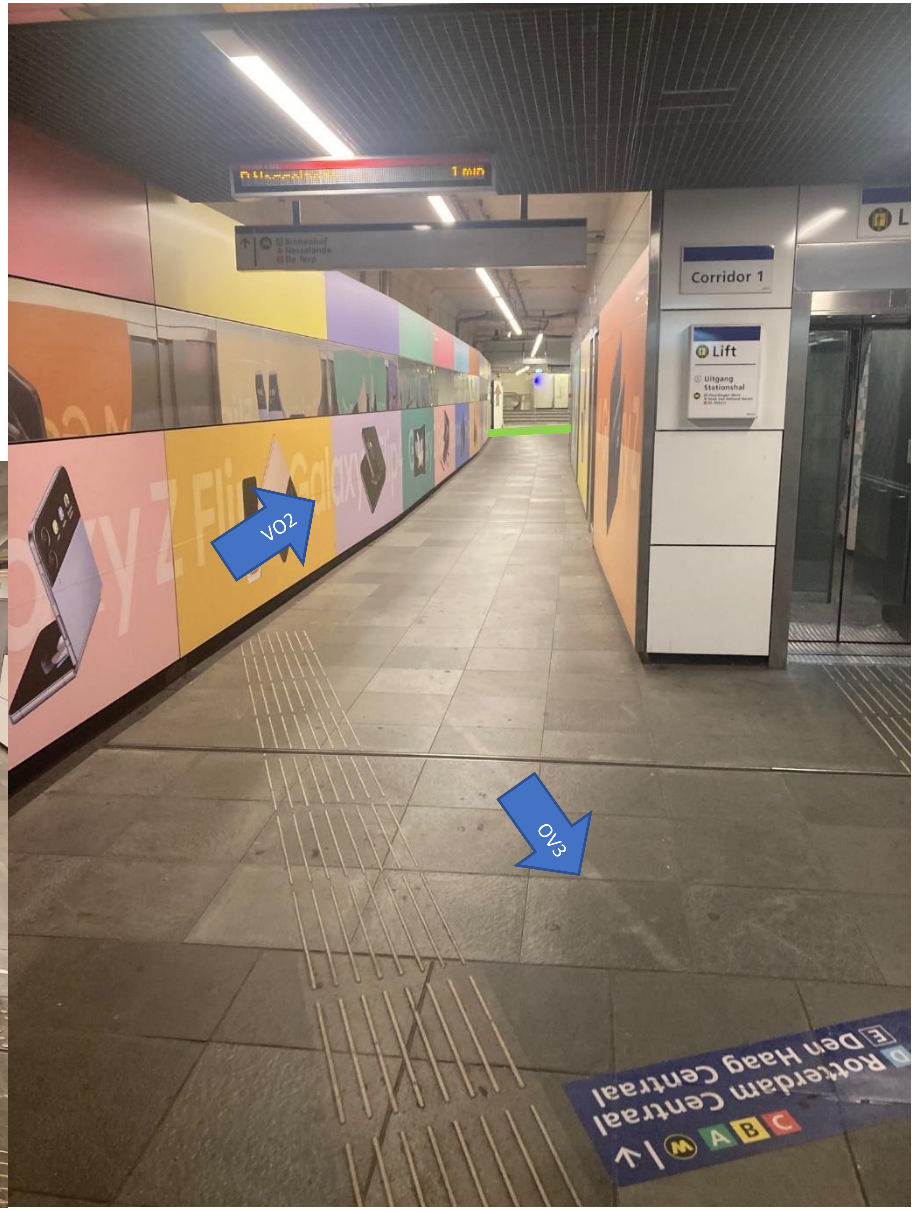
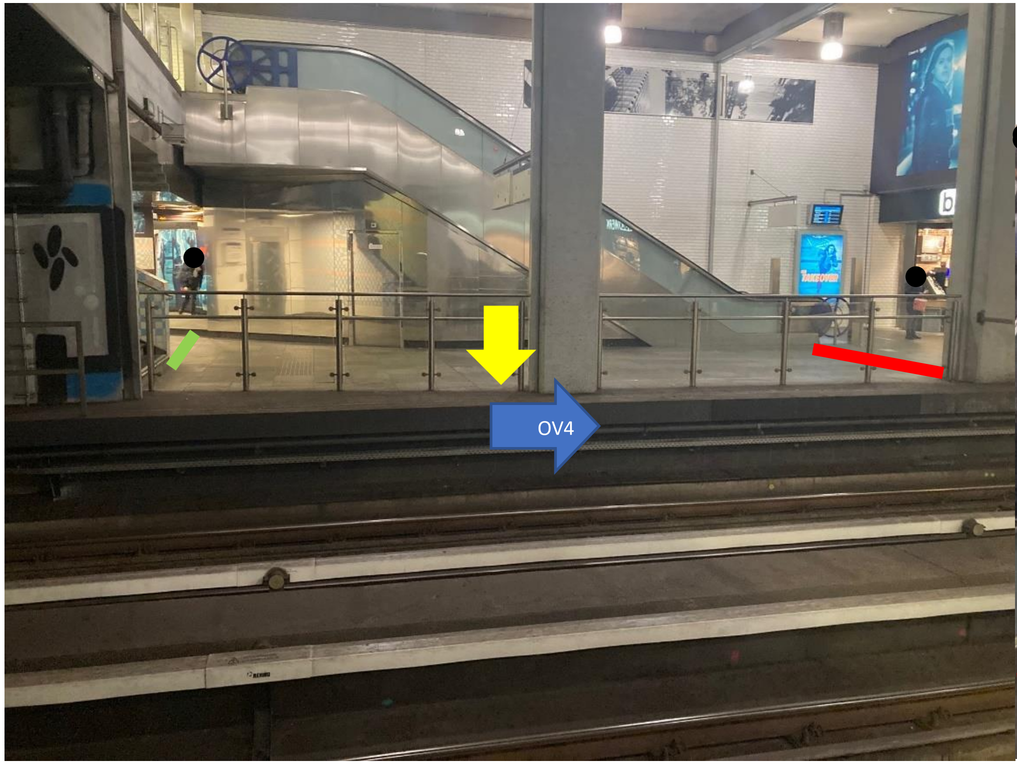


Figure: 8-4/8-5 (Left): Eastern corridor

Segment: OV4/VO1

Green line: OV4) Start timing VO1) Stop timing
Red line: OV4) Stop timing
Blue line: VO1) Start timing
Standing location: Arrow



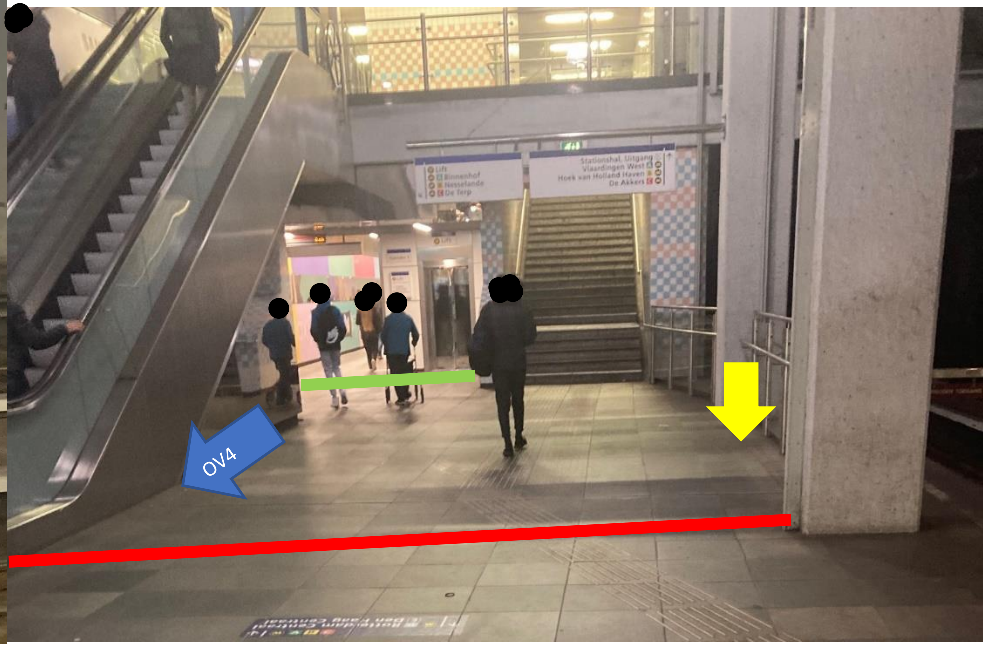


Figure 9-1/9-2: Overview segment and standing locations.

Figure 9-3: Standing location overview (and standing location for segment OV3/VO2).

Segment WV1/VW4:

Green line:	WV1) Start timing	VW4) Stop timing or when sample stopped to wait in segment
Red line (center of fare gate):	WV1) Stop timing	VW4) Start timing
Standing location: Arrow		

Figure 10-1/10-2: Segment overview and standing location:

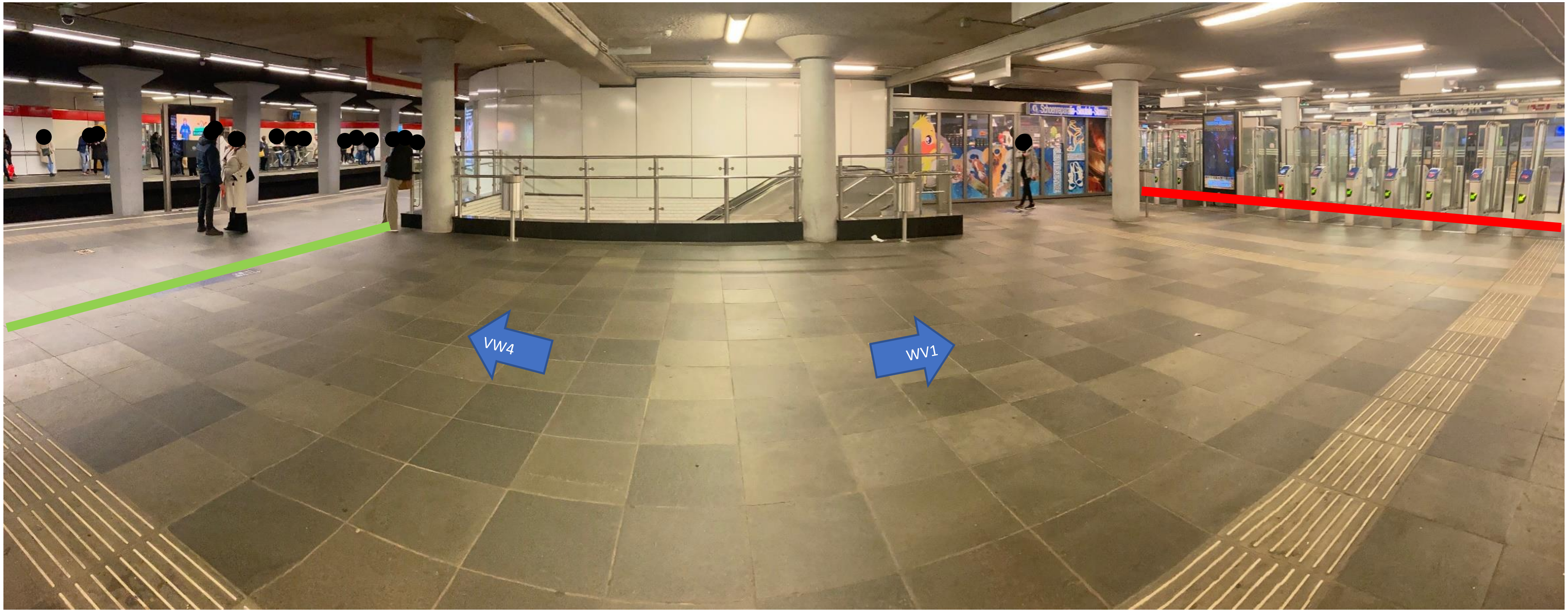


Figure 10-3: Segment overview

Segment WV2/VW3:

Green line (center fare gates):	WV2) Start timing	VW3) Stop timing
Red line (center fare gates):	WV2) Stop timing	VW3) Start timing
Standing location: Arrow		

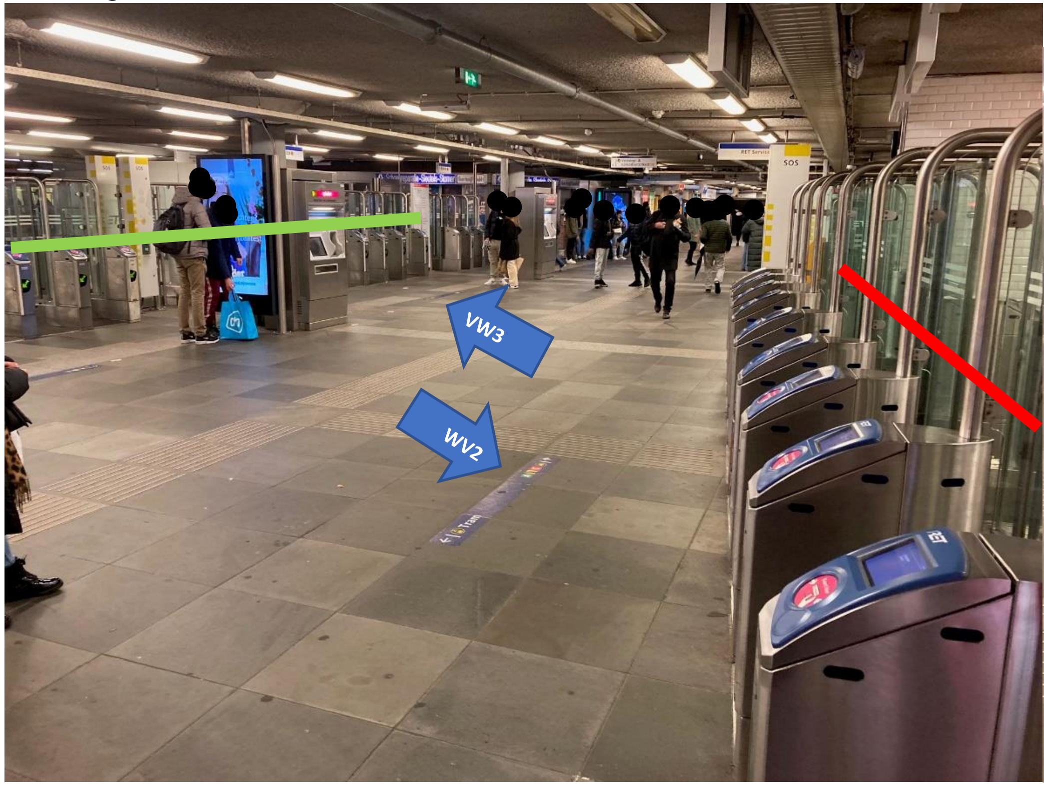


Figure 10-4: Eastern side, right fare gates to D/E Northbound

Figure 10-5: Western side, right fare gates to D/E Southbound

Segment WV3:

Start timing: Green line (center of fare gates)
Stop timing: Red lines

Figure 11: Overview from standing location.

Segment VWR2/VWL:

Start timing: VWR2) Green line VWL) Edge of segment VO1
Stop timing: (Both) Red line (center of fare gates)

Figure 12: Overview from standing location

Segment VWR1:

Start timing: Green line (a black joint in the platform)
Stop timing: Red line (samples board escalator)

Figure 13-1: Eastern side segment from standing location.

Figure 13-2: Western side from standing location.

C

Pre-pilot and pilot data collection results

In this Appendix, the pre-pilot and pilot data collection results are given.

Pre-pilot: Travel times escalators and lifts in Beurs

The escalators to/from A/B/C Eastbound and to/from D/E platforms have the same height but a different travel time. The travel times from Table C.1 are not always used in each segment.

Table C.1: Travel times for lifts and escalators in segments.

Segment	Vertical transport mode	Travel time
OV1/VO3, : To/from A/B/C Eastbound platform	Escalator Lift	15 s 20 s
WV3/VWR1/VWL: To/from D/E platforms	Escalator Lift	16 s 26 s
OVI2: Under A/B/C to ticket hall	Escalator	20 s

Some vertical transport segments were split up, and some were not (see Appendix B). A quick summary of the used travel times in each segment and motivation is in Table C.2. The travel times could be used for the escalators because, during the collection, all passengers and thus samples stood still on the escalators.

Table C.2: Used travel times of vertical transport modes in each segment.

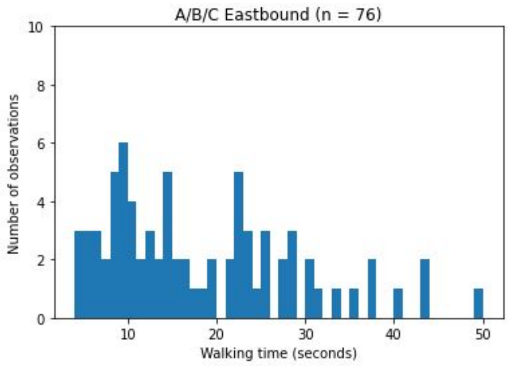
Segment	Used travel time	Motivation
OV1 Escalator	8 s	End boundary was at the halfway point with adjacent stairs.
OV1 Lift	20 s	End boundary was when lift doors opened.
OVI2 Escalator	10 s	End boundary was at the halfway point with adjacent stairs.
VO3 Escalator	8 s	End boundary was at the halfway point with adjacent stairs.
VO3 Lift	20 s	End boundary was when lift doors opened.
WV3 Escalator	16 s	End boundary was at the bottom of the escalator.
WV3 Lift	-	End boundary was beyond the lift.
VWL	-	End boundary was beyond the lift.
VWR1	16 s	End boundary was at the top of the escalator.

Pilot results

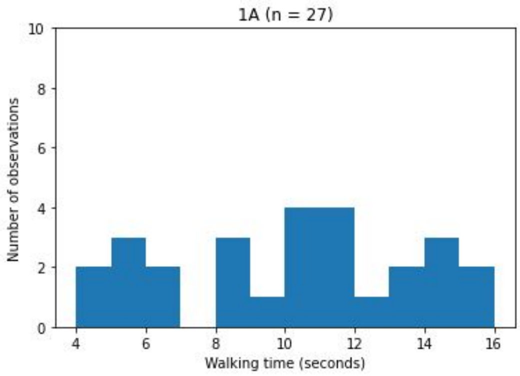
From the pilot at Beurs, Table C.3 shows the results on the number of samples and some descriptive statistics. The histograms of each segment in the pilot is in Figure C.1. Especially in the segments, which are not a platform, the type of distribution on the walking time is unclear from Figure C.1.

Table C.3: Pilot results on walking times (in seconds) per segment.

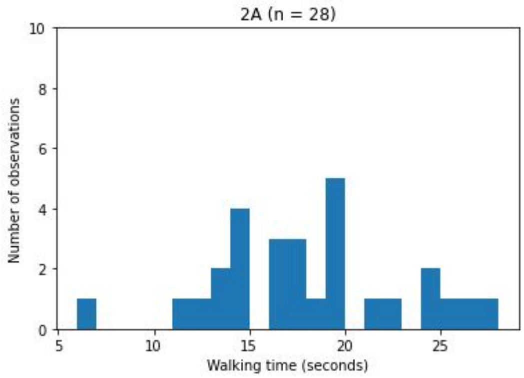
Segment	**Platform**	**OV1**	**OV2**	**OV3**	**OV4**
Number of samples (n)	76	27	28	28	33
Male / Female	35 / 41	13 / 14	11 / 17	15 / 13	16 / 17
No / Large	74 / 2	27 / 0	26 / 2	27 / 0	30 / 3
Alone / In group	67 / 9	24 / 3	25 / 3	23 / 5	26 / 7
Free / Crowded	20 / 56	10 / 17	17 / 11	20 / 7	30 / 3
Front / Middle / Back	35 / 36 / 5	-	-	-	-
Stairs / Escalator / Lift	-	19 / 8 / 0	-	-	-
Direct / Wait	-	11 / 16	-	-	-
Mean walking time ($\bar{x}$)	17.75 s	9.74 s	17.61 s	27.68 s	8.39 s
Standard deviation (s)	10.66 s	3.47 s	4.89 s	5.98 s	2.12 s
Minimal walking time	4 s	4 s	6 s	13 s	3 s
Maximum walking time	50 s	16 s	28 s	42 s	12 s



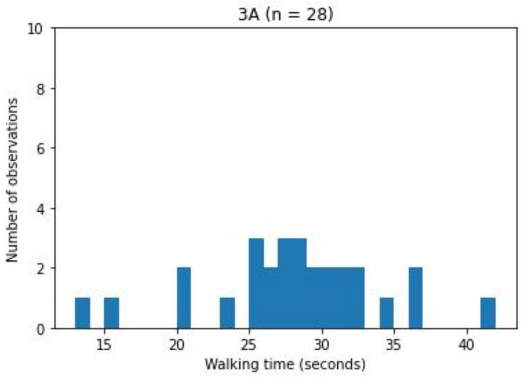
(a) Segment: A/B/C Eastbound platform.



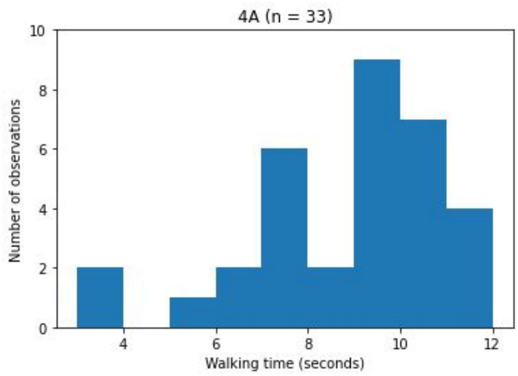
(b) Segment OV1.



(c) Segment OV2.



(d) Segment OV3.



(e) Segment OV4.

Figure C.1: Walking time histograms of pilot segments.

D

Collected sample sizes per segment and variable

Table D.1: Collected sample sizes on platform segments. (Red-marked cells violate the minimal sample size.)

Platform	Min required	D/E NB	D/E SB	A/B/C WB	A/B/C EB to eastern exit	A/B/C EB to western exit
Total (n)	95	109	115	105	108	102
Male	16	56	54	51	49	52
Female	16	53	61	54	59	50
No/Small	5	97	104	97	104	90
Large	5	12	11	8	4	12
Alone	14	86	77	77	95	86
In group	14	23	38	28	13	16
Free	15	25	42	23	33	35
Crowded	15	84	73	82	75	67
Front	14	39	37	36	44	7
Middle	14	32	42	37	58	41
Back	14	38	36	32	6	54

Table D.2: Total number of samples per variable for transfer segments to/from A/B/C West platform. (Red-marked cells violate the minimal sample size) *Segment VWL is a lift-only segment which few passengers use.

Transfer route:		A/B/C WB to D/E			D/E to A/B/C WB					
Segment	**Min required**	WV1	WV2	WV3	VO1	VWR1	VWR2	VWL*	VW3	VW4
Total (n)	95	102	107	100	100	98	102	24	116	107
Male	16	48	54	42	47	43	45	12	51	43
Female	16	54	53	58	53	55	57	12	65	64
No/Small	5	101	105	93	95	95	100	4	115	103
Large	5	1	2	7	5	3	2	20	1	4
Alone	14	81	97	75	77	77	64	14	102	78
In group	14	21	10	25	23	21	38	10	14	29
Free	15	45	21	86	45	63	76	23	77	91
Crowded	15	57	86	14	55	35	26	1	39	16
Stairs	5	-	-	50	-	-	-	-	-	-
Escalator	5	-	-	45	-	98	-	-	-	-
Lift	5	-	-	5	-	-	-	24	-	-
Direct	-	-	-	93	-	85	-	15	-	-
Wait	-	-	-	7	-	13	-	9	-	-

Table D.3: Total number of samples per variable for transfer segments to/from A/B/C East platform.(Red-marked cells violate the minimal sample size)

Transfer route		**A/B/C EB to D/E**						**D/E to A/B/C EB**			
Segment	**Min required**	OV1	OV2	OVI2	OVI3	OV3	OV4	VO1	VO2	VO3	VO4
Total (n)	95	101	100	96	106	100	102	100	109	103	104
Male	16	43	45	43	46	49	43	47	55	46	46
Female	16	58	55	53	60	51	59	53	54	57	58
No/Small	5	81	85	90	103	97	91	95	105	101	101
Large	5	20	15	6	3	3	11	5	4	2	3
Alone	14	81	81	73	83	74	73	77	94	77	86
In group	14	20	19	23	23	26	29	23	15	26	18
Free	15	55	52	30	69	56	88	45	57	81	56
Crowded	15	46	48	66	37	44	14	55	52	22	48
Stairs	5	66	-	53	-	-	-	-	-	20	-
Escalator	5	12	-	43	-	-	-	-	-	82	-
Lift	5	23	-	-	-	-	-	-	-	1	-
Direct	-	44	-	88	-	-	-	-	-	97	-
Wait	-	57	-	8	-	-	-	-	-	6	-

E

Walking time histogram per segment

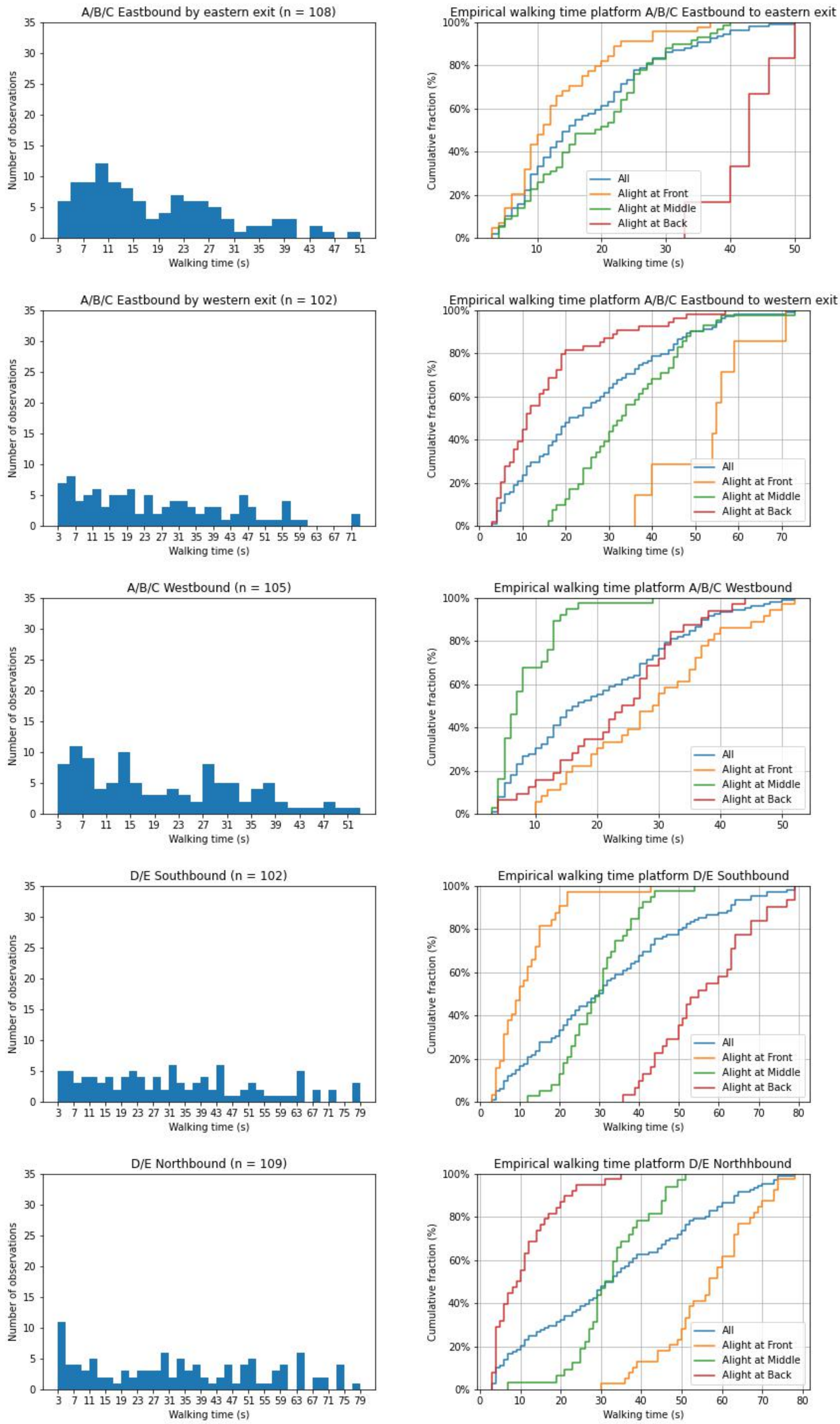


Figure E.1: Walking time histograms of the platform segments.

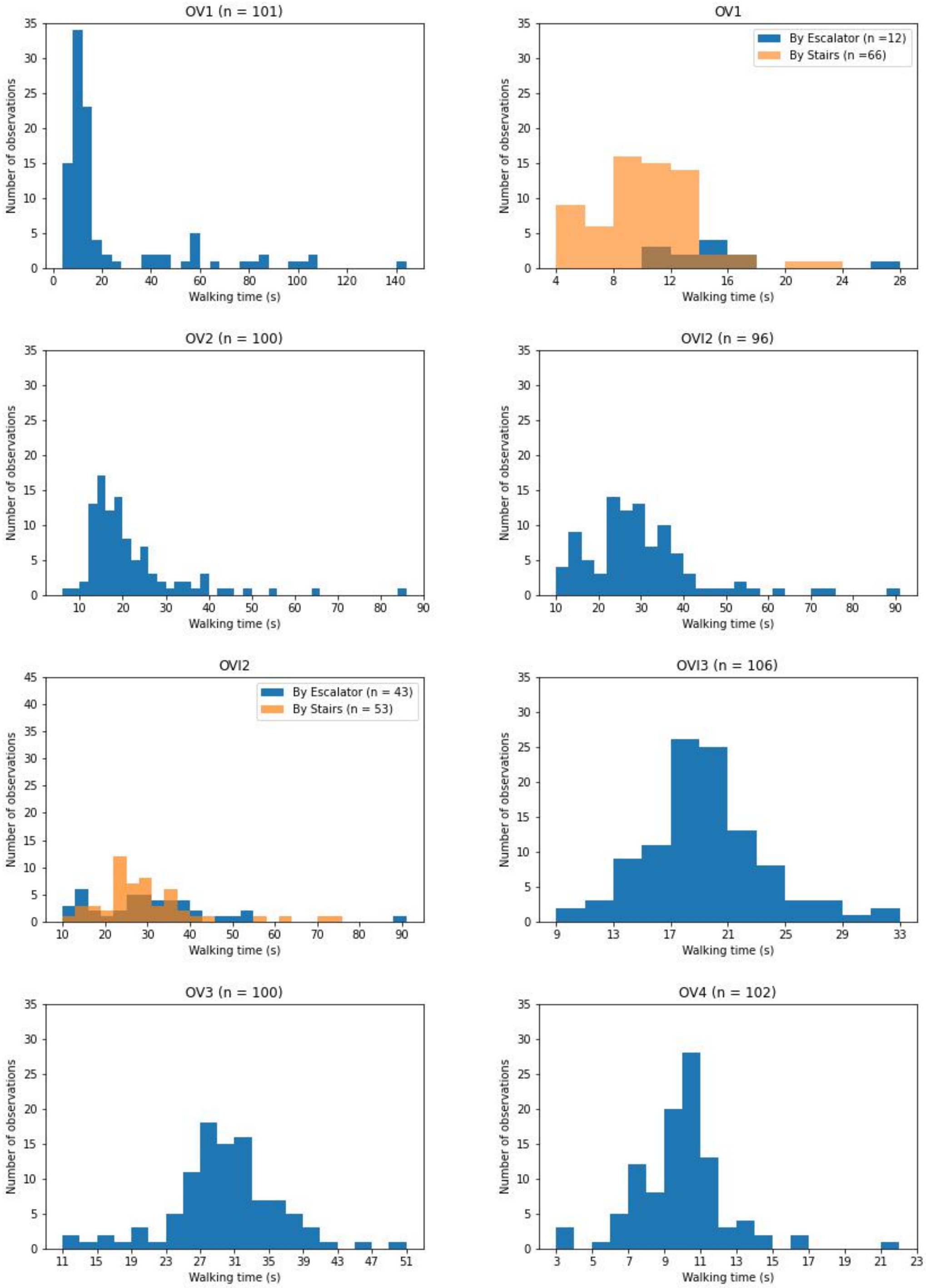


Figure E.2: Walking time histograms of segments from transfer path from A/B/C Eastbound to D/E.

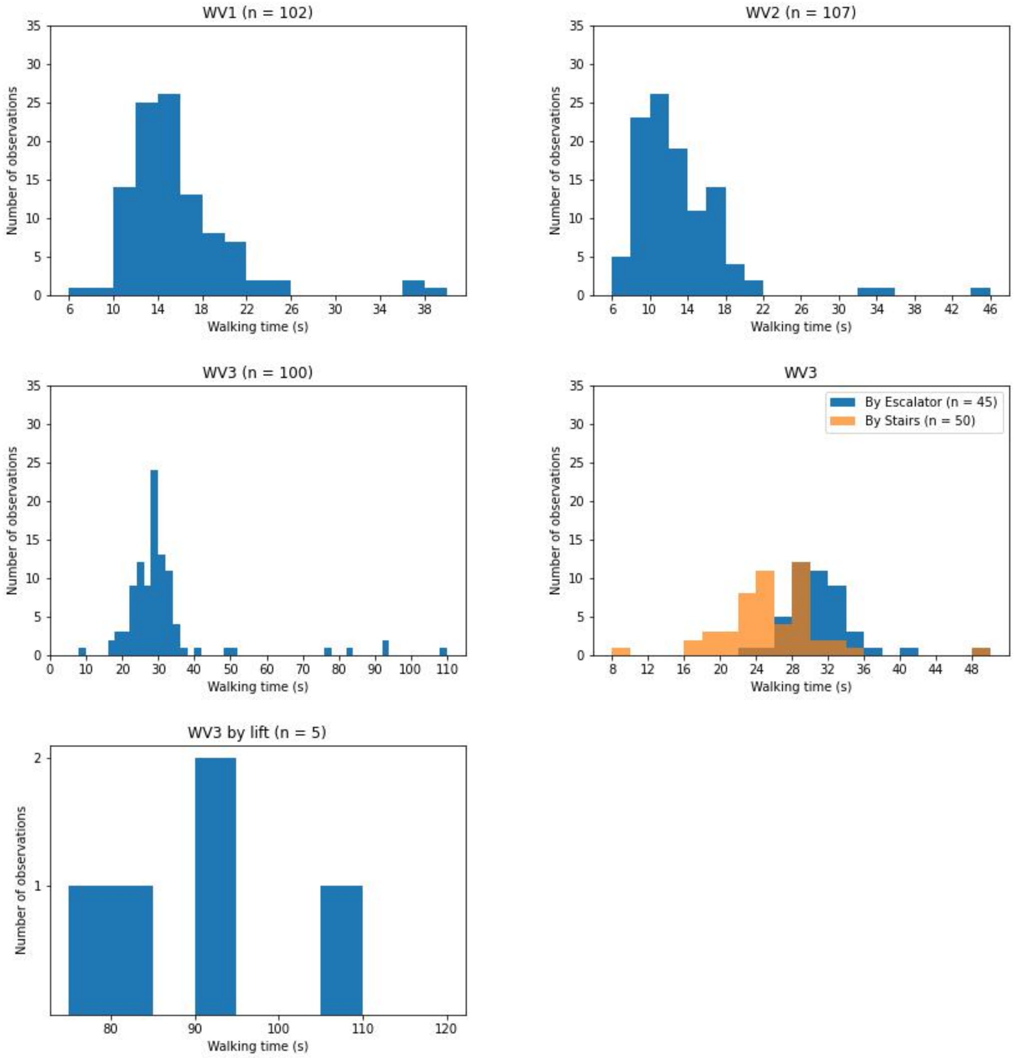


Figure E.3: Walking time histograms of segments from transfer path A/B/C Westbound to D/E.

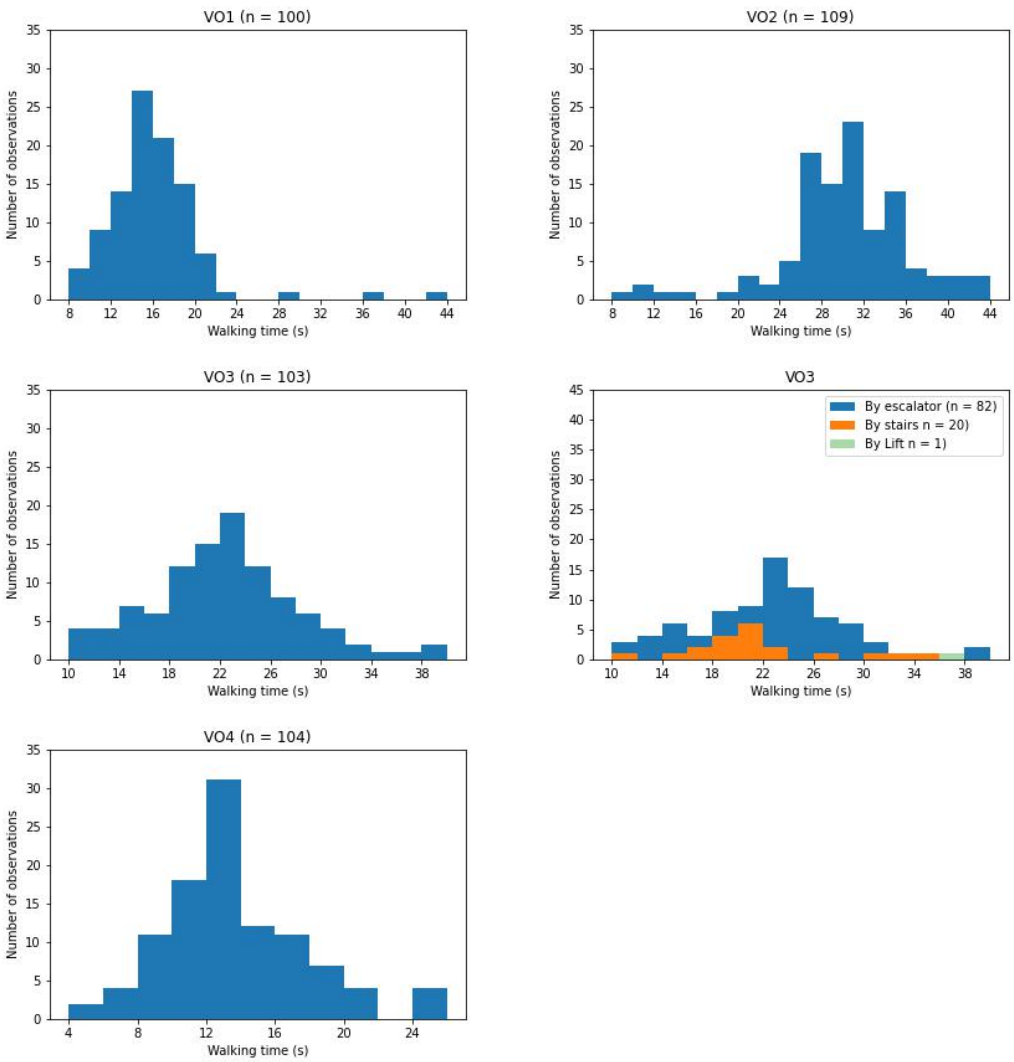


Figure E.4: Walking time histograms of segments from transfer path D/E to A/B/C Eastbound.

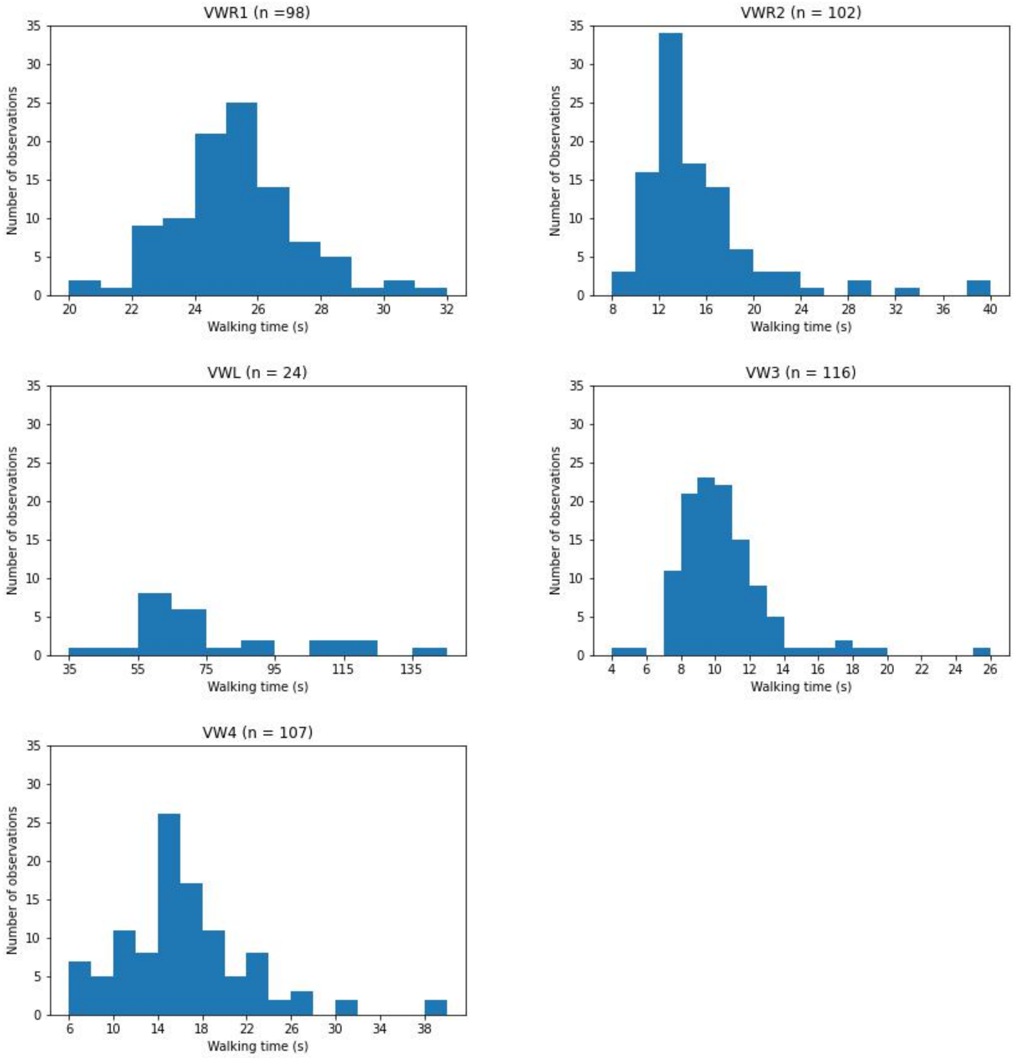


Figure E.5: Walking time histograms of segments from transfer paths D/E to A/B/C Westbound.

F

Boxplots of the walking time for each variable per segment

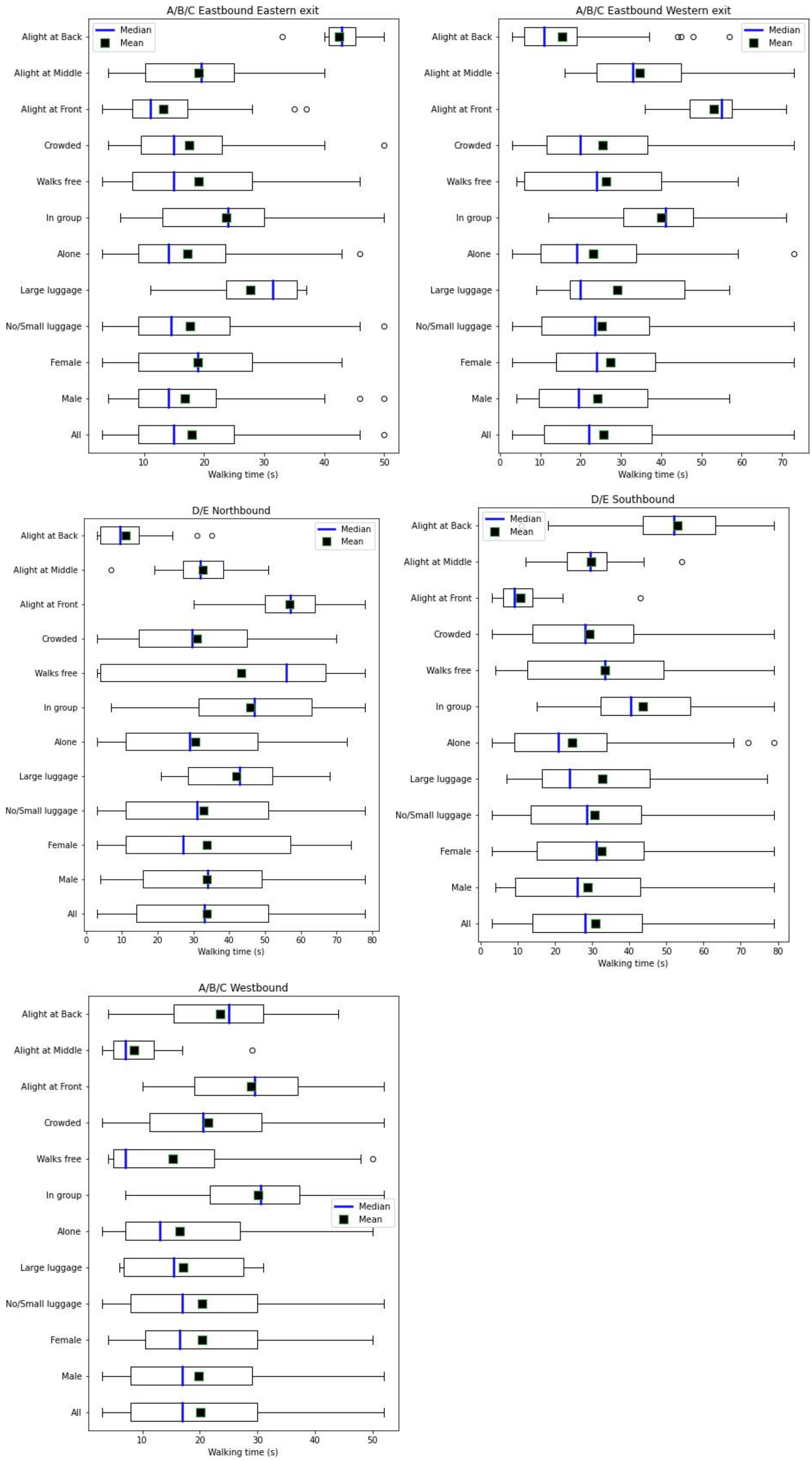


Figure F.1: Boxplots of platform segments.

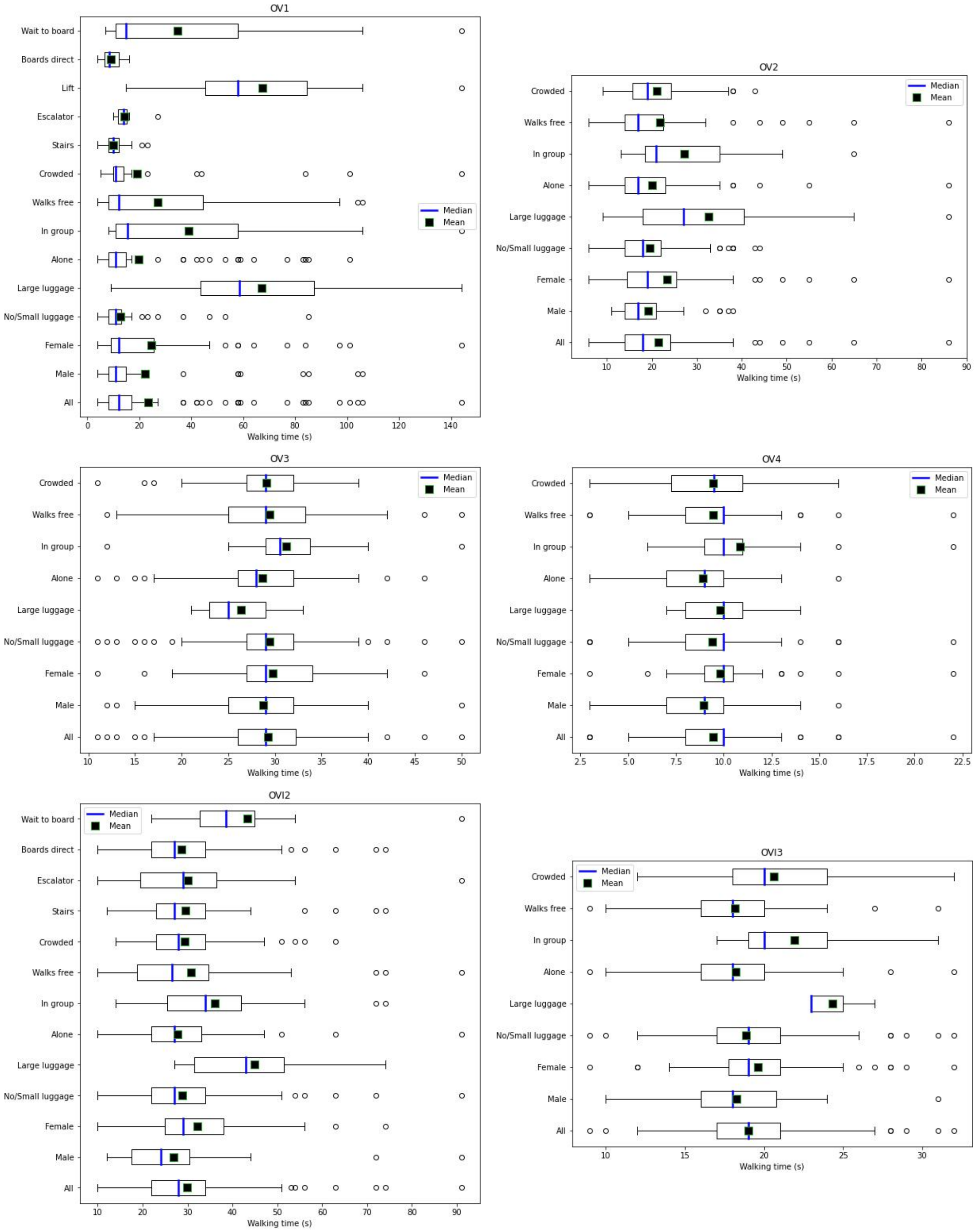


Figure F.2: Boxplots of segments A/B/C Eastbound to D/E.

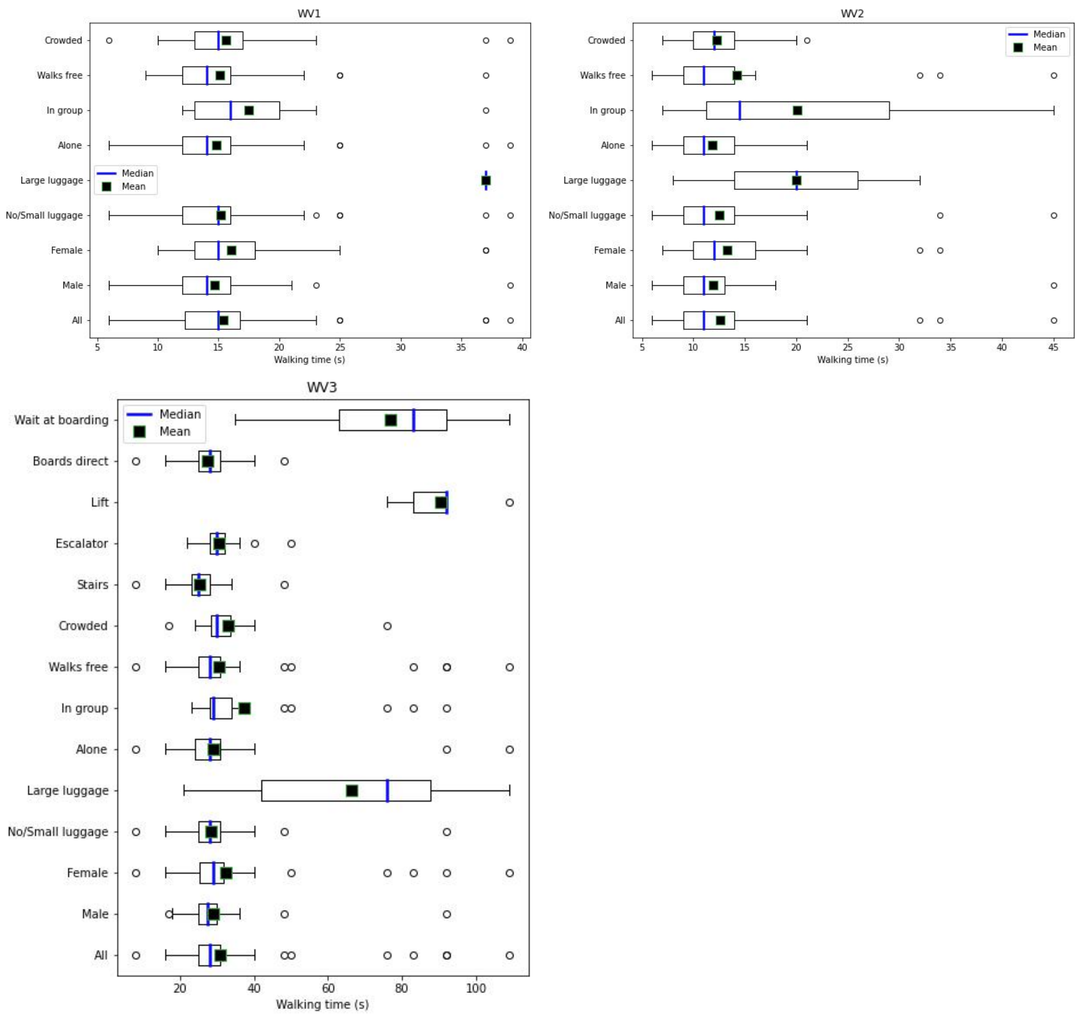


Figure F.3: Boxplots of segments A/B/C Westbound to D/E.

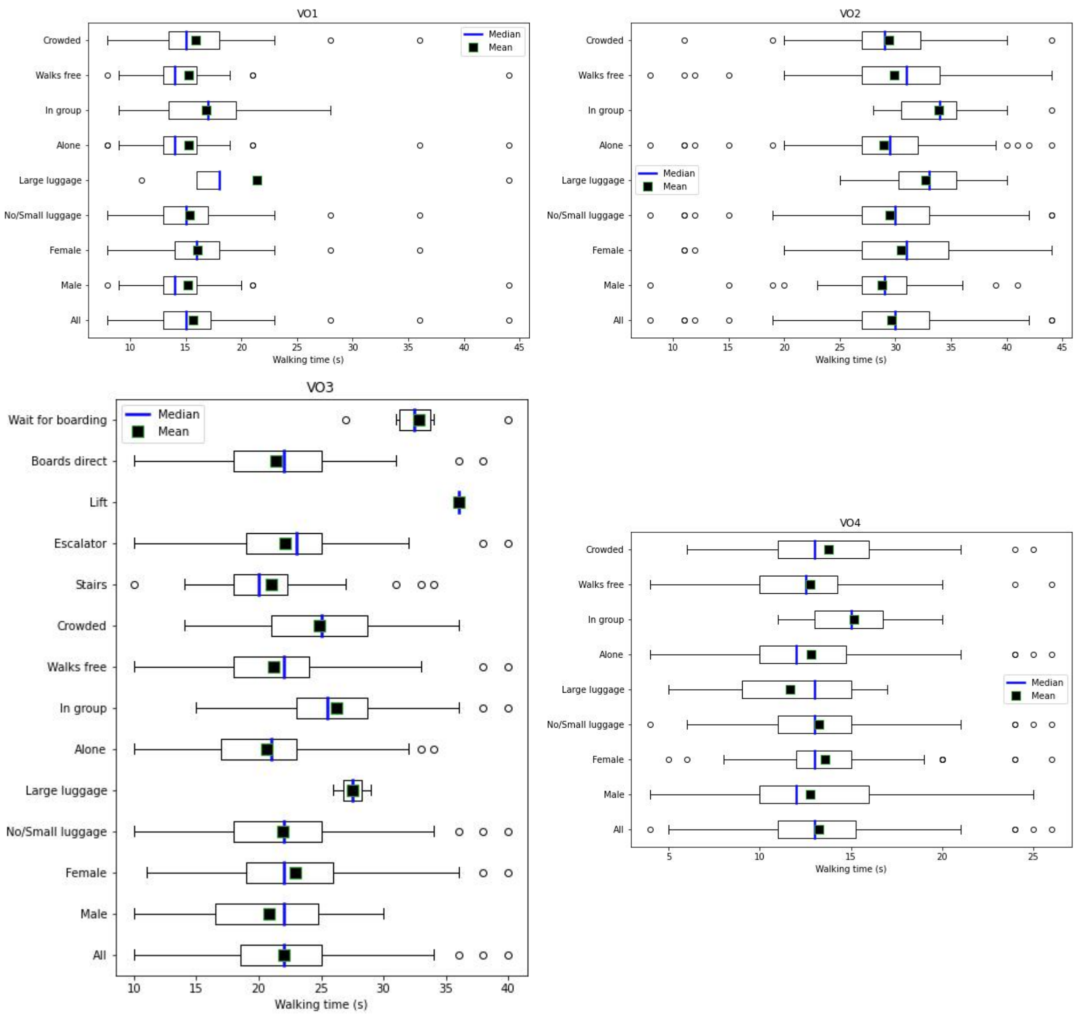


Figure F.4: Boxplots of segments D/E to A/B/C Eastbound.

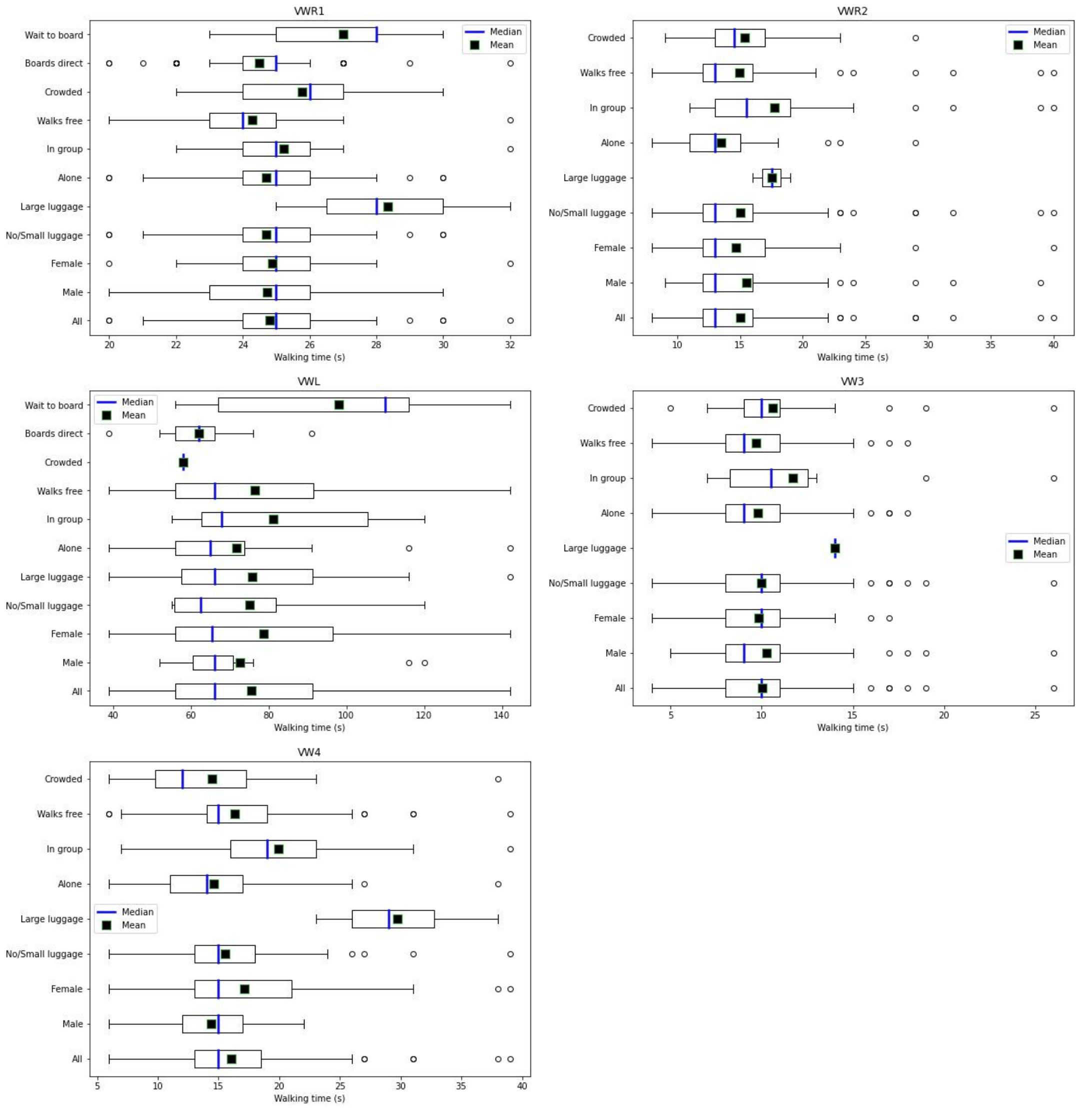


Figure F.5: Boxplots of segments D/E to A/B/C Westbound.

G

Expected distribution through skewness and Kolmogorov-Smirnov tests per segment

The skewness of each walking time distribution helps to predict the type of distribution. The calculation of the skewness is possible in two methods, the first one is through Equation G.1, with only the mean, median and standard deviation of the walking time.

$$\gamma_1 = \frac{3(\bar{x} - Median)}{s} \tag{G.1}$$

In the Python code, the skewness is calculated differently by Scipy (2022). Both calculations' results are in Table G.1. If the skewness lies between -1 and 1 (Field, 2013), then the data can be assumed to be symmetrical and normal distributed. A skewness higher than 1 has a right tail, where the mean is larger than the median value, which is expected for most of the segments because of the values of those variables in Table 5.1.

Table G.1: Skewness and expected distribution of walking time.

Segment	Skewness (γ_1) through Equation G.1	Skewness (Python)	Expected distribution
A/B/C EB by eastern exit	0.80	0.78	Symmetrical
A/B/C EB by western exit	0.65	0.61	Symmetrical
A/B/C WB	0.71	0.51	Symmetrical
D/E SB	0.26	0.42	Symmetrical
D/E NB	0.10	0.21	Symmetrical
OV1	1.24	2.13	Right-tailed
OV2	0.88	2.66	Right-tailed
OVI2	0.40	1.76	Unsure
OVI3	0.01	0.57	Symmetrical
OV3	0.13	-0.14	Symmetrical
OV4	-0.61	0.94	Symmetrical
WV1	0.23	2.42	Unsure
WV2	0.90	3.13	Right-tailed
WV3	0.59	3.42	Unsure
VO1	0.39	2.79	Unsure
VO2	-0.17	-0.78	Symmetrical
VO3	0.00	0.37	Symmetrical
VO4	0.15	0.67	Symmetrical
VWR1	-0.27	0.52	Symmetrical
VWR2	1.13	2.49	Right-tailed
VWL	1.09	1.04	Right-tailed
VW3	0.01	2.17	Unsure
VW4	0.52	1.13	Unsure

Three Kolmogorov-Smirnov tests (KS-test) are performed on the walking time per segment. In Table G.2 the hypothesises are given for each test and the significance level to accept or reject the null hypothesis. The first two KS-test gives insight if the walking time follows a type of general distribution. The last KS-test is a 2-sample test, to see if in peak periods (07:00-09:00 and 16:00-18:00) the walking time is significantly different from off-peak conditions.

Table G.2: Hypothesises for each Kolmogorov-Smirnov test (KS-test).

Normal KS-test	
Null hypothesis (H_0)	Walking time follows a normal distribution.
Alternative hypothesis (H_1)	Walking time is not normally distributed
Lognormal KS-test	
Null hypothesis (H_0)	Walking time follows a lognormal distribution.
Alternative hypothesis (H_1)	Walking time has not a lognormal distribution.
2 samples KS-test	
Null hypothesis (H_0)	No difference in walking time distribution between beak and off-peak.
Alternative hypothesis (H_1)	Difference in walking time distribution between peak and off-peak.
Significance level: 95 % ($\alpha = 0.05$)	

The results from both tests are in Table G.3, some could be categorised in both or none of the given distributions. Based on the conclusion from both tests in Table G.3 and the skewness values in subsection 5.1.3, all segments are assumed to have a log-normal distribution on the walking time.

Table G.3: Kolmogorov-Smirnov tests (KS-test) for normal and lognormal distribution on walking time per segment.

	KS-test for Normality		**KS-test for Log-Normal**		
Segment	**Test statistic D**	**p value**	**Test statistic D**	**p value**	**Conclusion**
A/B/C EB by eastern exit	0.13	0.048	0.10	0.23	Lognormal
A/B/C EB by western exit	0.11	0.14	0.09	0.31	Both
A/B/C WB	0.13	0.06	0.13	0.06	Both
D/E SB	0.09	0.31	0.11	0.10	Both
D/E NB	0.09	0.32	0.15	0.01	Normal
OV1	0.34	0.01	0.21	0.01	None
OV2	0.20	0.01	0.12	0.12	Lognormal
OVI2	0.14	0.03	0.08	0.47	Lognormal
OVI3	0.13	0.07	0.12	0.08	Both
OV3	0.11	0.20	0.16	0.01	Normal
OV4	0.17	0.01	0.20	0.01	None
WV1	0.20	0.01	0.13	0.05	Lognormal
WV2	0.18	0.01	0.12	0.08	Lognormal
WV3	0.32	0.01	0.22	0.01	None
VO1	0.16	0.01	0.12	0.10	Lognormal
VO2	0.13	0.04	0.20	0.01	None
VO3	0.08	0.43	0.10	0.25	Normal
VO4	0.16	0.01	0.11	0.16	Lognormal
VWR1	0.16	0.01	0.14	0.04	None
VWR2	0.19	0.01	0.16	0.01	None
VWL	0.26	0.06	0.22	0.18	Lognormal
VW3	0.18	0.01	0.13	0.04	None
VW4	0.18	0.01	0.14	0.03	None

The two samples KS-test in Table G.4 shows that most segments have the same walking time in off-peak and peak periods. A few segments do have few samples in a specific period, due to a mismatch in the data collection planning. For five segments, there is no conclusion because the data was collected only during a specific period. However, the conclusion is that there is no difference in walking time between busier periods in station Beurs.

Table G.4: Two samples Kolmogorov-Smirnov test between peak and off-peak walking time per segment.

Segment	Test statistic D	p-value	N off-peak	N peak
ABC EB to eastern exit	Only collected off peak			
ABC EB to western exit	0.15	0.79	80	22
ABC West	0.14	0.68	62	42
DE NB	0.07	0.99	71	50
DE SB	0.30	0.19	102	13
OV1	Only collected off peak			
OV2	0.47	0.01	70	46
OVI2	0.27	0.05	47	50
OVI3	0.80	0.03	46	60
OV3	0.21	0.55	78	16
OV4	Only collected off peak			
WV1	0.45	0.01	70	32
WV2	Only collected in peak			
WV3	0.13	0.88	73	24
VO1	0.17	0.31	54	60
VO2	0.27	0.15	20	89
VO3	0.14	0.60	45	58
VO4	0.37	0.01	30	74
VWR1	0.21	0.27	87	27
VWR2	0.20	0.81	92	10
VWL	Too few samples			
VW3	0.22	0.21	28	116
VW4	0.10	0.90	50	60

H

Monte Carlo simulation results

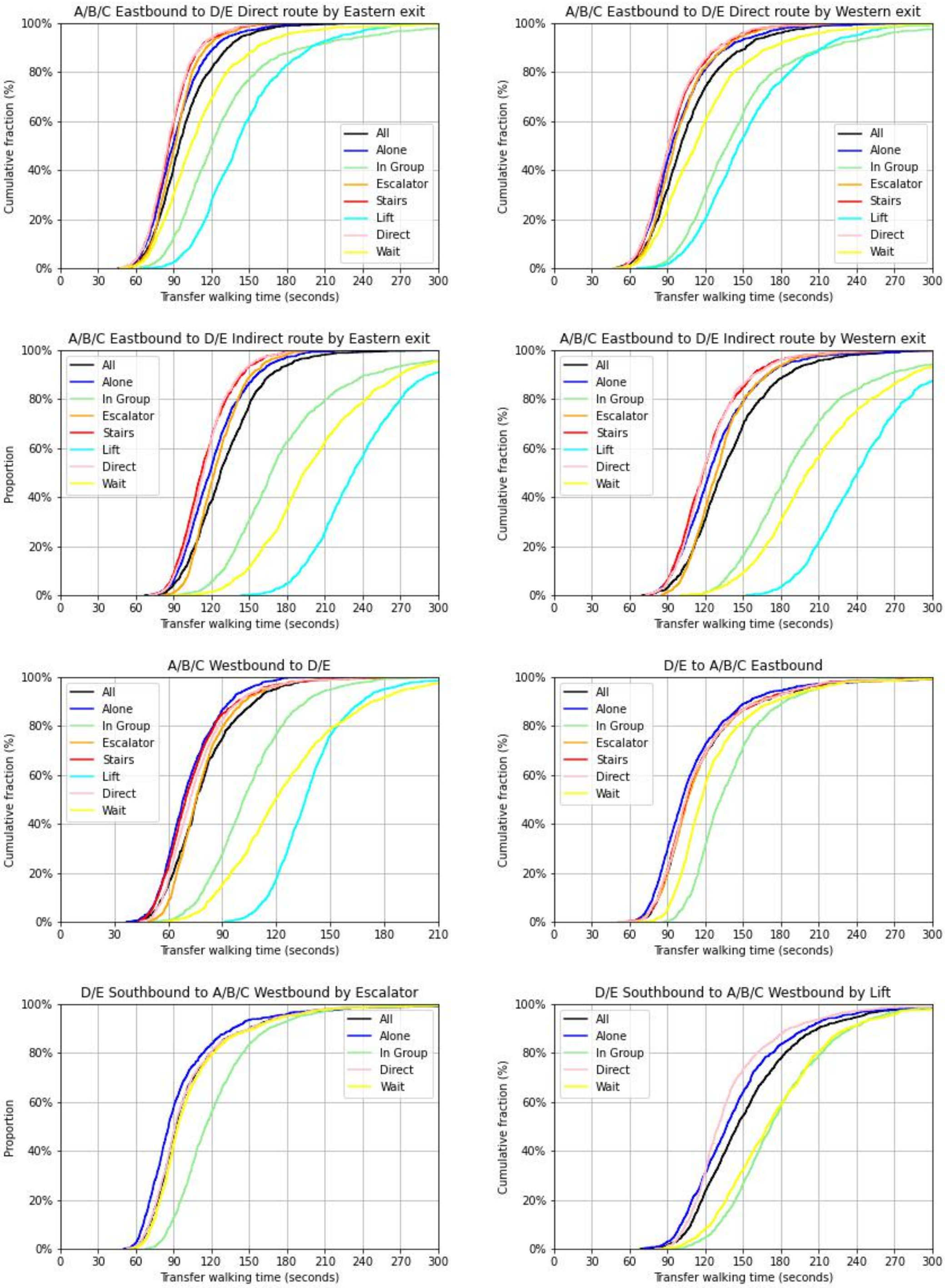


Figure H.1: Simulated transfer walking times for Beurs station.

I

Walking time models development

This Appendix presents the walking time model for each type of segment (corridor, vertical transport and platform). The modelling is done by estimating the β values and iterating each time additional values are included in the model. The model is an ordinal least squares fit through the walking time data. The general walking time estimation model is given in Equation I.1, were the X represents a certain category state. The βs are the linear ordinal least squares estimated parameters from the walking time data.

$$WT = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots \tag{I.1}$$

The estimation of the walking time consists of four steps, which are given in Figure 6.1 and are explained next. Beforehand some segments are excluded to the parameters because these are used in the model cross-validation (Picard and Cook, 1984). The chosen model approach is a multiple linear estimation with a backwards stepwise approach. The first step is to fit the walking time to all categories from the data collection. Not only the significant categories from the data analysis in section 5.2, but also the insignificant ones because these might help to improve the model's fit. The second step is to find additional categories that help explain walking time differences between the segments. Next is determining a possible correlation between variables because a strong correlation overfits the model (Field, 2013). Combined with the correlation results, the stepwise backward approach Tranmer et al., 2020 excludes insignificant variables from the estimation. The final model will only include the significant factors to the walking time.

The last step is model validation. This is done in two ways, with a fit of a passing speed model and the left-out segments in the model fit. The passing speed model is obtained similarly to the walking time model. From there, the passing speeds are translated to walking time and the results are compared to whether they give a similar fit. The validation segments are tested by comparing the estimated walking time from the model to the empirical walking times from that segment.

From the literature, the model in Zhou et al., 2016 used three segment types: corridors, vertical transport and platforms. In this study, a similar approach was used in the data collection, and from the data analysis results in chapter 5, the walking distribution types were different between these three segment types. Therefore, the walking time is modelled for these three segment types separately. Table I.1 shows the segments within each type and the split up between fitting the walking time and validation segment, as explained in the previous paragraphs. The validation segments are mainly identical to one of the training segments, only in the opposite direction.

For each segment type walking time model, the first step is to linear fit the categories from the data collection. Therefore, Table I.2 shows each standard variable to fit and the explanation thereon. Vertical transport and platform segments have additional standard variables, which are given in each specific section. As shown in Table I.2, the length is also part of the first fit because in the model of Zhou et al., 2016 the length was one of the main variables. Furthermore, the earlier study of Du et al., 2009 showed

Table I.1: Segments categorised for segment type and between training and validating the segment type walking time model.

Segment type	Corridors	Vertical transport	Platforms
Train the model with segments:	OV3, OV4, VO1, VWR2, VW3, WV1	OV1, OV2, OVI2, OVI3, VWL, VWR1, WV3	A/B/C EB to eastern exit, A/B/C EB to western exit, A/B/C WB, D/E NB
Validate model with segments:	VO2, VW4, WV2	VO3, VO4	D/E SB

a relation between walking time and speed in metro stations. However, in that study, the considered transfer walking lengths are far longer than the segments in this study, Therefore, a linear increase of walking time related to speed is expected rather than a quadratic increase in (Du et al., 2009 and Zhou et al., 2016) walking time models.

The second part of Table I.2 shows the additional layout element for all models except for the platform walking model. The fare gates showed in chapter 5 a reduction in the passage speed of passengers. This is no surprise because passengers must tap their fare card and wait for the gates to open. The width of a segment plays a role in the capacity of that segment and, thereon the walking time (Zhou et al., 2016). The decision is only to make the difference between wide and narrow corridors based on a threshold value of 3.5 m, the width of a metro platform in Rotterdam. Lastly, the "first segment" variable indicates if a segment is directly after a platform. The data collection reflection in section 4.5 explained that passengers walked slower when approaching the platform and slightly quicker when leaving it.

A non-layout related variable is introduced because of the mostly lognormal distributed walking times according to the data analysis is chapter 5. Therefore, X_O describes whether a sample from the data collection belongs to the quickest or lowest 20% from the observed samples or had a walking time within the 20 to 80 percentile.

Table I.2: Standard fitting variables in each model from the data collection (step 1) and additional layout variables (step 2).

Data collection category (step 1)	Description	State (and value)	Variable name	Variable type	Constant to linear fit
Gender	Defines gender	Male (0) Female (1)	X_G	Dummy	β_G
Luggage	Luggage size	No/Small (0) Large (1)	X_{LU}		β_{LU}
Group size	Defines group size	Alone (0) In Group (1)	X_{GR}		β_{GR}
Crowding	Defines crowding condition	Free (0) Crowded (1)	X_C		β_C
Layout variables					
Length	Horizontal walking distance in segment	-	L	Interval	β_L
Walking distribution related variable (step 2)					
Outlier	Quick or slower passenger	< 20th percentile walking time: (-1) Between 20-80th percentile (0) > 80th percentile walking time: (1)	X_O	Categorical	β_O
Linear constant					
Constant	Intersect walking time constant		None		β_{const}

Modelling corridors

The first model is with data collection variables and the length from Table I.2. The first model fit results are in Table I.3, where gender and luggage are signed as insignificant. Furthermore, the minimal walking time is estimated to be 2.6 seconds, which is assumed to be the lowest observed walking time in all segments. The model fit is already $R^2 = 0.61$, which is a relatively decent fit to lognormal distributed walking times.

$$WT_C = \beta_{const} + \beta_L L + \beta_G X_G + \beta_{LU} X_{LU} + \beta_{GR} X_{GR} + \beta_C X_C \tag{I.2}$$

Table I.3: First model fit with data collection categories and length as variables.

Dep. Variable:	time	**R-squared:**	0.656
Model:	OLS	**Adj. R-squared:**	0.653
Method:	Least Squares	**F-statistic:**	235.0
No. Observations:	622	**Prob (F-statistic):**	3.61e-140
Df Residuals:	616	**Log-Likelihood:**	-1846.1
Df Model:	5	**AIC:**	3704.
Covariance Type:	nonrobust	**BIC:**	3731.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	2.0660	0.495	4.171	0.000	1.093	3.039
length	0.6349	0.019	33.075	0.000	0.597	0.673
groupsize	2.6356	0.445	5.925	0.000	1.762	3.509
gender	0.3339	0.382	0.874	0.382	-0.416	1.084
luggage	1.2923	1.011	1.279	0.201	-0.692	3.277
crowding	0.4871	0.395	1.233	0.218	-0.289	1.263

Omnibus:	256.547	**Durbin-Watson:**	1.784
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1989.058
Skew:	1.637	**Prob(JB):**	0.00
Kurtosis:	11.126	**Cond. No.**	118.

The second model fit includes the layout variables from Table I.4 and the variable outlier. These include the width and presence of fare gates in the segment. Moreover, the effect of being the first or last segment after or before is tested. According to the result in Table I.5, the model fit has largely improved to $R^2 = 0.83$. Moreover, all newly added variables are significant. Especially the variable fast has a large attribute to the model with the absolute second highest Student t value. However, according to the result, adding a constant to the model is not required.

Table I.4: Additional layout variables to fit for the walking vertical transport time models.

Layout variables (Step 2)	Description	State (and value)	Variable name	Variable Type	Constant to linear fit
Width	Width of platform	< 3.5 m (0) ≥ 3.5 m (1)	X_{WI}	Dummy	β_{WI}
Fare gates	Presence of fare gates in segment	None (0) Present (1)	X_{FG}		β_{FG}
Last segment	Segment directly before the platform.	None (0) Before platform (1)	X_{LS}		β_{LS}

Table I.5: Second model fit with additional variables.

Dep. Variable:	time	**R-squared:**	0.851
Model:	OLS	**Adj. R-squared:**	0.849
Method:	Least Squares	**F-statistic:**	388.3
No. Observations:	622	**Prob (F-statistic):**	2.55e-246
Df Residuals:	612	**Log-Likelihood:**	-1586.0
Df Model:	9	**AIC:**	3192.
Covariance Type:	nonrobust	**BIC:**	3236.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	-0.8102	1.212	-0.668	0.504	-3.190	1.570
length	0.7422	0.032	23.015	0.000	0.679	0.806
groupsize	0.5596	0.307	1.822	0.069	-0.043	1.163
luggage	1.9793	0.677	2.922	0.004	0.649	3.310
gender	-0.1920	0.253	-0.759	0.448	-0.689	0.305
crowding	-0.0643	0.271	-0.237	0.813	-0.597	0.468
faregates	3.2820	0.561	5.849	0.000	2.180	4.384
lastsegment	2.3894	0.640	3.734	0.000	1.133	3.646
width	-0.5118	0.423	-1.209	0.227	-1.343	0.320
outlier	6.1513	0.228	26.963	0.000	5.703	6.599

Omnibus:	331.068	**Durbin-Watson:**	1.832
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	4289.368
Skew:	2.057	**Prob(JB):**	0.00
Kurtosis:	15.189	**Cond. No.**	256.

To see a possible correlation between certain variables, Figure I.1 shows the correlation values between all variables from Table I.5. Only a negative correlation is observed between the last segment variable and the length. The reason is that the last corridor segments before the platform are relatively short compared to the others. The conclusion is that all variables are independent because none show a strong correlation. Therefore, removing each insignificant variable should have a marginal effect on the model fit.

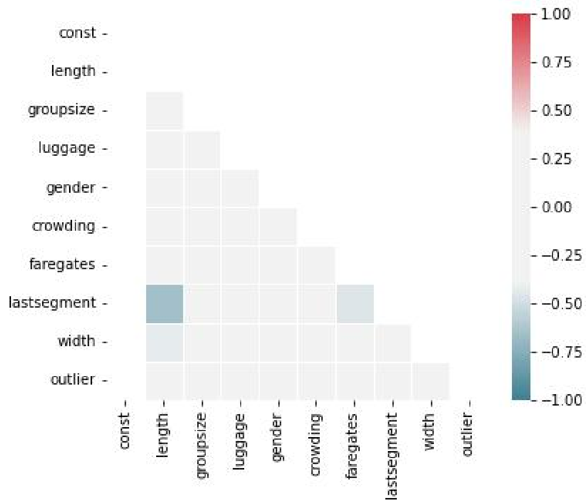


Figure I.1: Correlation between all variables in the corridor walking time model.

The backwards elimination of insignificant variables yields the final linear fit in Table I.6. As expected, the fit and the variable parameters remained the same. Most parameters add or subtract around 1 or 2 seconds. The bonus of 1.3 seconds for wider corridors seems counter-intuitive because one would expect that in narrower corridors, the walking time should be longer because of a larger probability of crowding. The fare gates value seems to be representative of the lost waiting time to pass the fare gates. However, the actual passing time of a fare gate was out of scope in the data collection.

Table I.6: Third iteration corridor segments walking time model.

Dep. Variable:	time	**R-squared:**	0.850
Model:	OLS	**Adj. R-squared:**	0.849
Method:	Least Squares	**F-statistic:**	582.9
No. Observations:	622	**Prob (F-statistic):**	6.06e-250
Df Residuals:	615	**Log-Likelihood:**	-1587.1
Df Model:	6	**AIC:**	3188.
Covariance Type:	nonrobust	**BIC:**	3219.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	-2.1194	0.735	-2.882	0.004	-3.564	-0.675
length	0.7710	0.022	34.672	0.000	0.727	0.815
groupsize	0.6096	0.303	2.011	0.045	0.014	1.205
luggage	1.9808	0.677	2.927	0.004	0.652	3.310
faregates	3.6733	0.466	7.881	0.000	2.758	4.589
lastsegment	2.7124	0.585	4.636	0.000	1.563	3.861
outlier	6.1188	0.226	27.027	0.000	5.674	6.563

Omnibus:	334.242	**Durbin-Watson:**	1.833
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	4409.357
Skew:	2.077	**Prob(JB):**	0.00
Kurtosis:	15.364	**Cond. No.**	176.

The final walking time model for corridor segments is given in Equation I.3 based on the coefficients from Table I.6. The most significant factor contributing to the walking time is whether someone belongs to the quicker walkers or not. Followed by the fare gate presence, while attributes from a passenger only marginal influence the walking time in station corridors.

$$WT_C(s) = -2.119 + 0.771L + 0.61X_{GR} + 1.981X_{LU} + 3.673X_{FG} + 2.712X_{LS} + 6.119X_O \qquad \text{(I.3)}$$

Passing speed model of corridors

The first fit is with the significant attributes from Table I.6. The resulting fit is in Table I.7 and all attributes are significant. Therefore, the resulting passing speed model for corridors is given in Equation I.4. The most significant factor on the passenger side is whether the objective is an upper bound of the passing speed. From a layout perspective, the presence of fare gates has the largest reduction in the passing speed. At the passenger attributes, large luggage affects the passing speed the most.

$$ps_C(m/s) = 1.42 - 0.102X_{GR} + 0.1805X_{LU} - 0.286X_{FG} + 0.158X_{LS} + 0.693X_O \quad \text{(I.4)}$$

Table I.7: Passing speed model fit with variables from Table I.6.

Dep. Variable:	v	**R-squared:**	0.517
Model:	OLS	**Adj. R-squared:**	0.513
Method:	Least Squares	**F-statistic:**	131.9
No. Observations:	622	**Prob (F-statistic):**	6.79e-95
Df Residuals:	616	**Log-Likelihood:**	-108.69
Df Model:	5	**AIC:**	229.4
Covariance Type:	nonrobust	**BIC:**	256.0

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	1.4197	0.023	63.068	0.000	1.376	1.464
groupsize	-0.1015	0.028	-3.679	0.000	-0.156	-0.047
luggage	-0.1805	0.063	-2.875	0.004	-0.304	-0.057
faregates	-0.2859	0.026	-10.928	0.000	-0.337	-0.235
lastsegment	-0.1579	0.036	-4.447	0.000	-0.228	-0.088
outlier	0.6928	0.032	21.758	0.000	0.630	0.755

Omnibus:	320.750	**Durbin-Watson:**	1.998
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3979.159
Skew:	1.987	**Prob(JB):**	0.00
Kurtosis:	14.737	**Cond. No.**	6.50

Modelling VTP segments

Besides the standard variables from the data collection Table I.2, two additional variables were collected, as seen in Table I.8. These are related to the vertical transport specific, the vertical transport mode and the waiting conditions to board it. The vertical transport mode is only a stair mode or the lift because of the insignificant walking time between stairs and escalators. The length variable is split up into the corridor length (L_h) and the actual (horizontal) vertical transport length (L_v).

Table I.8: Additional layout variables to fit for the walking vertical transport time models.

Data collection category (Step 1)	Description	State	Variable Value	Variable Name	Type Variable	Constant to linear fit
VT	Vertical transport mode	Stair mode (escalator or stairs) Lift	0 1	X_{VT}	Dummy	β_{VT}
Wait	Waiting condition to board VT.	Direct Wait	0 1	X_{WA}	Dummy	β_{WA}
v_length	Horizontal length of vertical transport.	-	-	L_v	Interval	β_{L_v}
h_length	Horizontal length before/after vertical transport	-	-	L_h	Interval	β_{L_h}
Layout variables (Step 2)						
Width	Width of platform	< 3.5 m ≥ 3.5 m	0 1	X_{WI}	Dummy	β_{WI}
Fare gates	Presence of fare gates in segment	None Present	0 1	X_{FG}	Dummy	β_{FG}
Last segment	Segment directly before the platform.	None Before platform	0 1	X_{FG}	Dummy	β_{FS}
Complete	Segment contains complete vertical transport length	Partly Complete	0 1	X_{CO}	Dummy	β_{CO}
Stairs only	Segment contains only stair modes	No Yes	0 1	X_{SS}	Dummy	β_{SS}
Part1	Location horizontal part	Before After Unclear	-1 1 0	X_P	Categorical	β_P
Slope	Direction of vertical movement.	Up Down Unclear	1 -1 0	X_S	Categorical	β_S
Bottleneck	Segment develops queues	No Yes	0 1	X_B	Dummy	β_B

The estimation in Table I.9 already gives a really good fit ($R = 0.70$), and all variables are relevant. This is already quite a good fit with only the data collection variables. Furthermore, only gender and crowding are insignificant for now. For the vt mode constant, there is already some waiting time included because the actual travel time of each lift was only 20-25 seconds. Besides, the waiting penalty of 7 seconds for each mode.

Table I.9: First iteration vertical transport walking time with data collection categories and length.

Dep. Variable:	time	**R-squared:**	0.705
Model:	OLS	**Adj. R-squared:**	0.702
Method:	Least Squares	**F-statistic:**	184.4
No. Observations:	625	**Prob (F-statistic):**	5.85e-158
Df Residuals:	616	**Log-Likelihood:**	-2338.2
Df Model:	8	**AIC:**	4694.
Covariance Type:	nonrobust	**BIC:**	4734.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.3644	1.483	0.246	0.806	-2.548	3.277
gender	1.2076	0.836	1.444	0.149	-0.434	2.850
luggage	4.1375	0.878	4.712	0.000	2.413	5.862
groupsize	5.1175	0.998	5.125	0.000	3.157	7.078
crowding	0.2804	0.902	0.311	0.756	-1.491	2.052
vt	45.5744	2.206	20.656	0.000	41.242	49.907
wait	7.2438	1.419	5.103	0.000	4.456	10.031
luggage	4.1375	0.878	4.712	0.000	2.413	5.862
v_length	1.4015	0.176	7.977	0.000	1.056	1.746
h_length	0.7838	0.094	8.334	0.000	0.599	0.969

Omnibus:	290.701	**Durbin-Watson:**	1.574
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3360.028
Skew:	1.756	**Prob(JB):**	0.00
Kurtosis:	13.802	**Cond. No.**	2.09e+17

In the second includes the additional layout variables from Table I.8 to explain all different layouts of the segments. The resulting fit in Table I.10 has improved. However, around half of the new layout variables are insignificant. Still, lift usage remains the most significant variable (X_{VT}). Whether the segment is a bottleneck and might cause longer waiting times to board vertical transport is insignificant. Therefore, indicating that the model works best for segments without capacity restriction.

Table I.10: Second iteration walking time model for vertical transport segments.

Dep. Variable:	time	**R-squared:**	0.797
Model:	OLS	**Adj. R-squared:**	0.792
Method:	Least Squares	**F-statistic:**	184.0
No. Observations:	625	**Prob (F-statistic):**	2.76e-201
Df Residuals:	611	**Log-Likelihood:**	-2222.5
Df Model:	13	**AIC:**	4473.
Covariance Type:	nonrobust	**BIC:**	4535.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	2.7050	0.563	4.803	0.000	1.599	3.811
gender	0.2247	0.701	0.321	0.749	-1.152	1.601
luggage	3.7146	1.499	2.478	0.013	0.770	6.659
groupsize	2.2954	0.851	2.697	0.007	0.624	3.967
crowding	-1.9725	0.805	-2.451	0.015	-3.553	-0.392
vt	39.7877	2.295	17.338	0.000	35.281	44.295
wait	5.3147	1.337	3.975	0.000	2.689	7.940
v_length	2.3182	0.225	10.315	0.000	1.877	2.760
h_length	0.6640	0.106	6.267	0.000	0.456	0.872
faregates	2.4053	0.410	5.859	0.000	1.599	3.211
width	-2.0034	1.004	-1.996	0.046	-3.975	-0.032
stairsonly	-5.5342	1.834	-3.017	0.003	-9.136	-1.932
lastsegment	-6.1910	1.841	-3.364	0.001	-9.806	-2.576
complete	0.7842	0.906	0.866	0.387	-0.994	2.563
part1	4.5791	1.108	4.131	0.000	2.402	6.756
slope	-0.6411	0.480	-1.337	0.182	-1.583	0.301
bottleneck	1.3909	0.666	2.087	0.037	0.082	2.700
outlier	10.6031	0.678	15.630	0.000	9.271	11.935

Omnibus:	378.123	**Durbin-Watson:**	1.763
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	7733.080
Skew:	2.279	**Prob(JB):**	0.00
Kurtosis:	19.619	**Cond. No.**	9.33e+16

With these many variables, the probability of correlation is high. This assumption is true according to Figure I.2. The variable fare gates has strong relations with the bottleneck, slope, part1 and width variables. Based on the layout of segments with fare gates, most have indeed similarities. However, the decision is to remove the variable in the backward elimination. Other variables which are considered to remove are width and slope. However, these are already insignificant in the previous model fit. The relation between vt and luggage can be explained by the fact that most passengers with large luggage take the elevator.

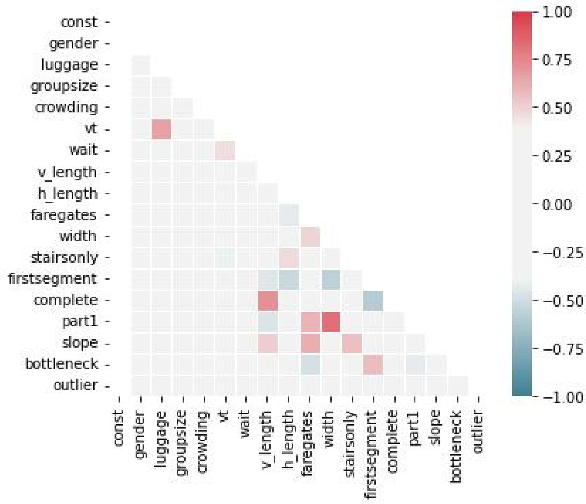


Figure I.2: Correlation between all variables in the vertical transport walking time model.

The backward elimination order is: bottleneck, gender, width, slope, fare gates and last segment. After that, most correlation was gone, and the remaining layout variables were mostly insignificant. Only the variable describing the location of the corridor part was left. The final fit with only significant variables is given in Table I.11.

Table I.11: Third iteration.

Dep. Variable:	time	**R-squared:**	0.792
Model:	OLS	**Adj. R-squared:**	0.790
Method:	Least Squares	**F-statistic:**	293.6
No. Observations:	625	**Prob (F-statistic):**	1.65e-204
Df Residuals:	616	**Log-Likelihood:**	-2229.1
Df Model:	8	**AIC:**	4476.
Covariance Type:	nonrobust	**BIC:**	4516.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	3.2515	1.203	2.703	0.007	0.889	5.614
luggage	4.5633	1.476	3.091	0.002	1.664	7.463
groupsize	2.4592	0.855	2.876	0.004	0.780	4.138
vt	43.7111	1.878	23.278	0.000	40.024	47.399
wait	4.6326	1.288	3.596	0.000	2.103	7.163
v_length	1.8151	0.178	10.200	0.000	1.466	2.165
h_length	0.4811	0.089	5.386	0.000	0.306	0.657
outlier	9.9441	0.652	15.256	0.000	8.664	11.224
part1	2.4093	0.572	4.215	0.000	1.287	3.532

Omnibus:	368.682	**Durbin-Watson:**	1.768
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	7392.450
Skew:	2.204	**Prob(JB):**	0.00
Kurtosis:	19.261	**Cond. No.**	93.5

The walking time model for the vertical transport segment is given Equation I.5. The most significant factor in the model is using the lift. Regarding passenger characteristics, having large luggage is the most influential variable, followed by walking in a group. Lastly, the vertical transport segment length adds almost a second for every additional meter in length.

$$WT_V(s) = 3.252+0.481L_h+1.815L_v+2.459X_{GR}+43.711X_{VT}+4.633X_{WA}+4.563X_{LU}+2.409X_P+9.944X_O \quad \text{(I.5)}$$

Passing speed for vertical transport segments

The first fit is with the variables from the previous walking time model fit in Table I.11. The fit as shown in Table I.12 shows that luggage and first segment variables are insignificant. Therefore, these are removed in the second fit and the remaining variables are indeed all significant in Table I.13. According to the adjusted $R^2 = 0.55$, the fit is adequate. The passing speed is then shown in Equation I.6. The most influential factor is taking the lift.

$$ps_V(m/s) = 0.86 - 0.081X_{GR} - 0.567X_{VT} - 0.162X_{WA} + 0.537X_O \quad \text{(I.6)}$$

Table I.12: Passing speed model fit with variables from Table I.11.

Dep. Variable:	v	**R-squared:**	0.551
Model:	OLS	**Adj. R-squared:**	0.547
Method:	Least Squares	**F-statistic:**	126.6
No. Observations:	625	**Prob (F-statistic):**	4.04e-104
Df Residuals:	618	**Log-Likelihood:**	-70.582
Df Model:	6	**AIC:**	155.2
Covariance Type:	nonrobust	**BIC:**	186.2

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.8579	0.015	57.132	0.000	0.828	0.887
luggage	-0.0577	0.045	-1.269	0.205	-0.147	0.032
groupsize	-0.0784	0.027	-2.936	0.003	-0.131	-0.026
vt	-0.5396	0.057	-9.548	0.000	-0.651	-0.429
wait	-0.1330	0.038	-3.521	0.000	-0.207	-0.059
outlier	0.5376	0.030	17.730	0.000	0.478	0.597
part1	0.0283	0.014	1.962	0.050	-2.9e-05	0.057

Omnibus:	239.143	**Durbin-Watson:**	1.249
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1951.403
Skew:	1.475	**Prob(JB):**	0.00
Kurtosis:	11.138	**Cond. No.**	6.51

Table I.13: Passing speed model fit with variables from Table I.11.

Dep. Variable:	v	**R-squared:**	0.548
Model:	OLS	**Adj. R-squared:**	0.545
Method:	Least Squares	**F-statistic:**	187.7
No. Observations:	625	**Prob (F-statistic):**	2.53e-105
Df Residuals:	620	**Log-Likelihood:**	-73.090
Df Model:	4	**AIC:**	156.2
Covariance Type:	nonrobust	**BIC:**	178.4

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.8599	0.015	58.278	0.000	0.831	0.889
groupsize	-0.0813	0.027	-3.055	0.002	-0.134	-0.029
vt	-0.5674	0.045	-12.687	0.000	-0.655	-0.480
wait	-0.1622	0.035	-4.664	0.000	-0.230	-0.094
outlier	0.5371	0.030	17.690	0.000	0.478	0.597

Omnibus:	248.391	**Durbin-Watson:**	1.270
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	2031.694
Skew:	1.546	**Prob(JB):**	0.00
Kurtosis:	11.274	**Cond. No.**	4.78

Modelling platform segments

Besides the standard categories from Table I.2, the alighting location is also part of the platform segment. However, a sample's covered length is unknown because the exact alighting location was not observed. In the category Alight X_A, the variable is transformed from the relative position in the metro (front, middle, back) to the relative position to the platform exit (closest 1/3 to the exit, middle and furthest to the exit). The position is based on the distance between the middle of a 30 m metro train and the exit location. Using fixed length categories to determine the position (average distance below or above threshold length) gave a poorer fit to the data. The base value is the middle section relative to the exit, the closest section has a negative dummy value, and the furthest is a positive value, as is set in Table I.14.

Table I.14: Additional layout variables to fit for the walking time platform models.

Data collection category (Step 1)	**Description**	**State (and value)**	**Variable name**	**Variable type**	**Constant to linear fit**
Alight	Relative average distance to (transfer) exit from a metro part.	Closest $\frac{1}{3}$ (-1) Middle $\frac{1}{3}$ (0) Furthest $\frac{1}{3}$ (1)	X_A	Categorical	β_A
Layout variables (Step 2)					
Exit	Number of exits at platform (for transfer)	1 Exit (0) 2 Exits (1)	X_E	Dummy	β_E
Width	Width of platform	< 3.5 m (0) ≥ 3.5 m (1)	X_{WI}	Dummy	β_{WI}

One platform (A/B/C Eastbound) in Beurs has two exits to the transfer facilities, so X_E describes the number of exits. The A/B/C Westbound theoretically also has two exit points for the same exit, but those are very close to each other and are modelled as a platform with one exit.

In the first estimation with only the data collection categories gives the following results in Table I.15. This model does not require a constant. Besides the insignificance of gender and crowding, the variable alight is also insignificant. An explanation for the insignificance is assumed that the length to the exit is only important. Various dummy configurations values have been tried, but the alighting remains insignificant. The fit is already quite good with and adjusted R of 0.64. The adjusted R^2 is now a better indicator of the model fit than the previous segment types because all platforms had a similar range of observed walking times.

Table I.15: First iteration walking time model platform segments.

Dep. Variable:	time	**R-squared:**	0.645
Model:	OLS	**Adj. R-squared:**	0.640
Method:	Least Squares	**F-statistic:**	126.3
No. Observations:	424	**Prob (F-statistic):**	1.43e-90
Df Residuals:	417	**Log-Likelihood:**	-1592.1
Df Model:	6	**AIC:**	3198.
Covariance Type:	nonrobust	**BIC:**	3227.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	-0.1254	2.314	-0.054	0.957	-4.673	4.422
gender	0.4785	1.019	0.469	0.639	-1.525	2.482
luggage	4.4706	1.843	2.426	0.016	0.849	8.092
groupsize	5.8067	1.348	4.307	0.000	3.156	8.457
crowding	0.9007	1.146	0.786	0.432	-1.351	3.153
alight	-2.9390	1.514	-1.941	0.053	-5.915	0.037
length	0.8104	0.061	13.188	0.000	0.690	0.931

Omnibus:	52.616	**Durbin-Watson:**	1.372
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	83.805
Skew:	0.786	**Prob(JB):**	6.34e-19
Kurtosis:	4.508	**Cond. No.**	175.

The second iteration shows with the additional variables exit, width and outlier show a major improvement to the fit in Table I.16. The adjusted R^2 is now 0.84, which is a very good fit. The variable luggage is significant at the platform as well. Now, a constant is significant, but the luggage variable became insignificant. From the new variables, only the outlier seems to be important.

Table I.16: Second iteration platform segment walking time model.

Dep. Variable:	time	**R-squared:**	0.839
Model:	OLS	**Adj. R-squared:**	0.836
Method:	Least Squares	**F-statistic:**	240.2
No. Observations:	424	**Prob (F-statistic):**	2.82e-158
Df Residuals:	414	**Log-Likelihood:**	-1424.2
Df Model:	9	**AIC:**	2868.
Covariance Type:	nonrobust	**BIC:**	2909.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	4.0238	1.851	2.174	0.030	0.385	7.663
gender	-0.2241	0.689	-0.325	0.745	-1.579	1.131
luggage	1.7769	1.256	1.414	0.158	-0.693	4.246
groupsize	2.5629	0.928	2.762	0.006	0.739	4.387
crowding	-0.6018	0.791	-0.760	0.448	-2.158	0.954
alight	-7.1341	1.445	-4.937	0.000	-9.975	-4.294
length	0.6975	0.064	10.971	0.000	0.573	0.823
width	1.5209	1.471	1.034	0.302	-1.370	4.412
exits	-0.5373	1.000	-0.537	0.591	-2.503	1.428
outlier	15.3408	0.695	22.065	0.000	13.974	16.707

Omnibus:	17.793	**Durbin-Watson:**	1.438
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	20.428
Skew:	0.428	**Prob(JB):**	3.67e-05
Kurtosis:	3.652	**Cond. No.**	235.

Four strong relations are shown in the correlation matrix between variables of Figure I.3. The strongest is between alighting location and length, which is obvious because the further a passenger alighted the metro the longer he has to walk on the platform. Furthermore, a similar correlation can be explained between outlier, alight and length variables. The first passenger to leave the platform was the closest to the exit, had the shortest length to walk and thus have the quickest walking time. Based on the different platform layouts, all three variables will remain in the model besides the strong correlation.

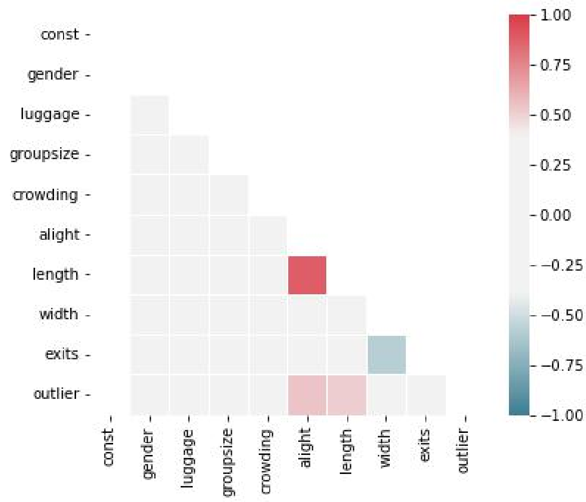


Figure I.3: Correlation between all variables in the platform walking time model.

In the third step, the backward elimination order was X_G, X_E, X_C. In each fit, the remainder of the insignificant variables were still insignificant. In the fit after that, luggage became insignificant as well. The final fit with only significant ones is in Table I.17. Again, a constant for the model is insignificant. The length constant is identical to the one for the corridors.

Table I.17: Third iteration platform segment walking time model

Dep. Variable:	time	**R-squared:**	0.837
Model:	OLS	**Adj. R-squared:**	0.835
Method:	Least Squares	**F-statistic:**	537.6
No. Observations:	424	**Prob (F-statistic):**	1.73e-163
Df Residuals:	419	**Log-Likelihood:**	-1427.2
Df Model:	4	**AIC:**	2864.
Covariance Type:	nonrobust	**BIC:**	2885.

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	1.7893	1.332	1.343	0.180	-0.829	4.408
groupsize	2.5804	0.922	2.800	0.005	0.769	4.392
alight	-8.2165	1.035	-7.941	0.000	-10.250	-6.183
length	0.7629	0.041	18.620	0.000	0.682	0.843
outlier	15.0751	0.669	22.544	0.000	13.761	16.389

Omnibus:	16.692	**Durbin-Watson:**	1.429
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	18.983
Skew:	0.412	**Prob(JB):**	7.55e-05
Kurtosis:	3.629	**Cond. No.**	156.

The walking time model for the platform thereon consists of just three variables from Table I.17. As Equation I.7 shows, the most significant factor is whether the passenger is an outlier, followed by the alighting location.

$$WT_P(s) = 0.763L + 2.580X_{GR} - 8.217X_A + 15.075X_O \quad \text{(I.7)}$$

Passing speed model of platforms

The first fit is with (the significant variables) from Table I.17. The resulting fit is in Table I.18, the adjusted R^2 is now only 0.55 and all variables remained significant. Therefore, the passing speed model with the constants from Table I.18 is given in Equation I.8. Of course, the outlier for faster passengers is the most significant factor in this model.

$$ps_P(m/s) = 1.107 - 0.177X_{GR} + 0.177X_A + 1.7791X_O \tag{I.8}$$

Table I.18: Passing speed model fit with variables from Table I.17.

Dep. Variable:	v	**R-squared:**	0.552
Model:	OLS	**Adj. R-squared:**	0.549
Method:	Least Squares	**F-statistic:**	172.7
Date:	Fri, 17 Mar 2023	**Prob (F-statistic):**	6.15e-73
Time:	11:00:57	**Log-Likelihood:**	-421.45
No. Observations:	424	**AIC:**	850.9
Df Residuals:	420	**BIC:**	867.1
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	1.1067	0.042	26.198	0.000	1.024	1.190
groupsize	-0.1766	0.086	-2.050	0.041	-0.346	-0.007
uitstap	0.1771	0.044	4.044	0.000	0.091	0.263
outlier	1.7791	0.081	21.867	0.000	1.619	1.939

Omnibus:	221.881	**Durbin-Watson:**	1.835
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1818.285
Skew:	2.097	**Prob(JB):**	0.00
Kurtosis:	12.238	**Cond. No.**	3.23