Technische Universität München
TUM School of Engineering and Design

Advances in Pedestrian Travel Demand Modeling

Innovation of Data, Modeling Approaches and Outcomes

Qin Zhang

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Constantinos Antoniou

Prüfende der Dissertation:

1. Prof. Dr.-Ing. Rolf Moeckel
2. Prof. Dr. Kelly J. Clifton

Die Dissertation wurde am 15.11.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 01.03.2023 angenommen.

Acknowledgement

This dissertation marks the end of a long and memorable journey that I could not have completed without the dedicated support of my supervisors, peer colleges, family, and friends.

First, I would like to express my sincere gratitude to my supervisors Prof. Rolf Moeckel and Prof. Kelly Clifton for their consistent and helpful support throughout my doctoral studies. They always provided me with great assistance and guidance, constructive feedback to help me better my work as well as excellent chances for me to participate in international conferences and academic communications. I consider myself lucky to have two excellent supervisors and benefited from their knowledge and character.

Furthermore, I would like to thank my mentor Ana for her constant support, advice, and encouragement to pursue my doctoral degree. Similarly, I would like to thank my teammates Alona, Carlos, Cat, Corin, Joanna, Karin, Nico, Matthew, and Wei-Chieh for their fruitful and enjoyable collaboration.

Finally, I wish to thank for the constant support and encouragement from my family and friends. Special thanks go to my husband Chong. He supported me through all the hard times. His support allowed me to focus on my studies and finish my dissertation. I thank my dear mother and grandmother for their strong and unwavering love, support, and patience. Last but not least, I dedicate this dissertation to the memory of my father, who was the first to encourage me to pursue my doctorate.

Abstract

Walking is an important mode for sustainability, health, environment, urban vitality, and equity. There is a critical need for pedestrian planning tools that operate at a scale that is sensitive to pedestrian needs and the quality of walking environments. Some efforts have been taken to develop stand-alone pedestrian planning tools or consider pedestrians in travel demand models that operate at an urban scale. However, pedestrian travel demand models still face challenges, such as limited understanding of travel behavior variability due to the lack of longitudinal data, poor spatial and temporal transferability of built environment measurements, and insufficient work in applying pedestrian travel behavior knowledge into modeling practices. Those gaps could limit the ability of researchers, planners, and policymakers to assess future transport demand and evaluate various future plans and policies.

This dissertation improves pedestrian travel demand modeling in terms of model accuracy, sensitivity, transferability, and efficiency. It targets to advance pedestrian modeling both theoretically and practically.

To begin, this work improves the stand-alone pedestrian planning tool Model of Pedestrian Demand (MoPeD) established in the Portland context. It explores the appropriate spatial resolution for measuring pedestrian environment and modeling behavior at the scale of a metropolitan area. It concludes that there is no optimal spatial resolution, and the selection of spatial resolution depends on the type of applications and the availability of spatial data needed to implement. More importantly, efforts are made to enhance the model performance of MoPeD, particularly for model transferability and efficiency. The model run time is upgraded to a few minutes for running large-scale areas at fine spatial units. The new pedestrian accessibility measurements are easy to construct and transfer to other contexts. Two-step walk destination choice models are estimated using small-scale destination zone alternatives. Besides, multiple built environment variables such as network density, accessibility to shops and retail, cross motorway, slope, and proportion of industrial jobs are incorporated in walk destination choice models.

Following that, the integrated modeling framework that incorporates MoPeD into the agent-based transport model (MITO) is developed in the Munich context. MoPeD and MITO have their respective strengths, and the integration can complement each other. The integrated model benefits from MoPeD's fine spatial resolution and its better

representation of the built environment, as well as MITO's agent-based environment and its good representation of motorized modes. The technical integration of the two models is straightforward though involving a large amount of coding work. The spatial transfer of econometric models from the estimation context (Portland) to the application context (Munich) involves several steps. First, the input data containers are prepared for the Munich context, including fine zone system, pedestrian network, and built environment data. Then, models are calibrated at an aggregate level by updating constants. After applying the integrated model to the Munich study area, model performances are compared to the MITO stand-alone model. It concludes that the integrated model can provide more accurate travel outcomes such as shares of walk trips, walk trip length distribution, the spatial distribution of walk trips, and physical activity volumes.

Finally, novel and longitudinal data Google Location History (GLH) is collected. 27 valid GLH datasets are used in this research. Due to privacy concerns and the effort required for in-person recruitment, the number of recruited individuals is limited. However, the data is rich in a wide range of time period. Most of the GLH datasets cover a large number of successive days over a two-year period. The mean number of days recorded is 481 days. The collected GLH data is employed to have a closer investigation of travel behavior variability. The analysis of travel behavior variability proves that individuals have a great deal of day-to-day variability. Week-to-week travel behaviors have relatively low dispersion, while people tend to have periodical behavior on a monthly scale. The analysis of GLH data confirms that household travel surveys were poor at capturing walking activities and biased in self-reported travel distance. The work also attempts to find out the potential determinants of travel behavior variability. It is proved that socio-demographics have impacts on travel behavior variability, particularly students have higher intrapersonal variability than workers, and individuals who have no car are more variable than those who have cars. Weather and public holidays can also disrupt an individual's travel behavior routine.

Overall, this dissertation makes strides towards a more accurate, sensitive, transferable, and efficient pedestrian planning tool for delivering travel outcomes as well as evaluating policies and scenarios. The pedestrian planning tool is an open-source model which can be further applied to other contexts. Furthermore, this work illustrates the use of longitudinal data. Although the analyses are conducted with a small sample size, the findings still reveal some innovative and exploratory insights into travel behavior and perhaps more importantly provide some better guidance on the design of future data collection efforts and the utility of GLH for transportation analysis.

Table of Contents

List of Figures

List of Tables

Acronyms

Coefficient of variation	CV
Excess commute	EC
German Health Update	GEDA
Google Location History	GLH
Global Positioning System	GPS
Home-based education	HBE
Home-based other	HBO
Home-based recreational	HBR
Home-based shopping	HBS
Home-based work	HBW
Household Travel Survey	HTS
JavaScript Object Notation	JSON
Multi-Agent Transport Simulation	MATSim
Mobilität in Deutschland	MiD
Microscopic Transportation Orchestrator	MITO
Marginal metabolic equivalent	mMET
Mobilitätspanel	MOP
Model of Pedestrian Demand	MoPeD
Metropolitan Planning Organization	MPO
Non-home-based other	NHBO
Non-home-based work	NHBW
Oregon Household Activity Survey	OHAS
Open Street Map	OSM
Physical activity	PA
Pedestrian Analysis Zone	PAZ
Pedestrian catchment area	PCA
Pedestrian catchment ratio	PCR

Coefficient of variation	CV
Pedestrian Index Environment	PIE
Root mean square error	RMSE
Transportation Analysis Zone	TAZ
Travel time budgets	TTB

1. Introduction

1.1. Background

Walking, as a low/no-tech mobility, is the most traditional mode we have experienced in everyday travel. More use of non-motorized modes helps to achieve better health outcomes in a number of ways. Walking can be an important source of physical activity, and better accommodation of these modes can lead to fewer injuries and deaths due to conflicts with motorized vehicles. Encouraging walking over private car use also reduces energy consumption and mitigates transportation's negative environmental impacts, such as greenhouse gas and other emissions. In terms of livability, the inclusion of these modes allows for a broader range of transportation choices for residents and visitors, lowers the costs of accessing destinations, and contributes to vibrant local commercial districts.

To achieve these ends, urban planners are increasingly interested in having more sophisticated models to estimate pedestrian travel, which can help evaluate the outcomes of various policies. As a result, there is a critical need for pedestrian planning tools that operate at a scale that is sensitive to pedestrian needs and the quality of walking environments. A better pedestrian travel demand model can improve sensitivity to pedestrian-relevant factors, provide better prediction of mode shifts, yield results that are responsive to socioeconomic and policy changes, and make outputs useful for policy making. In addition, efforts are keen to link these model outcomes to other tools for assessing other issues such as air quality, public health, and road safety.

However, pedestrians have often been left out of the modeling process. This is hampered by either misguided omission in favor of motorized planning goals, lack of fundamental research on pedestrian travel behavior, inadequate pedestrian data and corresponding information about the built environment at appropriate scales, or inability to process pedestrian demand due to computation limitations.

Today, many of the barriers mentioned above have been surmounted. First, pedestrian research has advanced over the last two decades thanks to the vast improvement of pedestrian data availability, particularly in the literature on linking travel behavior to the built environment (Cervero, 2003; Ewing & Cervero, 2010; Hahm et al., 2019; Khan et al., 2014; Targa & Clifton, 2004; Tian & Ewing, 2017). These studies have identified many factors that influence how frequently people walk, whether people walk, and activity locations where people walk. For example, Singleton et al. (2014) created the Pedestrian Index of the Environment (PIE), which included six measures such as activity density, transit access, block size, sidewalk extent, bicycle facilities, and urban living infrastructure. PIE was found to be significant and positively related to whether people walk or not.

Some efforts were taken to develop stand-alone pedestrian planning tools or consider pedestrians in travel demand models that operate at an urban scale. Most of the studies take a simulation approach and model individual microscopic pedestrians under specified locations, such as train stations, public streets, airports, and shopping malls, with a focus on wayfinding, evacuation, and response to crowding and queuing (Borrmann, Kneidl, Köster, Ruzika, & Thiemann, 2012; Erdmann & Krajzewicz, 2015; Kielar & Borrmann, 2016). Some travel demand models have begun to consider walk mode into the framework that is primarily designed to simulate car traffic (Delaware Valley Regional Planning Commission, 2011; Moeckel et al., 2020; San Francisco County Transportation Authority, 2002). Only a few models emphasize estimating the amount and nature of pedestrian travel (K. J. Clifton et al., 2016b). For example, the Model of Pedestrian Demand (MoPeD) is one of the first pedestrian-centric transport models. One of the contributions of MoPeD is using a finer-grained scale called Pedestrian Analysis Zone (PAZ) rather than coarse transportation analysis zones (TAZs). This can better represent the pedestrian behavior and built environment variables.

Despite pedestrian research and modeling practice have attracted more attention in transportation planning and have contributed to a number of improvements, pedestrian travel demand models still face several challenges.

Fundamentally, further theoretical studies are needed to investigate individuals' walking activity, including not only the interpersonal differences but also intrapersonal changes in behavior over time and space. An additional challenge for investigating intrapersonal variability is the scarcity of longitudinal travel behavior data. Collecting longitudinal travel behavior data and exploring the variation in pedestrian travel would be useful for predicting individuals' weekly walking levels, allowing for a more accurate assessment of health benefits.

Additionally, research on built environment measurements such as the walkability index attempted to cover a wide range of spatial factors while overlooking the complexity and transferability when applying those metrics to modeling practice for exploring future scenarios.

Because the availability and quality of data vary greatly across study areas, those metrics may be a barrier to incorporate built environment influences into planning tools.

Finally, more efforts are needed to apply theoretical and empirical analysis of pedestrian behavior into practice for use in planning and policy decisions. While the pedestrian-centric planning tool MoPeD has a wide range of benefits to modeling pedestrian travel behavior, many challenges remain in further development and implementation. For example, MoPeD has an issue in model efficiency due to the heavy computational burden of processing pedestrian demand at fine spatial scales. This also limits its ability to run large-scale scenarios. Moreover, MoPeD is limited to its nature of aggregated modeling. Average social-demographic attributes are simulated at the zonal level, which makes it less sensitive to demographic changes such as aging or car ownership in scenario application. Agent-based transport models have great opportunities to simulate pedestrian travel at a fine spatial resolution and at an individual level. However, most agent-based transport models either overlook pedestrian activities or have limited representation of pedestrian behavior. For example, the Microscopic Transportation Orchestrator (MITO) is a microscopic travel demand model that generates trips for every individual. In this modeling suite, the multinomial logit model that is part of MITO estimates the shares of walk trips but subsequently drops those trips from further analysis. Walk trips are processed in TAZs, and there are no built environment measurements related to pedestrians included in the models.

Based on the discussion above, the research gaps can be summarized as 1) lack of longitudinal travel behavior data; 2) limited understanding of travel behavior variability; 3) Poor transferability of built environment measurements; 4) insufficient work of incorporating pedestrian travel into modeling practices. Those gaps mentioned above could limit the ability of researchers, planners, and policymakers to assess future transport demand and evaluate various future plans and policies. A clear understanding of pedestrian behavior and better modeling of pedestrian activities could help to anticipate and plan for future transportation needs.

1.2. Research Objectives

This dissertation aims to address the gaps and limitations of the fundamental research and modeling practices mentioned above. Therefore, the goal of this dissertation is to improve pedestrian representation in travel demand models to better assess travel outcomes and evaluate policies and scenarios. To achieve this goal, four key requirements are identified to improve the performance of the pedestrian planning tools. They are model accuracy, sensitivity, transferability, and efficiency.

Model accuracy is one of the essential aspects. More accurate model outcomes such as mode shifts, walk trip lengths, and pedestrian flows can provide a clear picture of future transport demand, which results in better decisions in land use and transport infrastructure investments. Moreover, model outputs such as car volumes, physical activity volumes, and trip trajectories are increasingly valued as important inputs to emissions, climate, health, and equity analysis. Therefore, more accurate model outcomes can also help to assess other issues such as air quality, public health, and road safety. Model inaccuracy in pedestrian travel demand models can be caused by several reasons, including inaccurate fundamental data, improper modeling approaches as well as inappropriate spatial and temporal resolution.

Model sensitivity is an important indicator to show the explainable power of the model. It also reflects the model's accuracy. If key influencing factors are missing from the model, it can lead to model inaccuracy and reduce the interpretability of the policies and scenarios that have changes in the missing factors. The effects of the built environment on pedestrian travel behavior have been extensively investigated, but the incorporation of pedestrian-related built environment factors into transport planning tools is still lacking. Furthermore, the influences of time-varying factors such as the day of the week and weather are rarely explored or considered in travel demand models.

Model transferability plays an important role in transportation planning tools. The spatial transfer of an existing model from one location (estimation context) to another (application context) can save a deal of time and efforts in data collection and model development (Karasmaa, 2003). Particularly when pedestrian travel behavior data is scarce, spatial transferability can help to boost the application of pedestrian-centric planning tools. Moreover, temporal transferability is critical for almost all planning tools because the primary purpose of the pedestrian travel demand model is to evaluate future policies and scenarios by using models estimated from historical data. The methods for constructing built environment measurements are complex and vary largely. This may cause issues in spatial and temporal transferability.

Model efficiency is also essential for developing practical planning tools. Despite significant advances in computer technology, using fine spatial scales in large-scale study areas can boost the computational burden. The size of the context and the processing zone system have a large impact on model run time. In addition, the modeling approach can also have an effect on model efficiency. For example, pedestrian route choice is typically more expensive in terms of run time and memory usage.

Each research gap mentioned in Section 1.1 can lead to one or more issues in pedestrian modeling in terms of these four model requirements. When the gaps are addressed, and these four requirements are fulfilled, pedestrian planning tools will become widely applicable and useful for evaluating policies and scenarios.

1.3. Approaches

To achieve the end, research gaps are addressed through different theoretical, empirical, or analytical approaches involving advanced model development and novel longitudinal data collection.

1.3.1. Model Enhancement and Integration

To fill the gaps in modeling practices, an integrated modeling framework is proposed that incorporates a fine-grained resolution model of pedestrian demand (MoPeD) into a sparser spatial resolution of an agent-based transport model (MITO). MoPeD and MITO have their respective strengths and can complement each other. The integrated model would benefit from MoPeD's fine spatial resolution and its better representation of the built environment, as well as MITO's agent-based environment and its good representation of motorized modes.

Before the integration, it is critical to enhance MoPeD and resolve some of its limitations. First, the appropriate spatial resolution for measuring pedestrian environment and modeling behavior is explored at the scale of a metropolitan area. This investigation is useful for determining the spatial resolution requirements when transferring MoPeD to other contexts. To further improve the transferability of MoPeD, a new pedestrian accessibility measurement is created, and it replaces PIE in walk mode choice models. Then, the existing walk destination choice models are enhanced by using the whole universe of choices and a two-step multinomial logit model technique. From the technical view, the run time and memory saving of the model is upgraded by building up the whole model in the Java environment.

Once MoPeD is enhanced, it is incorporated into MITO. The integrated modeling framework is developed in the context of the Munich Metropolitan area. Thus, the first step is to gather input data for pedestrian modeling, such as fine zone systems, pedestrian street networks, and data for built environment measurements. Then, the step of integration is not straightforward due to the different spatial units and modeling sequences in these two models. Coding interfaces are created to transfer MITO trip locations to PAZs used in MoPeD. Moreover, a hybrid trip decision process is developed based on the modeling sequences in both models. Finally, the integration work also includes a large part of the development of source code in Java.

1.3.2. Novel Data Collection and Fundamental Research

To fill the gaps in fundamental data and research of travel behavior variability, a survey is designed to collect an individual's longitudinal travel behavior in a way that can enrich the data sources of week-long travel behavior and can investigate how and why an individual's travel behavior varies over time. The survey collects longitudinal and passive data – Google Location History (GLH) – that records an individual's trip diaries with coordinates and a board time horizon, which is a major innovation compared to previous studies. Besides the individual and household characteristics, the survey also asked questions about the occurrence of major life events (marriage, children, changing jobs, etc.) in the past 18 months. This intends to explore the relationship between life events and the change in travel behavior. Next, the collected GLH data is employed to have a closer investigation of week-long travel behavior and travel behavior variability. Specifically, four research tasks will be carried out to provide a better understanding of intrapersonal differences in behavior over time.

First, the specific measurements of travel behavior variability are examined across various temporal scales (day-to-day, week-to-week, and month-to-month). This analysis addresses the empirical question about how different travel metrics vary over time, including the number of trips, total travel time, walk time, number of walk trips, and start time of day. In addition, the author attempts to find out the correlation between socio-demographic attributes and an individual's travel behavior variability level.

Second, the assessment of weekly walking behavior (such as the weekly number of walk trips, weekly walk time, and weekly walk distances) is compared by using four different data sources, including GLH, a one-day travel survey, a seven-day travel survey, and a self-reported physical activity questionnaire. The comparison can highlight the strengths and limitations of each data collection method. Furthermore, there is a debate about using the single-day survey to estimate an individual's weekly travel behavior. This investigation may also provide some analytical evidence for the debate.

Following the empirical examination, the next task is to investigate the potential determinants of weekly travel behavior. The first investigation explores the explanatory factors such as socio-demographic characteristics, household attributes, travel-related information, weather-related attributes (e.g., weekly precipitation, weekly temperature), and time-related variables (e.g., month, has public holiday). Linear panel regression models are employed for predicting weekly walk time and physical activity volume. The majority of the samples have zero observation in weekly cycle time. Therefore, a binary logit model is first involved in determining whether people have cycle activities, and then a linear panel regression model is used to estimate weekly cycle time. The second investigation looked at how major life events influence weekly travel behavior. Due to the limited number of life events observed among the survey participants, the analysis is conducted through a series of descriptive statistics. These two studies can provide answers to the questions of the determinants of weekly travel behavior and the extent to which different attributes affect weekly travel behavior.

In contrast to the theory-driven approaches mentioned above, a data-driven approach is applied in the final step to further investigate an individual's travel routines and their potential disruptions. Unsupervised learning approaches are widely used in the transportation field to explore activity patterns (Cui et al., 2018; El Mahrsi et al., 2017). Here, the author will use cluster algorithms to group days with similar travel characteristics based on mode usage by time of day. The resulting clusters represent the regularity in travel behavior, while the difference between clusters represents some level of disruption.

1.4. Structure of the Dissertation

Figure 1 presents the road map of this dissertation. It reflects the dissertation goals and research gaps for defining four research objectives as well as the approaches to achieve these objectives. The structure of the dissertation is organized into several chapters that align with the different steps in the road map. The motivation and goals of this dissertation are presented in Chapter 1. The remaining chapters are summarized in the following paragraphs.

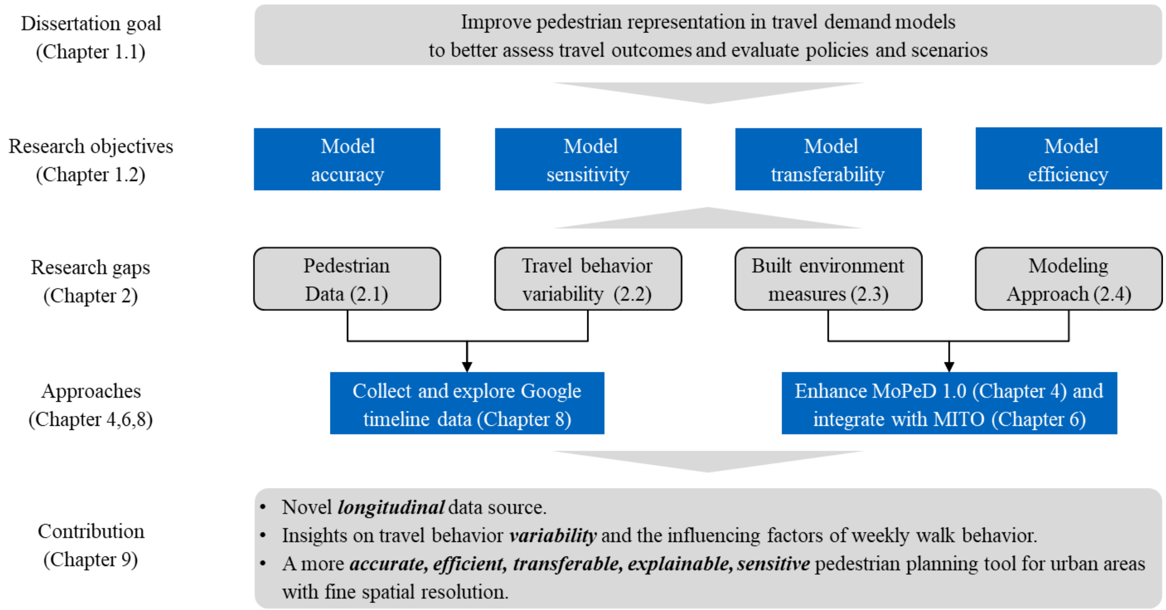


Figure 1 Framework and structure of the dissertation

Chapter 2 presents the progress of research on pedestrian behavior and existing modeling practices. First, various data collection methods for analyzing pedestrian behavior are reviewed. The strengths and limitations of different data sources are discussed. Next, the literature on built environment influences on walking behavior is reviewed, pointing out that there is a need to call for standardization and transferability of built environment measurements. Studies on travel behavior variability are also reviewed, with the conclusion that more emphasis should be placed on understanding how and why an individual's travel behavior varies over time. Finally, some existing travel demand models and their representation of pedestrian travel are demonstrated. Based on the existing literature and model practices, the author summarizes four major research gaps (as shown in Figure 1): 1) lack of longitudinal travel behavior data; 2) Poor transferability of built environment measurements; 3) limited understanding of travel behavior variability; 4) insufficient work of incorporating pedestrian travel into modeling practices.

Chapter 3 introduces the study areas that will be employed in this dissertation. MoPeD 2.0 was developed in the context of the Portland metropolitan area. The application of MoPeD 2.0 is carried out in the Portland Central City area. The Munich metropolitan area will be used for the development of the integrated model and its scenario applications.

Chapter 4 enhances the stand-alone pedestrian model MoPeD in different aspects. First, it explores the appropriate spatial resolution for measuring pedestrian environment and modeling behavior at the scale of a metropolitan area, concluding that there is no optimal spatial resolution, and the selection of spatial resolution depends on the type of applications and the availability of spatial data needed to implement. Then, efforts are made to enhance MoPeD in terms of model efficiency, transferability, accuracy, and sensitivity, including run-time upgrade, reconstruction of new pedestrian accessibility measures, and enhancement of the walk destination choice model.

Chapter 5 applies the enhanced MoPeD to the Portland Central City area to assess realistic urban and transportation scenarios from the Portland Central City 2035 Plan. The impacts of scenarios are evaluated in terms of changes in walk shares, average trip length, and pedestrian flows. The application demonstrates the capability of MoPeD in reflecting land use scenarios while also revealing the limitations of MoPeD, such as its insensitivity to the quality of network infrastructures.

Chapter 6 presents the integrated modeling framework that incorporates MoPeD into MITO. The author specifically addresses how the integration of these models improves the representation of pedestrians. The approach of integrating MoPeD and MITO is presented. Then, the input data containers are prepared for the Munich context. Since the mode choice and destination choice models were developed in the Portland context and then married into the Munich study area, a model transfer/calibration process is carried out. After applying the integrated model to the Munich study area, model performances are compared to the MITO stand-alone model. It concludes that the integrated model can provide more accurate travel outcomes such as walk shares, walk trip length distribution, the spatial distribution of walk trips, and physical activity volumes.

Chapter 7 applies three scenarios with the integrated model. The first application investigates the pedestrian travel demand changes and health impacts of a teleworking scenario. The second application investigates the changes in walking behavior under a radical scenario where jobs and workers are perfectly assigned to achieve minimal commuting time. The third application tests the concept of a car-free city under some draconic scenario settings, such as allowing only one car in the household. The application results prove the capabilities and sensitivity of the integrated model to various policies and scenarios. On the other hand, they also highlight the limitations of the model.

Chapter 8 demonstrates the novel longitudinal data from Google Location History and analyses of travel behavior variability. First, it documents the development and administration of the data collection effort. Due to privacy concerns and the effort required for in-person recruitment, the number of recruited individuals is limited. In the end, there are only 27 valid GLH datasets used in this research. Then, the methods of data filtering and processing are presented. After that, some descriptive statistics are presented, such as the socio-demographic distribution of samples and respondents' travel behaviors. Finally, the processed GLH data is employed to have a closer investigation of travel behavior variability. Four research tasks are carried out 1) empirical examination of travel behavior variability; 2) assessment and comparison of weekly walking behavior using four data sources; 3) investigation of the potential determinants of weekly travel behavior; 4) application of clustering algorithms to further investigate individual's travel routines and disruptions. The analysis of travel behavior variability proved that individuals have a great deal of day-to-day variability. Week-to-week travel behaviors have relatively low dispersion, while people tend to have periodical behavior on a monthly scale. The investigations of weekly walking behavior discovered its association with the weather, day of the week, public holidays, and life events. Although these analyses are conducted with a small sample size, they still provide innovative and exploratory insights into pedestrian travel behavior research.

Chapter 9 concludes the dissertation with a summary of the research motivation and key achievements. It then highlights the key contributions and implications of this research. In terms of fundamental research, this dissertation makes strides toward richer data sources and a better understanding of pedestrian travel behavior variability. In terms of modeling, this dissertation develops a more accurate, sensitive, transferable, and efficient pedestrian planning tool for delivering travel outcomes and evaluating policies and scenarios. Finally, the chapter concludes with a discussion of research limitations and opportunities for future work.

2. Literature Review

Early studies of pedestrian travel were limited and primarily hampered by the lack of pedestrian data and corresponding information on the built environment. There was also a notable gap between research on pedestrian behavior and modeling practices. This was mainly because of the car-centric design goals of most travel demand models, as well as their inability to process pedestrian demand at the appropriate scale due to computational limitations. However, many of these obstacles have now been overcome. The availability of data has improved significantly. Aided by a wealth of data, scholars have begun to investigate the mechanisms and impacts of walking behavior. In the meanwhile, researchers and transport planners also have made efforts to improve pedestrian planning tools.

The goal of this review is to present the progress of research on pedestrian behavior and existing modeling practices. Section 2.1 first presents various data collection methods for analyzing pedestrian behavior. Then Sections 2.3 and 2.2 concentrate on the behavioral research of pedestrian studies. They summarize the literature on built environment influences on walking behavior and temporal patterns of walking trips. Section 2.4.1 focuses on the progress of pedestrian modeling practices. The author discusses the spatial resolution used in pedestrian travel demand models. After that, some existing travel demand models and their representation of pedestrian travel are demonstrated. Based on the existing literature and model practices, the author summarizes some research gaps in Section 2.52.5, which also serves as the basis for this dissertation framework.

2.1. Data Collection Methods for Studying Pedestrian Behavior

In the early stage, one common problem that the researchers faced when modeling pedestrian demand was the barrier of the data collection on pedestrian behaviors and environments. Data sources for investigating pedestrian travel behavior are generally grouped into two categories: self-report surveys and objective measures.

Household travel surveys (HTS) are used primarily by transport scientists and practitioners to explore pedestrian travel behavior (K. Clifton & Muhs, 2012; Fairnie et al., 2016). HTS collects characteristics of individuals and their households and ask respondents to record trip diaries over a specified time period. Every interaction the respondents have with the transportation system, including active travel, is recorded. Trip distance, speed, mode, and departure/arrival time are all captured. HTS generally provides rich information on travel purposes, travel behavior, and sociodemographic attributes. For example, the national household travel survey, *Mobilität in Deutschland* (MiD), is the most widely used data source for transport studies in Germany (BMVI, 2019b). It is a large cross-sectional survey conducted in 2016-2017 consisting of 156,420 households containing 912,002 trips. Respondents complete 24-hour travel diaries to give on their interactions with the transport network over the course of a day. Trips recorded in MiD with main mode walking can be used to investigate pedestrian travel behavior. However, walk distances are self-reported and origins and destinations of trips are recorded with very coarse geographic locations. This is a barrier for pedestrian travel behavior studies to associate built environments with walk trips.

Although HTS has made improvements that do a better job of capturing pedestrian behaviors, including prompts for infrequent and short-distance trips, there would be several pitfalls in using household travel surveys in this way. First, due to the nature of self-reporting, HTS may still be biased in the trip distance (Aschauer et al., 2018; K. W. Axhausen et al., 2007; Reiffer et al., 2022; Wolf et al., 2003). This may cause inaccurate measures of walking behavior like the willingness to walk and destination choice of walking trips. Moreover, most HTSs only require respondents to record trip diaries on a single day (Kunert et al., 2002). As a result, they are limited in capturing important variations in an individual's travel across the day of the week or seasons of the year. In some areas and during specific seasons, weather can be a significant factor affecting pedestrian trips (Aultman-Hall et al., 2009; Miranda-Moreno & Lahti, 2013). On the other hand, understanding the regularity of travel behavior like walking and cycling is an important component of health research (Merom et al., 2010). For example, public health guideline recommendations use a week-long timescale for assessing walking, cycling, and other physical activities.

To overcome this issue, longer-term HTS or panel surveys were conducted and widely used in transport studies. The Uppsala household travel survey and Reading activity survey, which are widely used in the literature, were conducted in the early 1970s. The Uppsala data records travel diaries over 35 consecutive days, while the Reading data had seven days (Hanson & Huff, 1982; Pas, 1983). The UK national travel survey was carried out with a seven-day travel diary which is helpful in exploring pedestrian travel behavior patterns across weekdays and weekends. However, this dataset is limited in capturing the variance of short walk trips because respondents are only required to record walks of less than 1.6km on one of the seven days (The

Department for Transport, 2020). The Mobidrive six-week travel diary dataset collected in 1996 also provides a rich source for travel pattern variability studies (K. Axhausen et al., 2002). However, one major issue of longer-term travel diaries or panel surveys is respondent fatigue (K. W. Axhausen et al., 2007).

With advances in technology, objective measurements, such as global position systems (GPS), accelerometers, and pedometers, have emerged. GPS technology is widely used in studying the relationship between pedestrian travel behavior and the built environment. Lue and Miller (2019) investigated pedestrian route choice behavior by using smartphone-based travel survey data collected in Toronto. Their study demonstrates the feasibility of using GPS data for pedestrian travel behavior studies. Objective measurements are also broadly employed in health studies to capture physical activity such as walking and cycling. A recent study by Chaix et al. (2019) used GPS receivers and accelerometers to capture walking during the usage of public transport. These technologies facilitate the collection of location-based data at a high resolution over long periods of time without the issues of respondent burden. However, they are relatively expensive due to the cost of equipment and participant burden, so they are usually employed only with a small population. In addition, they require data processing and algorithms to infer the activity type and travel mode. Assumptions used for data processing may result in large differences in understanding pedestrian travel behavior.

2.2. Travel Behavior Variability

Transportation behavior and policy researchers raise important questions about how and why an individual's travel behavior might vary over time (Hanson & Huff, 1982; Jones & Clarke, 1988; Pas, 1983). The study of variability can, to some extent, improve the goodness-of-fit of the transport models (Pas, 1987). With an interest in crafting policy interventions for congestion, the environment, and health, it is important to understand how an individual's behavior might change under different circumstances (Jones & Clarke 1988). Understanding the regularity of walking travel behavior can also better predict individuals' level of walking, which in turn allows for a more accurate assessment of people's health benefits. Thus, there has been a long interest in analyzing travel behavior variability.

2.2.1.1. Measures of Travel Behavior Variability

Travel behaviors are complex and are influenced by many factors, some of which can change daily, including activity participation, weather, and availability of modes. Previous studies have investigated travel behavior variability across different travel metrics. Pas & Koppelman (1986) looked at the changes in daily trip frequency and found considerable differences. The degree of variation was dependent upon the socio-demographic characteristics. For example, individuals with fewer economic and role-related constraints (such as non-employed individuals) were observed to have higher levels of intrapersonal variability in daily trip frequency.

Most of the research uses a combination of travel measures to represent an individual's travel-activity pattern (Dharmowijoyo et al., 2017; Hanson & Huff, 1982, 1986, 1988; Manley et al., 2016; Pas, 1983; Schlich & Axhausen, 2003; Susilo & Axhausen, 2014; Tarigan et al., 2012). The most common travel measures are activity types, trip frequency, time use, and travel distance. For example, Hanson & Huff (1986) derived a set of travel measures to summarize an individual's travel-activity pattern, such as the proportion of different activity types and the number of stops per tour. Additional dimensions, including multimodal travel, location choice, and route choice, are considered in measuring travel patterns (Buliung et al., 2008; Chikaraishi et al., 2009; Heinen & Chatterjee, 2015; Manley et al., 2016; Pas & Sundar, 1995; Streit et al., 2015). As a result, different levels of variability were found for different travel measures. The variability of the travel-activity pattern increases if more complexity is considered in the travel pattern measurement (Schlich & Axhausen, 2003).

In more recent works, researchers found that the order of activities and travel events play an important role in the regularity of travel behavior (Goulet-Langlois et al., 2018; Prelipcean et al., 2018; Xianyu et al., 2017). Travel-activity sequences are similar among weekdays while being different from weekend sequences. (Xianyu et al., 2017).

2.2.1.2. Temporal Scales of Travel Behavior Variability

The level of variability could be quite different when analyzed at different temporal scales. The majority of previous studies examined the day-to-day changes in travel behavior. Hanson & Huff (1982) investigated daily travel activity patterns and concluded that employed men and nonworking women tend to have a higher level of repetition in their daily travel-activity patterns. Pas & Sundar (1995) analyzed the daily variability in trip chaining generation, departure time from home, and route choice. They found that individuals from two-person households have substantial day-to-day variability in travel behavior. Susilo & Axhausen (2014) measured the degree of repetition of an individual’s daily activity-travel-location pattern. The results indicated a significant variance in travel behavior patterns, which is influenced by personal and household characteristics and the accessibility of the activity locations.

Some studies focused on analyzing the day-of-week variability of travel behavior. In general, they found that weekend-to-weekday variability is much greater than weekday-to-weekday or weekend-to-weekend variability (Buliung et al., 2008; Dharmowijoyo et al., 2017). In terms of activity space, it is found that weekday activity spaces are more compact than those at weekends (Dharmowijoyo et al., 2014). This is expected since individuals have less flexibility in their schedules during workdays. Furthermore, Xianyu et al. (2017) discovered a similar pattern that weekday activity-travel sequences are more similar while being different from weekend sequences.

A few studies have investigated the week-to-week variability or even month-to-month variability due to the lack of data. Hanson & Huff (1988) attempted to find out the repetition of travel patterns from week to week by using a five-week travel survey. However, they didn't find a weekly cycle to be sufficient to define travel routines. Tarigan & Kitamura (2009) studied the week-to-week variability in leisure trip frequency. Their results revealed that the mean number of trips per week influences the week-to-week variability in the number of trips for socializing activity types.

2.3. Association with Built Environment

Since the 2000s, many scholars have considered theories about the relationships between the built environment and travel behavior (Boarnet & Crane, 2001; Ewing & Cervero, 2001). Thanks to the improvement of travel survey data and data collection technologies, a number of studies on the association between pedestrian behavior and the built environment have been carried out. These studies identified the various factors which impact pedestrian behavior. It was confirmed many times that environmental influences are related to walk trip frequency, walk mode choice, walk destination choice and walk route choice (Ewing & Cervero, 2010; Khan, M. Kockelman, & Xiong, 2014; Kuzmyak, Walters, Bradley, & Kockelman, 2014). However, the set of factors and the magnitude of their effects vary across different aspects of walking activity.

The role of the built environment on the choice to walk and how frequently people walk has been investigated in many studies. Researchers have identified a common set of built environment attributes that are related to walk trip frequency and walk mode choice. They are residential and employment densities, land use diversity, and pedestrian network connectivity (Ewing and Cervero, 2010; Saelens and Handy, 2008; Saelens et al., 2003; Guo et al., 2007). A few studies also pointed out the positive relationship between walk trip generation and sidewalk conditions (Ewing and Cervero, 2010; Desyllas et al., 2003; Boarnet et al., 2011). In addition to discussing influencing factors, researchers also worked on finding out appropriate spatial scales for measuring these attributes (Gehrke and Clifton, 2014). They emphasized the need to use small geographic scales in pedestrian travel behavior research.

The influence of the pedestrian environment on destination choice has been examined in a few studies. Khan, Kockelman, and Xiong (2014) found that intrazonal non-motorized trip likelihoods rose with higher street connectivity, transit availability, and land use entropy. K. J. Clifton et al. (2016a) developed pedestrian destination choice models for different trip purposes. They concluded that pedestrian destination choice was primarily influenced by distance, while it was also sensitive to the pedestrian environment (Ewing & Cervero, 2010; Saelens & Handy, 2008; Saelens et al., 2003). Their research found that retail employment, household density, and parks are positively related to the likelihood of choosing a destination. On the other hand, the degree of slope, the existence of freeways, and the share of industrial-type employment are barriers for pedestrians in choosing a destination.

Pedestrian route choice is an emerging topic. There is comparatively less research devoted to how pedestrians choose routes, and the built environment attributes that they consider for various trip purposes (Borst et al., 2009; Broach & Dill, 2015; Koh & Wong, 2013; Rodríguez et al., 2015). Lue and Miller (2019) investigated the influences of street infrastructure and built environment on pedestrian route choice behavior by using smartphone-based travel survey data collected in Toronto. They found that the number of signalized intersections and traveling along sidewalks were significant variables in the pedestrian route choice model. Broach & Dill (2015) presented a pedestrian route choice model estimated from revealed preference GPS data. Their research found that pedestrians are sensitive to attributes of the walking network, intersection crossing aids, and elements of the street environment. Rodríguez et al. (2015) examined the influence of the built environment on pedestrian route selection among adolescent girls. They pointed out that shorter distances had the strongest positive association with route choice. In addition, a set of built environment variables was found to be associated with better walking routes in their research, such as the presence of a greenway or trail, higher safety, the presence of sidewalks, and the availability of destinations along a route.

This review of walking and the built environment in different aspects of walking activity is useful for the selection of variables in the later stages of model estimations.

2.4. Modeling Practices

2.4.1. Spatial Resolution of Pedestrian Travel Demand Modeling

Finding an appropriate spatial scale at which to model travel and land use patterns is a critical and often challenging decision, made even more difficult with the additional constraint of modeling pedestrian travel. Singleton et al. (2018) conclude that one of the perspectives of the research on pedestrian modeling is a traditional four-steps pedestrian model with finer spatial resolution. However, there are only a few operational models estimating pedestrian demand with a fine zoning system. Generally, the zone system could be divided into two categories: gradual raster cells and uniform raster cells.

Moeckel and Donnelly (2015) develop a methodology to create a zoning system with gradual raster cells. The size of the raster cell depends on the population and employment density. Finer raster cells are used in areas with higher density, while coarse raster cells are generated in low-density areas. A most recent study by Okrah et al. (2017) implements this methodology into practice to find the optimal spatial resolution for handling non-motorized transport. They use the total network length in a zone to define the size of raster cells, and the threshold value of network length ranges from 50m to 5000m. By comparing the predicted traffic assignment volume to the reference volume, they conclude that a total network length of 1,000 m per zone as the optimal spatial resolution for their study area.

Regards the uniform raster cells, K. J. Clifton et al. (2016b) established a model of pedestrian demand (MoPeD) by using a uniform zoning system. It follows the traditional four-step model, but it changes the spatial unit from Transportation Analysis Zone (TAZ) to a finer spatial scale called a Pedestrian Analysis Zone (PAZ) defined by an 80 m × 80 m grid cell and an aggregation of these PAZs into 400 m × 400 m zones called superPAZ. In addition, there have been some studies examining the pedestrian behavioral response to built environment measures taken at various scales (Gehrke & Clifton, 2014). These studies operate a much courser resolution (400 m buffers and larger). To date, there has been no exploration of the responsiveness and efficiency of a pedestrian demand model to finer uniform raster cells (< 400 m × 400 m).

Although the studies mentioned above prove that finer spatial units can better represent pedestrian behavior and react to the changing pedestrian environment, they also identify the difficulties and challenges in computational burden and data collection when implementing finer scales. Some researchers contend that reducing zone sizes to parcel level will cause an exponential increase in the size of impedance matrices in the traffic assignment step, which could escalate the computational burden (Moeckel & Donnelly, 2015; Okrah et al., 2017). In terms of data collection, although the data quality of archived household travel behavior data and land use data has been improving at very fine scales, it could be challenging when forecasting these data at the same fine level of detail.

2.4.2. Model of Pedestrian Demand (MoPeD)

2.4.2.1. Overview

With the advances in data availability and computation power, some efforts were taken to develop stand-alone pedestrian planning tools or better represent pedestrians in travel demand models that operate at an urban scale. One of the recent works, which achieved this end, is the modeling framework MoPeD developed by Dr. Clifton and her team (Clifton et al. 2016a; Clifton et al. 2016b). The framework of MoPeD is illustrated in Figure 2. The pedestrian prediction tool is integrated with a four-step urban model, in this case, the Portland Metro Regional Model.

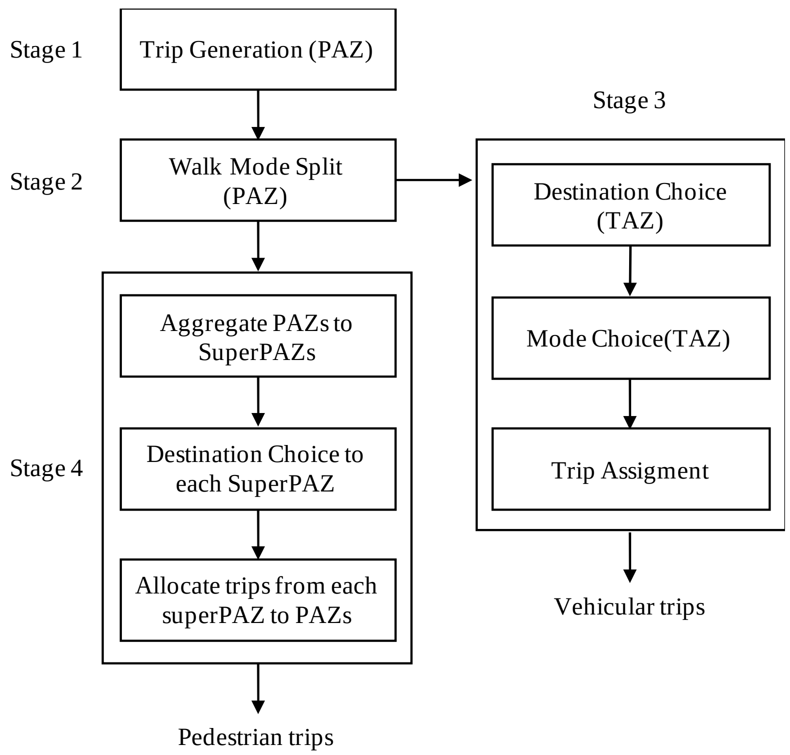


Figure 2 Modeling framework of MoPeD (adopted from (K. J. Clifton et al., 2016))

2.4.2.2. Contribution

One of the contributions of MoPeD is that stage 1 (Trip Generation), stage 2 (Walk Mode Split), and stage 3 (Pedestrian Destination Choice) are all modeled at a Pedestrian Analysis Zone (PAZ), which is represented as an 80m by 80m grid cell. See the comparison of the PAZ zonal structure with the larger Transportation Analysis Zone (TAZ) used in regional transportation models in Portland, OR, in Figure 3 below.

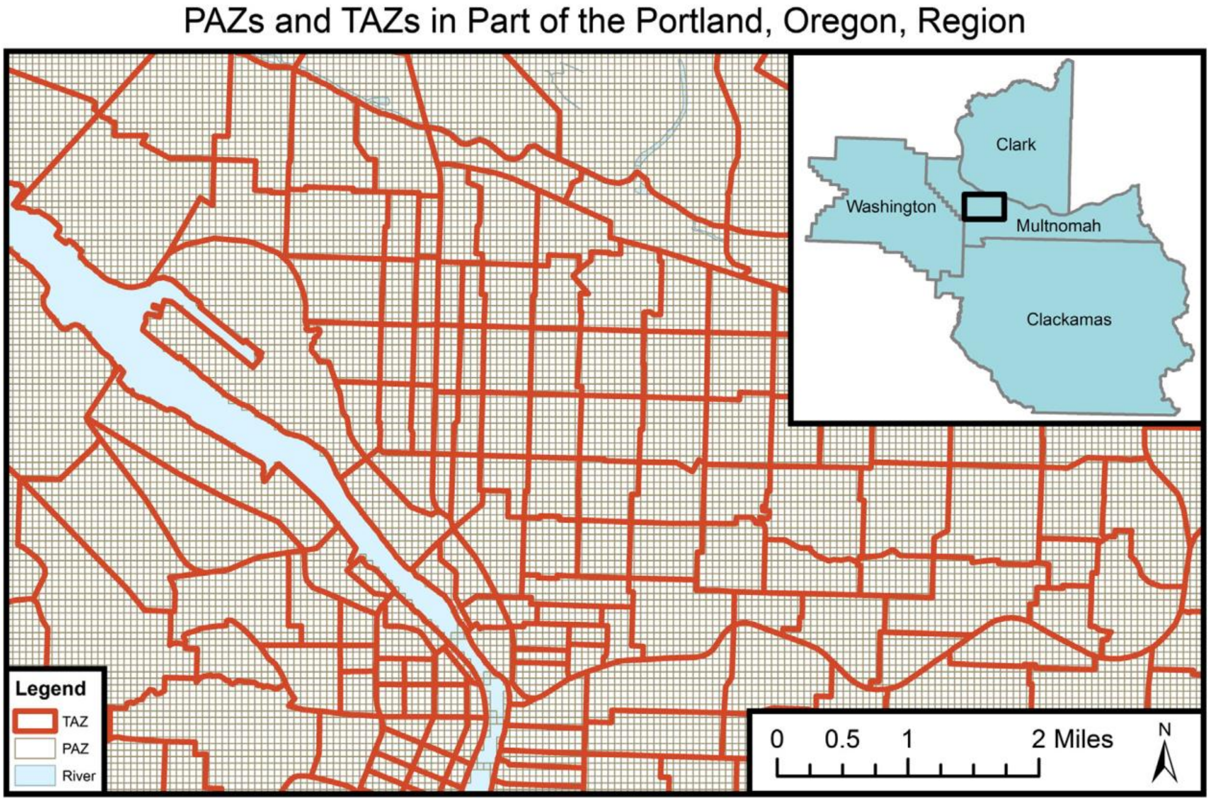


Figure 3 Comparison of two zonal structures—PAZs and TAZs—in part of the Portland, Oregon, region. (K. J. Clifton et al., 2016b)

MoPeD starts with trip generation at the finer spatial resolution. Stage 2 models the probability that these trips are made by walking. Those trips that are made by walking are then distributed to destination PAZs using a destination choice model in Stage 3. In the traditional four-step modeling process, mode choice commonly follows the trip distribution. Nevertheless, this opposite order or model steps was chosen in MoPeD because destination choice works substantially differently for walk trips and non-walk trips. By modeling the choice walk/non-walk first, the conditions for the destination choice model are largely improved. Also, since MoPeD implements at a fine spatial resolution, making mode choices prior to the destination choice can avoid dealing with massive distance matrices.

Trips that are not made by walking (vehicular modes) as determined by Stage 2 can then be aggregated to larger zonal structures (such as TAZs) and handled by the more typical transport demand modeling process.

Another contribution of MoPeD is that the model utilized a measure of the pedestrian environment (Pedestrian Index of the Environment, or PIE) at the PAZ scale (Singleton et al. 2014). It consisted of six measures that included activity density, transit access, block size, sidewalk extent (miles of continuous sidewalks within 0.25 miles), bicycle facilities, and urban living infrastructure, which means the number of shopping and service destinations in the neighborhood. PIE was used as one predictor in walk mode choice models, and some other pedestrian supports and barriers variables (e.g., slope, park, and freeway) were also considered in walk destination choice models. This improvement of model practice is aligned with the findings of built environment influences on walking behavior presented in Section 2.3. As a result, the MoPeD model can better represent pedestrian behavior and more accurately simulate walk trips.

The authors also pointed out that there is a wide range of applications for MoPeD. For example, the model can simulate urban-wide/regional-wide pedestrian activities, which can be used for various planning and policy applications. Moreover, there is the opportunity to link the outputs of these models to health assessment tools such as the Integrated Transport and Health Impact Modeling Tool (James Woodcock et al., 2013).

2.4.2.3. Discussion on MoPeD

Although MoPeD has a wide range of benefits to modeling pedestrian and transport planning. Many challenges remain in further development and implementation.

First, the first implementation of MoPeD in R had slow run times and could only run a small subset of the Portland region at a time. It was not able to handle the entire Portland metropolitan area due to the heavy computational burden of this fine spatial resolution.

Second, one of the limitations of the previous MoPeD model is the complexity of pedestrian-built environment representation. As described in 2.4.2.2, the pedestrian index of the environment (PIE) was used in MoPeD. With a rich spatial database, PIE was able to better represent walking behavior. However, it was less transferable to other applications due to the requirements of detail land use data at a fine spatial resolution. In addition, it was difficult to assess future scenarios and policies because it was challenging to predict the specific changes to the built environment, such as sidewalks and cycle lanes, at this fine spatial resolution.

Another limitation of the previous MoPeD model is the destination choice. The choice set generation method is a major challenge. The previous MoPeD model employed a simple random sample of ten zones for the choice set. There is an ongoing debate on destination choice sets. Singleton and Wang (Singleton & Wang, 2014) found that estimation results did not change substantially when using choice sets of 10 or 25 alternatives. Using the universe of destination alternatives might have conflicts with an individual's behavioral decision-making principle. However, some authors still suggested using full samples or sampling a larger number or proportion of alternatives (Nerella & Bhat, 2004). They argued that model performance is quite sensitive to the sampling approach. Using universe alternatives can eliminate the drawbacks caused by the sampling process. Furthermore, it is feasible to run a large number of alternatives in the discrete choice model (Travel Forecasting Resource, 2019).

In addition, the pedestrian trip assignment was eliminated in MoPeD. Pedestrian trips are not typically assigned to a network because of the computational complexity of considering the various route options, particularly in dense, connected urban areas. However, the state of the research in this area is rapidly changing (Broach & Dill, 2015; Rodríguez et al., 2015), and with the addition of new data and primary research, future work of MoPeD may consider adding this stage.

Finally, MoPeD is also limited to its nature of aggregated modeling. Average social-demographic attributes are simulated at the PAZ level, which makes it less sensitive to demographic changes such as aging or car ownership in scenario application.

2.4.3. Microscopic Transportation Orchestrator (MITO)

Agent-based transport models have been well-developed in the past decades. The Munich Model is a model suite with three modules, including the synthetic population (Moreno & Moeckel, 2018), the travel demand model - MITO (Moeckel et al., 2020), and the transportation simulation - MATSim (Horni et al., 2016). The synthetic population provides a list of households and persons with socioeconomic and demographic attributes, workplaces, and school places, which are then fed into MITO. MITO is a microscopic travel demand model that generates trips for every individual, which are then passed on to MATSim for trip assignment. In this modeling suite, the multinomial logit mode choice model that is part of MITO estimates shares of walk trips but subsequently drops those trips from further analysis (as shown in Figure 4). Most agent/activity-based models fall within this framework (Singleton et al., 2018). At this point, it is impossible to analyze the impact of the built environment on walking or the health benefits for travelers choosing non-motorized modes. The Munich Model uses 4,953 gradually-sized zones as its spatial unit (Molloy & Moeckel, 2017), which were designed to capture vehicle trips rather than relatively short walking trips. These short trips in MITO were usually considered as same length intrazonal trips. The length of intrazonal trips is half of the average distance to the 3 nearest neighboring zones (Okrah, 2016).

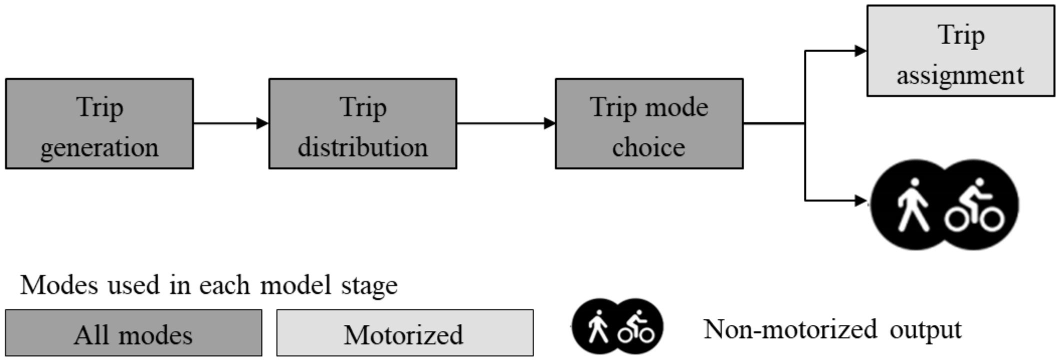


Figure 4 Pedestrian modeling framework used in the Munich Model (adapted from Singleton et al., 2018)

2.4.4. Other Travel Demand Models

Although the research on pedestrian travel behavior has lots of key findings, there is still insufficient work on developing pedestrian travel demand in transport models, especially at a regional level. The regional transport planning tools were originally oriented towards automobile travel. Non-motorized modes were first incorporated into regional transport models in the early 1990s (Liu et al., 2012). After that, many Metropolitan Planning Organizations (MPO) made progress in their regional models for pedestrian travel.

The DVRPC Regional Travel Demand Model is applied in the Philadelphia area (Delaware Valley Regional Planning Commission, 2011). In the DVRPC model, the non-motorized trip rates are directly modeled in the trip generation stage for three purposes. They are stored as zonal results and are dropped off in the distribution and assignment stage. The trip generation models for non-motorized trips include TAZ-level attributes such as the number of households, group quarters population, basic employment, and retail employment.

In the San Francisco County Model (SF-CHAMP), walk and bike trips are modeled separately in mode choice models (San Francisco County Transportation Authority, 2002). Pedestrian environment factors such as network continuity, ease of street crossing, perception of safety, and topological barriers are considered in mode choice. The route choice model is applied to cyclists but not to pedestrians. Trip purposes are only categorized into work, education, and others.

mobiTopp was developed and applied in the German context (Schnittger & Zumkeller, 2004). The model simulates activities over a week. Walk and bike tours/trips are modeled separately for eight personal purposes, including shopping and leisure. Mode choice is modeled after destination choice. There are no pedestrian environment factors including in mode choice models.

Portland Metro Model has enhanced non-motorized travel forecasting by incorporating walk and bike modes in the mode choice model (Portland METRO, 2020). Walk, and bike choices are made by these TAZ variables: number of local intersections, households, and total/retail employment. In addition, recreation trips are modeled, which comprise a significant number of pedestrian and bicycle trips.

A comprehensive report on urban models is conducted by Singleton et al. (2018). Among the 48 MPOs in the US, non-motorized trips are still excluded in 12 MPO models. The rest of the MPO models include walking and cycling separately or only as a combined non-motorized mode in mode choice. Although the shift to sustainable transport has created increasing interest in modeling pedestrian travel, there remain improvements that can better represent pedestrian behavior and evaluate health impacts more accurately. They conclude that the perspectives of the research on pedestrian modeling are 1) traditional pedestrian models with finer spatial resolution and 2) disaggregated models like agent-based and activity-based models.

2.4.5. Transport and Health Models

In previous studies, many researches have observed the health outcomes resulting from physical activity, including active transport, referring to walking and cycling (Evenson et al., 2012; Health Effects Institute (HEI), 2010; Maizlish et al., 2013). These studies have established the links between physical activity and health outcomes and exposure to pollutants and various respiratory diseases. A systematic review of health impact assessments (HIAs) from the U.S. included 21 HIA projects in the transportation sector (Rhodus et al., 2013). Most of them utilize the HIAs tools to examine the likely health impacts of transport-related policies and prioritize maximizing health impacts.

Walking-related indicators play important roles in the HIAs tools. One of the indicators widely used in the HIAs tools is the share of active transport mode. For instance, estimated walk and cycle share was used to calculate the number of lives saved in the Health Economics Assessment Tool for evaluating different road pricing policies (Rhodus et al., 2013). In addition, time spent/distance traveled in walking is an important indicator needed to calculate the total time of physical activity, the risk exposure to pedestrian-vehicular crashes and the contribution on air pollutant reduction (Lassarre et al., 2007; Maizlish et al., 2013; James Woodcock et al., 2013). The integrated Transport and Health Impact Modelling Tool (ITHIM) developed by James Woodcock is widely used in many countries for evaluating the health effects of different transport scenarios and policies (Centre for Diet and Activity Research, 2018). He implements the ITHIM to calculate the number of deaths by mean time spent in walking and cycling per week under three scenarios in the UK and India (Woodcock et al. 2009). They find that increasing active travel and lower-emission motor vehicles would give large benefits on health, especially for a reduction of ischemic heart disease. Besides walk share and time spend in walking, walkability/walk score is another common indicator for the health assessment at a local or country level. Interestingly, the CLARK county public health department found that the greater walkability can increase the access to healthy food, which can reduce the risk of obesity and many chronic diseases (Haggerty et al., 2011).

2.5. Research Gaps

This chapter presented state of the art on pedestrian studies in terms of data, behavioral research, and modeling approaches. In general, pedestrian research and practice have attracted more attention in transportation planning and have contributed to several improvements. However, there are still several issues that need to be taken up by the researchers to better understand and model pedestrian travel demand.

2.5.1. Pedestrian Travel Behavior Variability

Few studies have been conducted on travel behavior variability, particularly when accounting for pedestrian behavior.

One of the reasons is that the data for exploring travel behavior variability has historically been rare. Although a wide range of data sources exists for exploring pedestrian behavior, all these methods have strengths and limitations in assessing pedestrian behavior variability. HTS is costly and laborious. Due to the nature of self-reporting, HTS may still be biased in terms of the number of trips and trip distance. Also, conventional HTS data commonly provides only one day of travel and records the activity locations at a sparse scale/zonal level, which has poor coverage in terms of spatial and temporal resolution. Although panel surveys or longer-term HTS data can capture pedestrian travel behavior patterns across weekdays and weekends, they also face challenges in respondent fatigue and lack of fine-scaled location information. There are efforts made to combine HTS or panel surveys with GPS tracker technologies, but it is a relatively new approach with only a few examples. The quality of cell phone data highly depends on the density of cell phone towers, which varies by location. In general, there are few comprehensive datasets that can capture multi-temporal and spatial features simultaneously for understanding pedestrian travel patterns.

Another reason for the lack of travel behavior variability studies is that weekly or monthly (pedestrian) travel behavior is complex. It turns out that defining variability (and vice versa stability) is challenging. Furthermore, changes in travel behavior can be influenced by a variety of factors. Some of these factors can change on a daily basis, including activity participation, weather, and availability of modes. Some circumstances occasionally change during life, such as household relocation, car/bike purchases, and changes in employment status. Others could be exogenous changes brought about by the social and economic environment.

The lack of knowledge about pedestrian behavior variability can lead to some issues in model accuracy and model sensitivity. For example, those pedestrian planning tools mentioned in Section 2.4 simulated walk trips for a typical day. To calculate the physical activity volume for the health benefits assessment, the walk time/distance will be simply multiplied by seven under the assumption that people have the same active level on other days. Without capturing the walking behavior variability, pedestrian travel demand models tend to overmeasure people's physical activity volume (James Woodcock et al., 2009). Furthermore, in scenarios where employees work partially from home, their travel behavior will be related to the day of the week. For example, a day without commuting might lead to more active travel for non-commuting purposes. However, the existing pedestrian travel demand models are not sensitive to the partially teleworking scenarios since the temporal patterns are not considered.

Given the scarcity of longitudinal travel behavior data and the limitations of household travel surveys, this dissertation will carry out a survey to collect longitudinal and passive data – Google Location History (GLH) – that records an individual's trip diaries with coordinates and a board time horizon. Moreover, the collected GLH data will be employed to have a closer investigation of travel behavior variability. The survey design, empirical examination, and some analysis approaches of travel behavior variability are demonstrated in Chapter 8.

2.5.2. Built Environment Measurements

Built environment measurements are constructed differently across studies, which may be a source of discrepancy in results. As a result, there is a need to call for standardization of measuring built environment that is relevant to pedestrians.

In addition, it is important to account for how these measures may be used in various scenarios and forecasting applications. For example, the PIE measure developed for MoPeD combines six aspects of the pedestrian environment into an index scaled from 20-100. While the PIE can statistically avoid the high correlation between built environment variables, this index suffers from an inability to understand what sets of policies/actions can result in a 10-point increase in PIE.

Another issue of built environment measurements is that it is defined by the context of different study areas and is difficult to be transferred or applied to other regions. For example, it is difficult to collect built environment information to construct a walkability index or PIE for areas where data are scarce.

To improve the transferability of MoPeD, a new pedestrian accessibility measurement will be constructed to replace PIE and be used as a predictor in walk mode choice models. The details are presented in Section 4.2.3. The new measurement is easier to construct than PIE, more transferable to other study areas due to fewer data requirements, and easier to interpret in future scenarios.

2.5.3. Modeling Approaches

Although the research on pedestrian travel behavior has made great progress, there are only a few publications on pedestrian modeling practices. Most of these studies focus on simulating microscopic pedestrian movements in a specific situation, such as crowding and queuing at a single intersection or pedestrian evacuation at train stations and shopping centers (Borrmann, Kneidl, Köster, Ruzika, & Thiemann, 2012; Erdmann & Krajzewicz, 2015; Kielar & Borrmann, 2016). Only a few studies focus on pedestrian travel demand models at the urban scale. A comprehensive review of urban travel demand models in the U.S. pointed out that only over half of the transport planning tools account for walking or non-motorized mode shares (Singleton et al., 2018).

Some models, such as MoPeD, MITO, and SF-CHAMP, have improved the representation of pedestrians in travel demand models. While they have a wide range of benefits to pedestrian modeling and transport planning, many challenges remain in further development and implementation.

Pedestrian-centric planning tool MoPeD has a better representation of pedestrian behavior by employing finer spatial units and incorporating built environment variables. However, due to the heavy computational burden, model performance is limited. Model transferability is hampered by the complexity of built environment measurements. The nature of aggregation modeling limits model sensitivity to individual attributes, while the lack of representation of non-walk modes also limits model sensitivity to other modes.

The existing travel demand models, such as MITO and mobiTopp, have great opportunities to simulate pedestrian trips in an agent-based environment and along with other modes. However, most agent-based transport models either overlook pedestrian activities or have limited representation of pedestrian behavior. Walk mode is usually skipped or combined with cycling as one non-motorized mode (Waddell, 2002). When it comes to measuring physical activity energy expenditure, walking and cycling have significant differences. As a result, combining walking and cycling as the non-motorized mode will cause some issues when evaluating the health benefits of transport-related physical activity.

Moreover, the majority of the transport models are applied with coarse zone systems (e.g., block group, TAZ). This is sufficient for understanding car demand on transport infrastructure. However, the walk trips are usually too short to be neglected or be considered as intrazonal trips in the car-oriented transport models. Intrazonal trips are always difficult to measure. The great number of intrazonal trips will also reduce model accuracy and sensitivity because intrazonal trips are usually assigned to the same length. To simulate pedestrian demand in an agent-based transport model, it is important to apply a fine-grained spatial resolution to better capture shorter trips as well as attributes of the built environment that influence walking.

Another limitation of agent-based transport models is the poor understanding of important factors that influence walking. Previous studies have shown a significant influence of built environment factors on walking behavior. This knowledge has not been incorporated into agent-based transport models in reliable, predictive methods for use in planning and policy decisions.

The author concludes that the pedestrian-centric planning tool MoPeD and agent-based transport models have their respective strengths and limitations, which can complement each other. However, the incorporation of pedestrian-centric modeling with an agent-based transport model in an urban region has not been attempted.

To fill this gap, MoPeD and MITO are chosen as the two starting points of model development. To begin, efforts will be made to enhance the stand-alone MoPeD model. Its limitations, as mentioned in Section 2.4.2.3, will be addressed one by one in Chapter 3. Once MoPeD is improved, it will be incorporated into MITO. An integrated modeling framework will be developed in Chapter 6, resulting in a more accurate, sensitive, transferable, and efficient pedestrian planning tool.

3. Study Areas

This chapter introduces the study areas that will be employed in the subsequent chapters. MoPeD 2.0 described in Chapter 4 is developed in the context of the Portland metropolitan area. Portland Central City area will be used in scenario applications in Chapter 5. Munich metropolitan area will be used in Chapters 6 and 7 for the development of the integrated model and its scenario applications.

3.1. Portland Metropolitan Area

The Portland metropolitan area has more than two million population. It has made large investments in public transportation and in pedestrian public spaces that are experienced every day.

MoPeD 2.0 is developed in the context of the Portland metropolitan area, which is delimited by the urban growth boundary defined by Portland Metro Regional Governance (see Figure 5). The total population of the three counties of this region is over 1.6 million, and the area covers about 1,048 square kilometers.

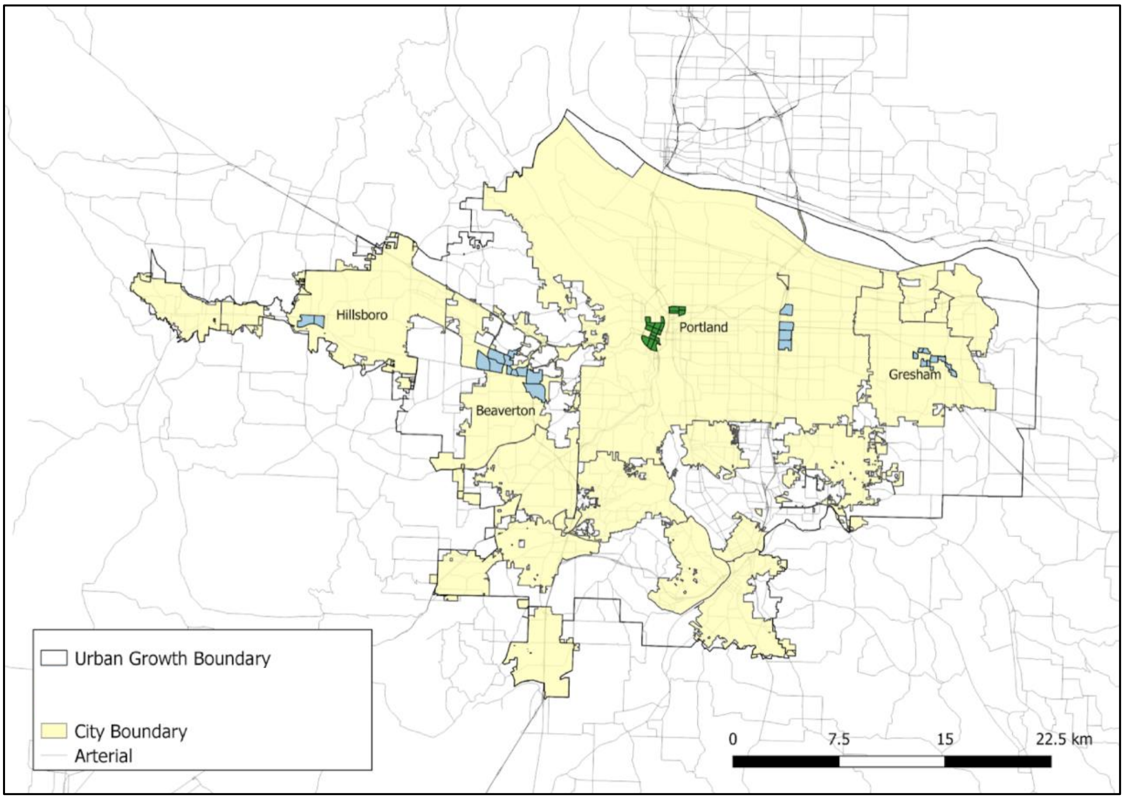


Figure 5 The urban growth boundary of the Portland metropolitan area

3.2. Portland Central City

Portland, OR, as the heart of the metropolitan area, has the densest concentration of people and jobs. With an increase in urbanization, Portland city will continue to experience population and employment growth. The city projects it is going to gain approximately 38,000 new households and about 51,000 new jobs by 2035 (City of Portland, 2018).

The study area for scenario application in Chapter 5 is Portland Central City, shown in Figure 6. The Central City area consists of ten different neighborhoods and stretches from the West Hills to East 12th Avenue and from the Pearl and Lower Albina to the South Waterfront area and Powell Boulevard (City of Portland, 2018). Although Central City only covers about 12 square kilometers in land area, it accounts for almost 20% of the total population in the metropolitan region. It is the densest area of people and jobs in Oregon. The Willamette River divides this area and is spanned by several bridges, including the non-automobile bridge Tilikum Crossing, completed in 2015.

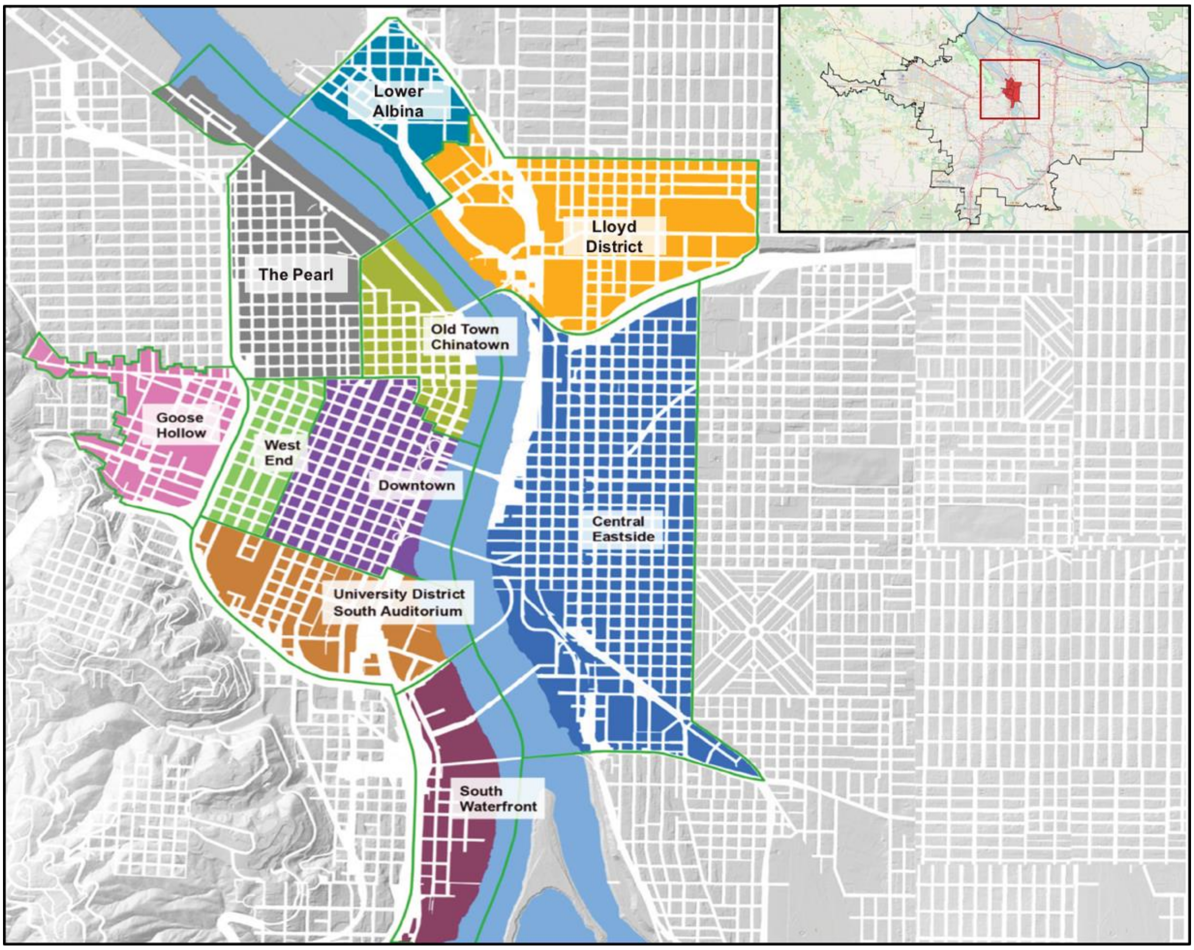


Figure 6 A map of the Portland Metropolitan area and the Portland Central City City of Portland, 2018)

3.3. Munich Metropolitan Area

The MITO/MoPeD integration model is developed in the context of the Munich metropolitan region. Munich is the capital of Bavaria state in Germany. It is also the third-largest city, with a population of 1.58 million (Landeshauptstadt Muenchen, 2022). The study area is delineated based on commute flows to five core cities: Munich, Augsburg, Ingolstadt, Landshut, and Rosenheim (see Figure 7). The study area includes all municipalities from where 25% or more of the working population commutes to one of the five core cities. This 25% threshold was chosen to include all major commute flows while keeping the study area size computationally feasible.

The synthetic population consists of geo-referenced microscopic households, persons, jobs, and dwellings, which were generated using an iterative proportional updating approach (Moreno & Moeckel, 2018). The synthetic population consists of 2.1 million households with 4.4 million persons (of which 1.7 million are workers) and 2.5 million jobs.

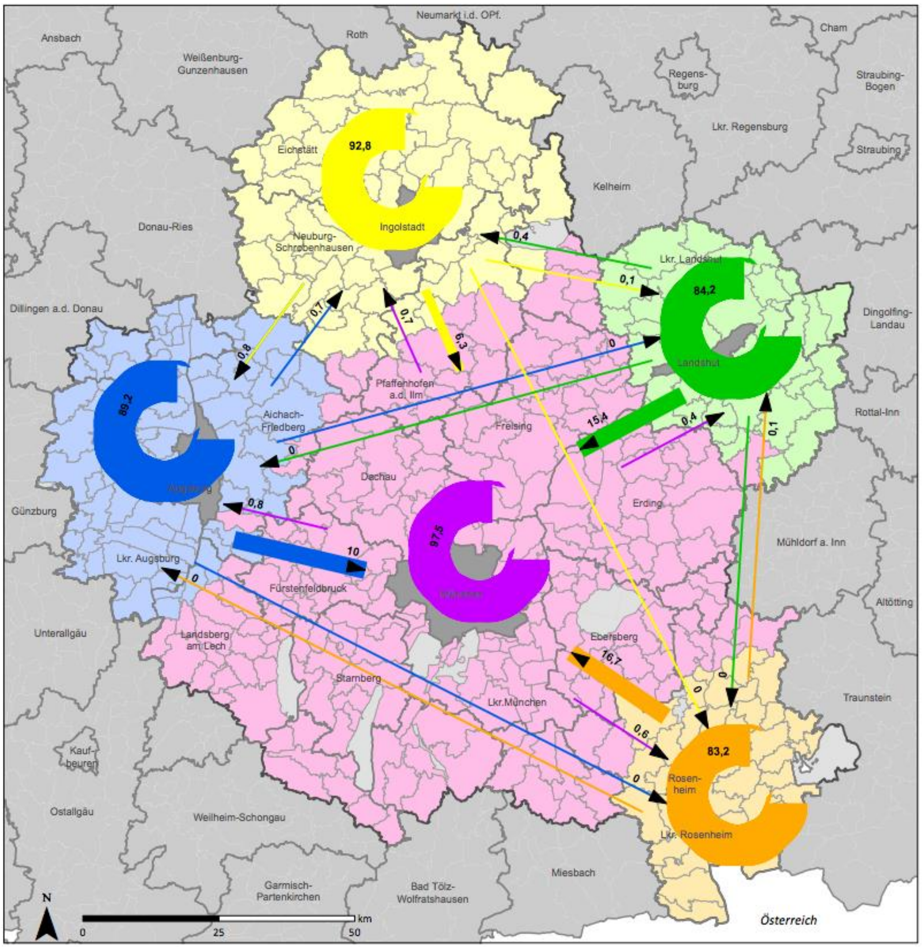


Figure 7 Extent of the study area and major commute flows between core cities

A zoning system of 4,096 zones was developed with a quad-tree-based automated zone system generator that iteratively creates smaller raster cells in densely populated areas and larger raster cells in rural areas while respecting administrative boundaries (Molloy & Moeckel, 2017). This leads to zones with similar population sizes but different sizes (see Figure 8 and Figure 9).

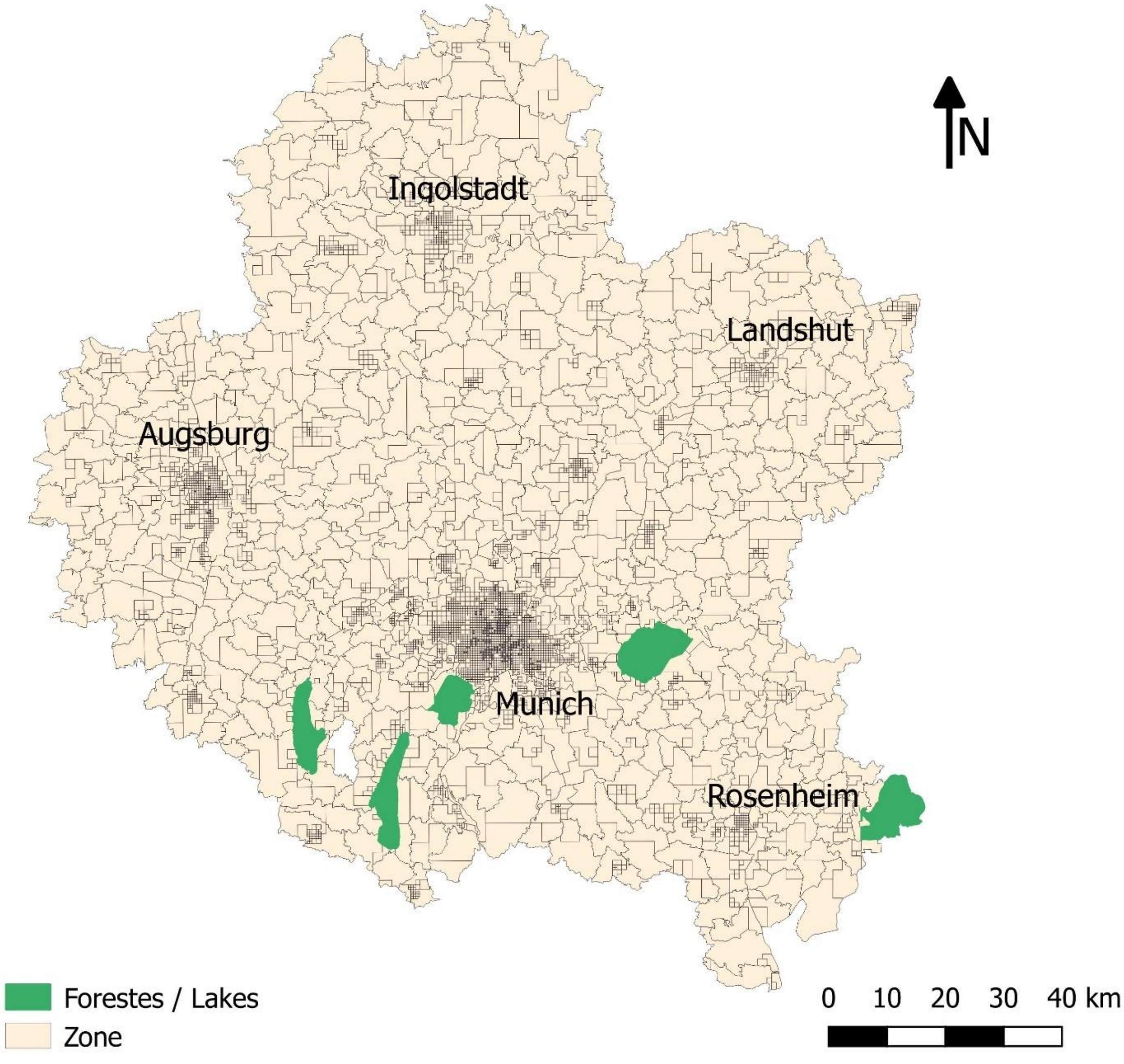


Figure 8 TAZ Zone system of the Munich Metropolitan Area

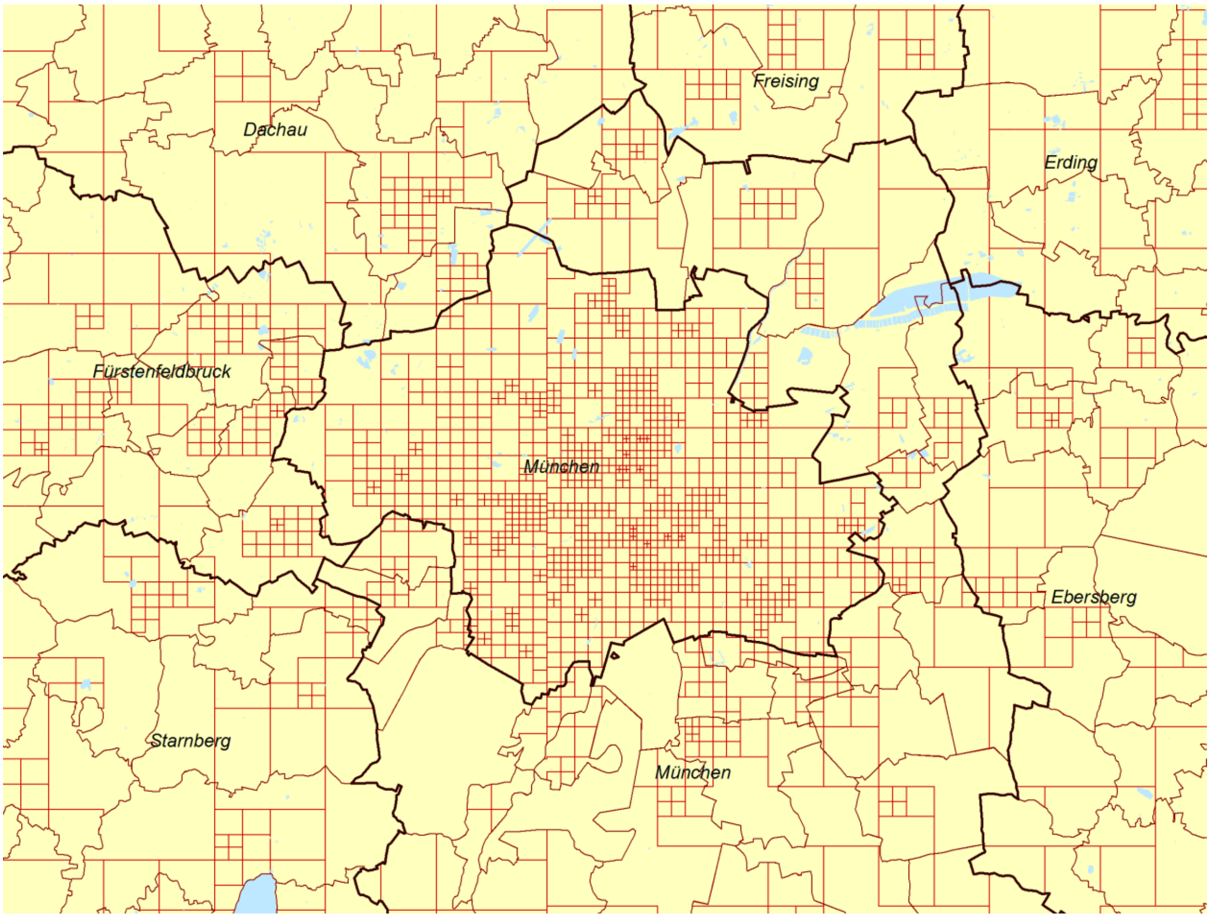


Figure 9 A closer look at the TAZ zone system of the Munich metropolitan area

4. Advancing Pedestrian Demand Modeling: Towards MoPeD 2.0

In this chapter, the existing pedestrian modeling tool – Model of Pedestrian Travel (MoPeD) – will be enhanced. The aim is to improve the stand-alone pedestrian planning tool MoPeD to be more efficient, more sensitive to the built environment, more powerful in predicting and explaining future scenarios as well as easier to be transferred to other areas/applications. The limitations of the MoPeD 1.0 discussed in 2.4.2.3 are the barriers to achieving this goal. In the following sections, the author presents the approaches for addressing these issues.

In Section 4.1, the appropriate spatial resolution for measuring pedestrian environment and modeling behavior is explored at the scale of a metropolitan area. Once the appropriate spatial scale is determined, Section 4.2 presents some key improvements the author made in MoPeD 2.0. First, the run time and memory saving of the model are upgraded (Section 4.2.1). Then in Section 4.2.2, the author implemented trip generation models based on the latest version developed by the Oregon Metropolitan Planning Organization (MPO). Furthermore, the new pedestrian accessibility measures are developed and integrated into the walk mode choice models as a strong predictor (Section 4.2.3 and 4.2.4). More important, walk destination choice models are re-estimated by using the whole universe of choices and a two-step multinomial logit model technique (Section 4.2.5).

4.1. Spatial Resolution Investigation

Along with the rich literature on influencing factors on pedestrian travel behavior, many scholars also emphasis the appropriate spatial scales for measuring and modeling pedestrian travel (Gehrke and Clifton, 2014). It is important for models to set a finer spatial resolution to capture variations in walking conditions, leading to a better representation of pedestrian demand over space (Gehrke & Clifton, 2014).

However, it might be impossible to define a singular solution of the “optimal” spatial resolution that fits various types of applications, including but not limited to transportation investments and land use policy. Each of these uses has different requirements with implications for spatial resolution. For example, regional land use scenarios may require a scale that can detect modal responses to the changes in local and regional accessibility. There is increasing interest in conducting health impact assessments for planned transportation investments with an emphasis on safety, air quality exposure, and physical activity. The latter requires some estimation of total minutes spent in physical activity from active transport modes, which could be approximated from the trip distance. The error introduced by increasing the spatial resolution of the model could have an impact on its ability to inform these various studies. As a result, the appropriate spatial resolution highly depends on the research questions that the model aims to address, the run times that the model shall not exceed, and the availability of spatial data needed to implement. Therefore, this work is to try to balance these considerations and understand the various trade-offs involved by testing various spatial scales and providing recommendations of the appropriate resolution for a specific scenario. The results presented here have been published in Zhang, Clifton, & Moeckel (2019)

Eight different scales are examined (Figure 10). They are 80 m to 400 m in increments of 80 m, as well as 800 m, 1,200 m, and 2,400 m. The current size of PAZ (80 m × 80 m) is considered the reference scale since it is the minimum grid cell size. The grid cell sizes smaller than 80 m are not considered in this study. This is because the archived land use and household travel data are available in the 80 m level of detail. The data disaggregation spatially needs an additional process. Besides, an 80 m grid cell was hypothesized to be small enough to capture fine-grained attributes of households and the physical environment, as well as variation within those attributes, in order to accurately represent walking (K. J. Clifton et al., 2013).

The Portland study area is divided into 54,329 raster cells having a dimension of 80 m × 80 m. It is a challenging amount for the computational burden. The varying test scales decrease the number of grid cells from 54,329 to 89. Table 1 lists the number of raster cells for each test PAZ scale. It shows an exponential decrease trend in the number of raster cells when applying coarser spatial resolution.

Regards to the impedance of intrazonal trips, there are several approaches to calculating the intrazonal distance appropriately. In this study, we define the intrazonal distances as half of the cell sizes, which also represent half of the distances to the adjacent zones. Therefore, the intrazonal impedance varies in different test scales. Table 1 gives the overall impedance of each PAZ scale.

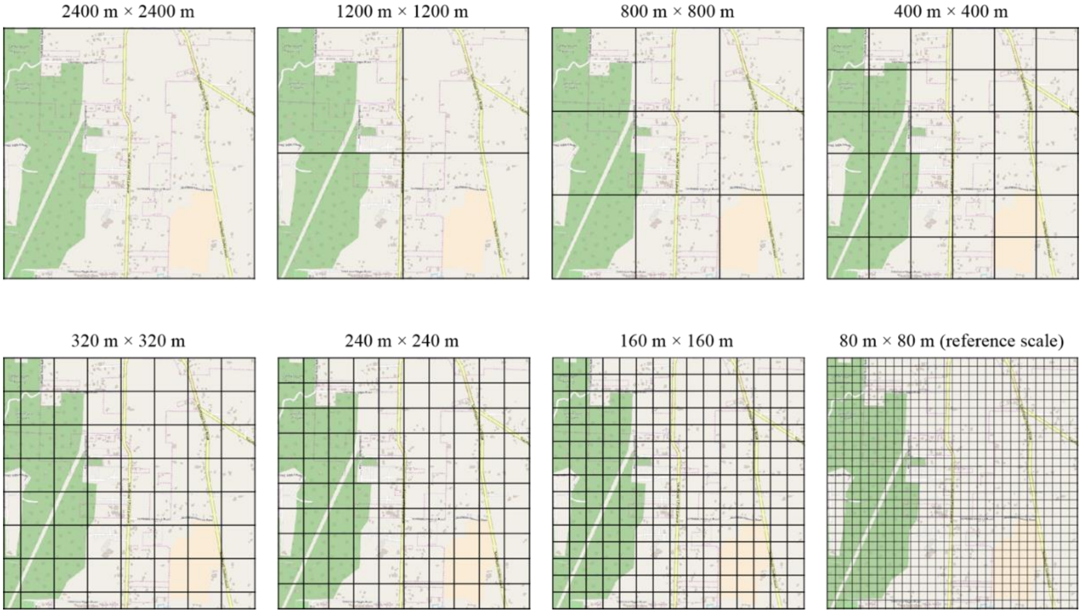


Figure 10 Eight test scales

Table 1 Number of raster cells and intrazonal impedance of each test scales

Cell Size	Number of Raster Cells	Intrazonal Impedance
80 m × 80 m	54329	40 m
160 m × 160 m	14057	80 m
240 m × 240 m	6433	120 m
320 m × 320 m	3722	160 m
400 m × 400 m	2430	200 m
800 m × 800 m	649	400 m
1200 m × 1200 m	311	600 m
2400 m × 2400 m	89	1200 m

After defining the test PAZ scales, we prepare the input data corresponding to the test scales. The mode choice and destination choice model used in MoPeD measures relationships between walking and built environment, traveler characteristics, and socioeconomic variables. Therefore, basic inputs to MoPeD include the number of households with different household characteristics, the number of jobs by employment type, and various measures of the built environment, such as the pedestrian index of the environment variable called PIE (Singleton et al., 2014), slope and freeway. Most of these data are available in 80 m × 80 m level of details.

The input data of a larger PAZ scale is an aggregation of the original data of each reference scale. The aggregation methods are different according to the data type. For count data (households, jobs, and shopping center areas), aggregated data is a sum of each reference PAZ data. For PIE and length of the freeway, we assume that the data in larger scales are the mean of each reference PAZ within them. For binary data (freeway and parks), provided there is freeway/parks in the correlated reference PAZ, the data in larger scales is set to one. Different from the other input data, the degree of slope is only available in the level of details of 400 m × 400 m. To disaggregate the value to the reference scale, we assume that each reference PAZ within the same 400m × 400 m grid cell has the same value of the slope.

When implementing different zoning systems, the impedance matrices need to be recalculated between new O-D pairs. The centroid coordinates represent the location of each raster cell. Network data is obtained from OpenStreetMap Contributors (OpenStreetMap Contributors, 2017), which skims off the motorway and includes both bicycle and pedestrian paths. MATSim (Horni et al., 2016) is used for calculating the impedance matrices. With an increase in the number of raster cells, calculating time and computer memory increase dramatically. It is most challenging to calculate the reference zone system, which spent over four hours to generate the impedance matric.

4.1.1. Impacts on Model Efficiency

Undoubtedly, model complexity is inversely proportional to model efficiency. In this study, model complexity refers to the reduction of the cell size as well as the increase of the number of raster cells in the model. Figure 11 shows how model complexity impacts the run time (seconds) and the maximum memory usage (Gigabytes). Both indicators are generated after implementing the model 20 times.

As seen in Figure 11 (a), the run times decrease as the spatial resolution becomes coarser. The drop of run times is extremely steep from 80 m to 240 m scale and becomes relatively gentle between 240 m and 800 m scale. Beyond 800 m scales, it reaches a stable level. As varying the cell size, the run times improve from four and half minutes to 0.1 seconds. Although the run times with 80 m cell size are already very promising, Figure 11 (a) seeks to highlight the dramatic change between 80 m and 160 m. Maximum memory usage (Figure 11(b)) shows an approximate exponential growth as the increase of number of raster cells. It is worth mentioning that using the finest scales (80 m) occupies more than 40 GB, which already accounts for two-third memory of the experimental computer. The great amount of memory is caused by the large size of the impedance matrix that depends on the number of OD pairs. Therefore, it would imply that memory exceed would be happened soon when we increase the number of raster cells to more than 60,000.

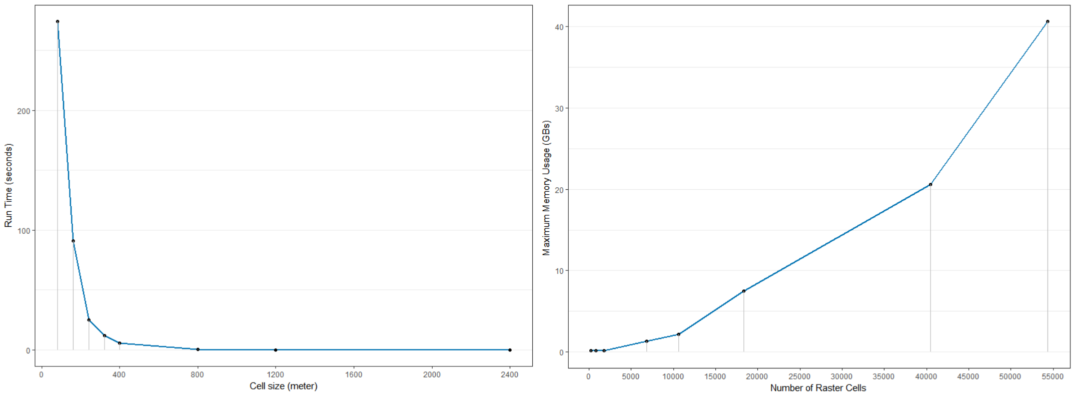


Figure 11 (a) Run time in seconds by varying cell sizes; (b) Maximum memory usage in gigabytes by a varying number of raster cells

4.1.2. Impacts on Model Performance Indicators

The errors of walk share and average trip length are computed as percentage root mean square error (%RMSE) between the predicted data of each test scale and the results from the reference scale (80 m x 80 m). Average trip length is the total trip length divided by the total walk trips, including both interzonal and intrazonal trips. To have a better understanding of the errors of both performance indicators, Figure 12 illustrates the normalized %RMSE by difference scales, while the data labels are the %RMSE in correlation with the test scales.

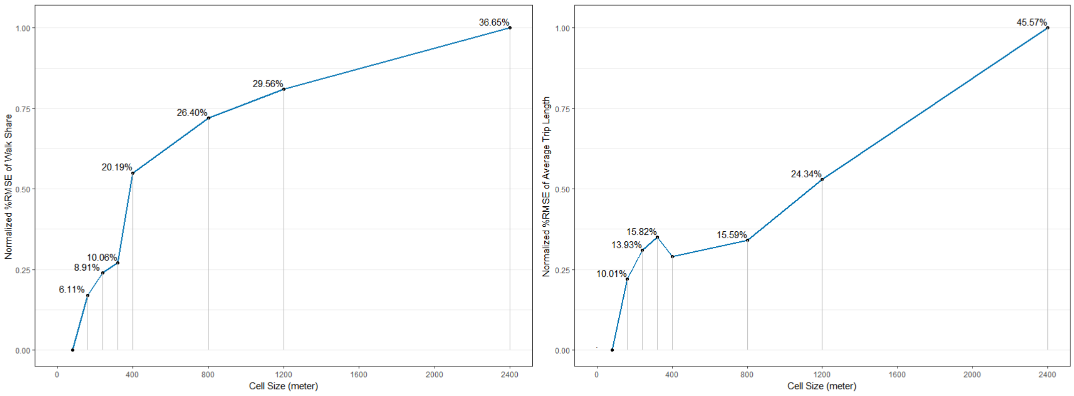


Figure 12 (a) Normalized %RMSE of walk share; (b) Normalized %RMSE of average trip length.

Figure 12 (a) figures out how spatial resolution impacts the deviation of walk share compared to the reference data. We can find that the error of walk share increases sharply when using 400 m, but they maintain the error under the level of 10% when using cell size under 320 m. As seen in Figure 12 (b), although the deviations of average trip length show a reduction at the point of 320 m, they generally illustrate a rising trend with the varying cell sizes. Compared to walk share, the varying spatial resolutions have larger impacts on the average trip length. Almost all the test scales generated more than 10% RMSE.

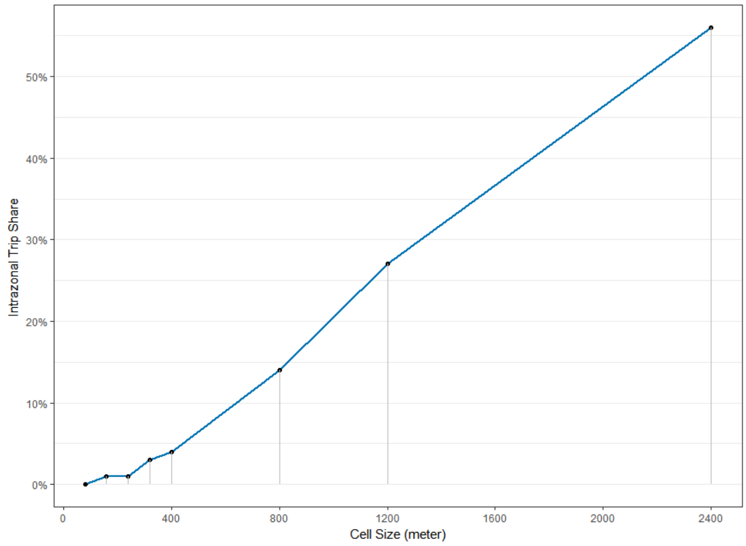


Figure 13 Intrazonal trip share

For a zonal-based travel demand model, intrazonal trips cannot be ignored and are always difficult to measure. The great number of intrazonal trips will to some extent, reduce the accuracy of the model estimation. Walking is a relatively short distance trip, so intrazonal walking trips are much more sensitive to spatial resolution. As expected, the intrazonal trip share increase approximately linearly as the spatial resolution gets coarser (Figure 13). The steepest increase occurs between 400 m and 1,200 m. It is unusual to use uniform 2,400 m × 2,400m raster cells in the pedestrian demand model, but it hints at the great number of intrazonal trips when modeling walking trips in the regional demand model with TAZ or block groups. With a threshold of 5%, this research suggests that the zone systems with raster cells equal to or under 400 m are supposed to be good for avoiding intrazonal trips.

To sum up, it is proved in this section that the use of fine levels of spatial resolutions generally results in improved representation of walk share and average trip length and a reduction of intrazonal trips. However, this comes with an exponential increase in run time and memory usage.

4.1.3. Impacts on Scenario Applications

Previous sections give us general ideas for how spatial resolutions impact the model performance and efficiency, but it cannot conclude an optimal spatial resolution at all since the spatial resolution depends on the type of application. In this section, we implement these scales into a local land use development scenario. This is a sensitivity analysis to explore the impact of spatial resolution on scenario results. In terms of the redevelopment in Portland city, a large parking center in downtown is supposed to be built as a mixed residential and commercial area. The assumed land use scenario allocates 500 new households with a population of about 800 and 1000 new jobs to the downtown.

Regarding the input data, household and job distribution are changed in the scenario case. Meanwhile, PIE is recalculated for the entire study area in correlation with the scenario assumption since activity density is one of the measures for estimating PIE.

The deviation of walk shares and average trip length is computed as percentage root mean square deviation (%RMSD) between the scenario case and the base case of each test scale. Figure 14 illustrates the normalized %RMSD of both indicators by different scales. It tells us the sensitivity of the spatial resolution response to the scenario changes.

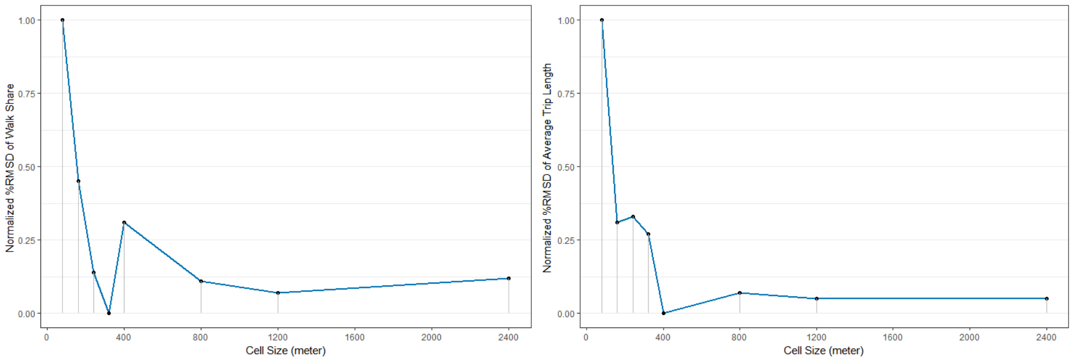


Figure 14 (a) Normalized %RMSD of walk share; (b) Normalized %RMSD of average trip length.

Since the scenario assumption is allocated in a high activity density area, it is expected to be more sensitive to spatial resolution. Figure 14 proves that finer scales are more sensitive to detect the changes between the scenario case and base case. Moreover, it is remarkable that the impacts of the land use scenario on walking behavior cannot be represented well when using coarser spatial resolutions that are equal to or larger than 800 m. In terms of the spatial resolution under 800 m, the sensitivity to the scenario change differs from these two indicators. Undoubtedly, the 80 m scale is the best to represent the changes between the scenario and base case. However, the finer scales reduce the model efficiency, and increase the difficulties in calculating impedance matrices and assembling input data, so it is worthwhile to limit the cell size. Instead of the finest spatial scale (80 m), 160 m is the second appropriate spatial resolution for estimating the scenario changes in walk share, while 160 m to 320 m could all be alternative resolutions for predicting the change of average trip length.

4.1.4. Summary

This research provided an exploration into pedestrian demand model performance and efficiency at various geographic scales. Different from previous studies on examining the pedestrian behavioral response to various spatial resolutions, such as gradual raster cells (Okrah et al., 2017) and uniform raster cells larger than 400 m (Gehrke and Clifton, 2014), this study focused on exploring the uniform raster cells ranging from 80 m to 2400 m. It fills the research gap and provides recommendations for the appropriate resolution, which is ever more important as the development of pedestrian models is on the rise (Singleton et al. forthcoming).

The results presented in this section proved that a finer spatial resolution is beneficial for model accuracy in estimating walking-related indicators and sensitivity to the scenario analysis. Coarser scales that are larger than 800 m generally cause high errors in predicting walk share and average trip length and lose a great amount of information due to the high share of intrazonal trips. However, the finer scales reduce the model efficiency, increase the difficulties in calculating impedance matrices and have a heavier computational burden in memory usage. Particularly, using the finest spatial resolution (80 m) is the risk of the memory exceeding issue. Besides, it would meet the challenge of hardware limitations when expanding the model functionality. Therefore, the findings pointed out that the finest spatial resolution may not be the appropriate scale, and it is reasonable to enlarge the cell sizes up to 400 m, which produces an acceptable error as well as higher efficiency and maintain the intrazonal trip share under a level of 5%. These findings also give the locations where the land use data are not available at a fine-grained level a better understanding of choosing alternative spatial resolutions when developing pedestrian travel demand models.

The debate about the appropriate spatial resolution will continue since it highly depends on the type of application that the model aims to address and the level of details of the available data differing from locations. In this study, a land use scenario was applied at various scales. The scenario assumption is a block-level land use development in the downtown area. 80 m scale is the best to represent the changes between the scenario and base case. Nevertheless, fine scales of 160 m and 240 m are relatively more sensitive than the other scales, which can be considered as the second appropriate spatial resolution.

To have a better understanding of defining an appropriate spatial resolution, more scenario analysis should be conducted in future work. Besides the land use policy, they should cover other various types of applications such as transportation investments and safety and health assessment. In addition, the areas with different activity densities and pedestrian environments will have diverse requirements for spatial resolution. In this study, the results of the urban area show a high sensitivity to the spatial resolution, while the results may have little difference between fine-grained cells and coarse grid cells in a suburban area. As a result, scenario analysis in suburban areas will be conducted in future work.

4.2. Enhancements in MoPeD 2.0

We enhanced the model performance and made some modifications to the model stages from the version previously described by Clifton et al. (2015, 2016a, 2016b, 2018). These improvements were made to overcome the limitations mentioned in Section 2.4.2.3.

4.2.1. Upgrade of Model Run Performance

MoPeD 1.0 was developed in R, which is easy to construct and accessible with limited knowledge of programming. However, MoPeD 1.0 was facing the computational burden of running large-scale study areas with fine spatial resolution. The computational burden highly depends on the number of PAZs in the study area, which is discussed in Section 4.1.1. It was only managed to test scenarios for a subset of the Portland central city area

To address the model efficiency issue, the author migrates the full model from R to a Java project. Benefits from the nature of Java as a high-level language, this change made the model efficient and operational for the entire Portland region with a runtime of a few minutes (Presented in Section 4.1.1).

4.2.2. Extension of Trip Generation Models

For trip generation models, we adopted Metro's existing trip production models, which use cross-classification to calculate trip production for all home-based purposes (Vogt et al., 2015). While Metro designed a model for TAZ-level inputs, we assumed model scalability and applied the trip rates at the PAZ level. The detailed methodology and coefficients of the trip generation model can be found in the report Vogt et al. (2015). Home-based work (HBW), home-based education (HBE), home-based shopping (HBS), home-based recreational (HBR), home-based other (HBO), non-home-based work (NHBW) and non-home-based other (NHBO) purposes are modeled in MoPeD 2.0. Take HBW trips as an example, they are produced based on the number of workers in a household. For each zone, the number of households by the number of workers (0 workers, one worker, two workers, and 3+ workers) is multiplied by the corresponding production rate. Afterward, the number of trips is scaled up to match the total regional number of jobs by applying a calibration factor of 1.36 for HBW trips. For the entire Portland Metro area, the model generates a total of 1,039,872 HBW trips.

4.2.3. Reconstruction of Pedestrian Environment Index

The Pedestrian Index of the Environment, called PIE, was less transferable to other applications due to the requirement of detailed land use data at a fine spatial resolution (K. J. Clifton et al., 2019). Also, it was challenging to predict changes to the built environment due to its construction. Furthermore, PIE limits the model's transferability to other regions due to its relative scale. For example, an area in Portland with a PIE scale of 100 represents the best conditions for walking in that region. However, for the less walkable cities, the best areas are not comparable with those in Portland. With this study, the author contributes a more transferable measurement of pedestrian activity that is more usable for policy analysis.

The new pedestrian accessibility that is used in mode choice is defined as activity density (employment and population) that can be reached within an 800-meters network distance. Often, accessibility is calculated based on a circular buffer (Greenwald & Boarnet, 2001; Lee & Moudon, 2006). However, it neglects that people do not make travel decisions in a circular buffer, but rather following street grids can be imperfect and lack connectivity. The network distance to reach the outer rim of the buffer area is likely to be longer than the radius of the circle. To overcome the limitation, we use network-distance-based isochrones instead of circular buffers. The isochrones are also known as the pedestrian catchment area (PCA) and represent a buffer defined by the pedestrian network distance instead of a fixed radius. For this task, we generated the PCA for each PAZ in the sample using a fixed network distance of 800 meters. According to Oregon Household Activity Survey (OHAS) data, an 800-meter distance covers 80% of the walk trip distances, and an 800-meter distance is equivalent to about 10 minutes' walk, which is sufficient to represent the pedestrian accessibility in the neighborhood. The method is based on street segments to generate the buffer polygon. For this task, every block that is enclosed by street segments becomes part of the polygon. If the street segment does not enclose an area, a buffer of 25 meters of the street segment is added to the polygon. We completed this task for the urban growth boundary of Portland which consists of about 160,000 PAZ. Afterward, we calculated the total number of jobs by type and population that live within each polygon, resulting in the measurement of pedestrian accessibility within 800 meters for the mode choice model. The built environment data used in this study for Portland, Oregon, was provided by the Portland Metro MPO. The scale of the data is the PAZ level, which consists of a grid of 80 by 80 meters. For each zone, the population and the number of employees by eight job types were provided. The results of pedestrian accessibility for the Portland metropolitan area are shown in Figure 15.

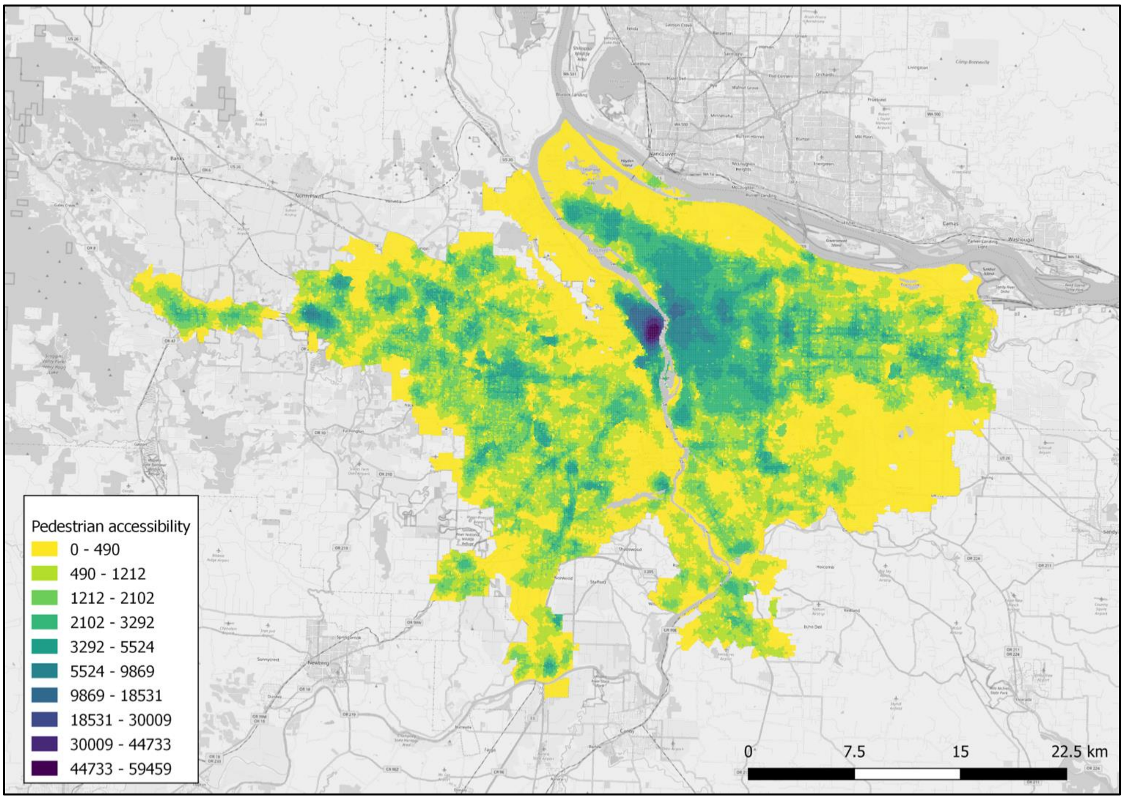


Figure 15 Map of the pedestrian accessibility

4.2.4. Estimation of Walk Mode Choice Models

The model choice models in MoPeD are estimated with data from the Oregon Household Activity Survey (OHAS). OHAS is a one-day household travel survey collected for the entire state, including the Portland metropolitan region. The survey was conducted in the fall of 2011 and included 6,450 households in the Portland Metropolitan area. Personal and household characteristics and their travel behavior were collected.

MoPeD employs a binary logit model to estimate the probability of choosing to walk. Note that in contrast to traditional four-step models, the MoPeD mode choice model does not know the destination yet. Since MoPeD is an aggregate model, the utility is computed for each household type located in every PAZ. Previously, the model for HBW included five household attributes (household size, income, age of household head, number of vehicles, and children) and PIE as independent variables. In this study, we replaced PIE with the pedestrian accessibility measurement, which does represent not only the activity density but also the network connectivity between PAZs. In addition, the model includes household car ownership and a dummy variable for having children. After removing the non-significant variables, the final model estimation results are shown in Table 2. Some variables with a significance slightly below 90% were also highlighted when theory supports their inclusion

Table 2 Binary logit mode choice model estimation in MoPeD 2.0

	Home-based purposes			Non-home-based purposes		
	Estimate	Pr(>|z|)		Estimate	Pr(>|z|)	
(intercept)	-8.392	0.000	***	-7.411	0.000	***
Income category 2				-0.205	0.261	
Income category 3				0.222	0.046	*
Income category 4			.	0.448	0.000	***
Number of vehicles (0)	1.001	0.000	***	1.375	0.000	***
Number of vehicles (2)	-0.226	0.002	**	-0.898	0.000	***
Number of vehicles (2+)	-0.394	0.000	***	-0.963	0.000	***
Number of children (0)	-0.554	0.000	***			
Number of children (2)	-0.574	0.000	***			
Number of children (2+)	-0.718	0.000	***			
Child (Yes)			***	-0.162	0.039	*
log(pedestrian accessibility)	0.754	0.000	***	0.686	0.000	***
HBS	1.029	0.000	***			
HBO	1.046	0.000	***			
HBR	1.566	0.000	***			
NHBW				-0.362	0.000	***
Log-Likelihood:	-4189			-2624		
McFadden R^2:	0.135			0.228		
Accuracy	10.14%			11.96%		

Household characteristics had significant effects in the model. The number of vehicles was the most significant predictor of walking. Zero-car households had a more positive association with walking than households that owned cars. Households with two or more vehicles had increasingly negative coefficients, confirming that household with more vehicle is less likely to walk. Household income had no significant impact in the model. It also highly correlated with the number of vehicles, and we removed it from the final model. An interesting effect was observed for the dummy variable of having children. In the model, living in a household with children had a positive impact on walk mode choice for work trips. Household size was removed from the model due to its high correlation with the number of children.

Pedestrian accessibility was transformed into a log-form to improve the model fit. It shows a positive impact in the model, which indicates that households living in denser neighborhoods with better street networks tend to be more likely to walk. The log-transformation suggests that differences in pedestrian accessibility matter a lot at the lower end of accessibility. Once a certain level of pedestrian accessibility has been reached, additional growth in accessibility has less impact on the likelihood of walking. This suggests that pedestrians need a certain level of accessibility. Once this level is satisfied, commuters are much more likely to walk to work. The statistical significance also indicates that the new measurement of the built environment is a good indicator of walking activity while controlling for all other variables. The R^2 appears low in comparison to traditional mode choice models. However, traditional mode choice models appear after destination choice, and thereby they can model the walk share based on travel distance. In this case, the walk share is modeled first to skim off walk trips from the total trip production. As the travel distance is not known yet, the challenge for MoPeD's mode choice model is substantially higher. After implementing the mode choice model to the entire study area, the estimated walk share in this region is 3.85% (observed walk share in OHAS: 3.50%). Figure 16 presents how the walk share is spatially distributed in the Portland metropolitan area.

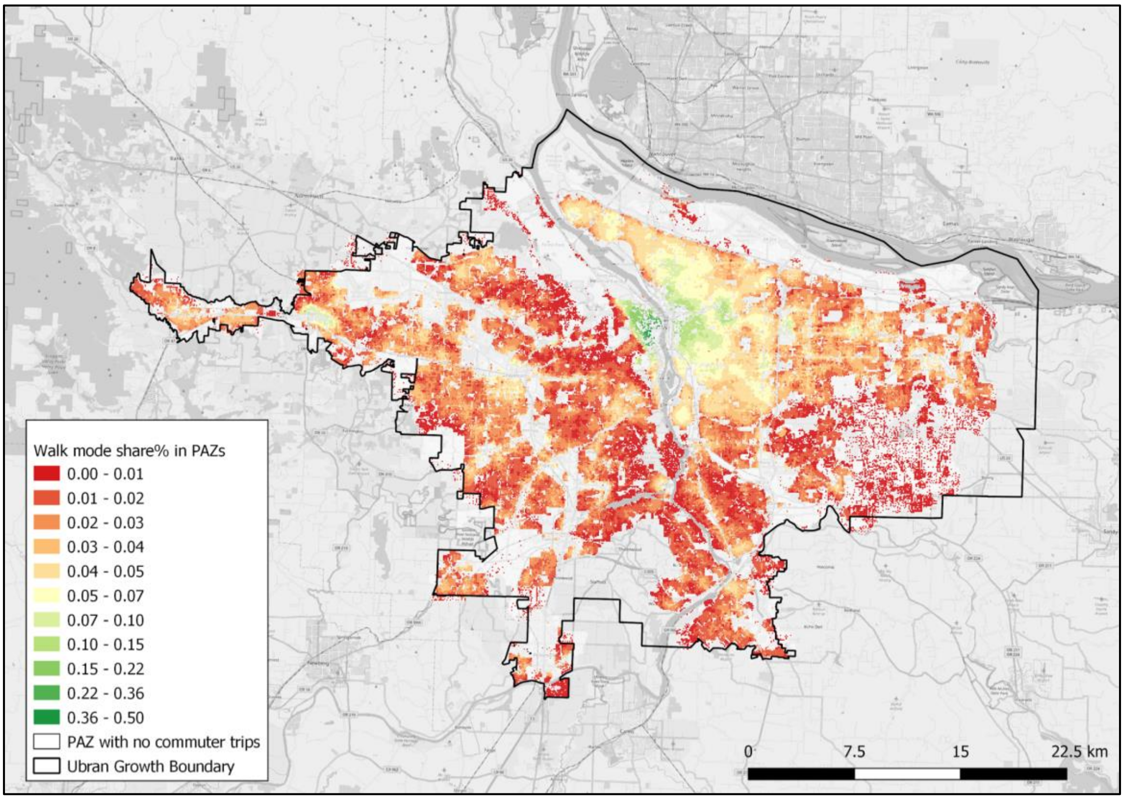


Figure 16 Estimated walk mode share of commute trips in the Portland metropolitan area

4.2.5. Estimation of Walk Destination Choice Models

For the estimation of the choice model, the definition of the choice set is a major challenge. There is little literature regarding this topic for pedestrian models, and there is no conclusive evidence of the preferable approach. The debate continues in the literature between different sampling methods or to include the whole universe of choices. Clifton et al. (K. J. Clifton et al., 2016a) used a random sample of 10 trips that were shorter than the distance of the 99% percentile in the observed data, Berjisian and Habibian (Berjisian & Habibian, 2019) used a 90% percentile threshold with a complete sample at a different spatial scale. We took the percentile 99% of the network distance that was 4.8 kilometers, and we used the whole choice set to estimate the destination choice model.

Using PAZ, the choice set would include an average of 10,000 possible destinations. Such a large choice set violates the assumption of discrete choice models that the number of choices should be small enough to be comparable against each other. Therefore, we aggregated the PAZ structure to a new scale that we called superPAZ. A superPAZ consists of 5 by 5 PAZ. For the purpose of estimation, we calculated the network distance shortest path between all superPAZ. Next, we assigned origin and destination superPAZ to each observation in the

OHAS survey data for model estimation. In terms of superPAZ, the possible choice set for each trip is about 576 (all superPAZ within a 4.8-kilometer buffer). By aggregating PAZ to super-PAZ, the issue of very large samples was compensated to some degree.

The specification (see equation below) was inspired by the previous MoPeD model yet simplified slightly to improve model sensitivities. Newly introduced were a log transformation of total employment, the proportion of industrial jobs, and the distance to be controlled by car ownership. We also introduced a β_0 as a constant of intrazonal trips. The estimation results are presented in Table 3 and Table 4.

$$U_{ij} = \beta_0 + \beta_{dist} DIST_{ij} Auto + \beta_{size} \ln(Total\ jobs_j) + \beta_{industrial} \frac{Industrial\ jobs_j}{Total\ jobs_j}$$

Distance was a significant and sensitive factor in the model. If the destination is one kilometer further, then its probability of being chosen is reduced by 75% compared to a destination at the origin of the trip. Supporting the previous destination choice model, we found a significant interaction between distance and auto ownership. Households with no cars tend to walk slightly longer than those who own cars. The size variable shows a significant and positive impact, while the share of industrial jobs has a barrier impact on choosing a destination. It suggests that destinations with more retail and service jobs and fewer industrial jobs are more attractive to choose. In contrast to the previous destination choice model, we added a dummy variable for checking if the destination zone is equal to the origin zone. In other words, it is a constant for intrazonal trips. The result represents that the origin zone has a higher probability of being chosen. This confirms expectations as we are using a superPAZ of 400 by 400 meters in the destination choice model, which leads to a significant number of intrazonal trips.

Table 3 Results of the SuperPAZ destination choice model (HBW, HBS, HBR)

	HBW			HBS			HBR		
	Estimate	Pr(>|z|)		Estimate	Pr(>|z|)		Estimate	Pr(>|z|)	
Distance (Km)									
x Auto (Yes)	-1.536	0.000	***						
x Auto (No)	-1.372	0.000	***						
x Child (Yes)				-2.182	0.000	***	-2.321	0.000	***
x Child (No)				-1.776	0.000	***	-1.955	0.000	***
Network density (Km)	0.141	0.008	**	0.049	0.209				
Size term (ln)									
Service	0.445	0.000	***				0.133	0.000	***
Retail				0.977	0.000	***			
Finance	0.352	0.000	***						
Government									
All other non-industrial									
Household							0.054	0.126	
Industrial prop.	-1.249	0.021	*	-1.306	0.003	**			
Slope (mean)	-0.167	0.024	*	-0.386	0.000	***	-0.139	0.001	***
Crossing Motorway	-0.321	0.18		-0.279	0.149		-0.568	0.032	*
Park (Yes)							0.662	0.000	***
Sample size	289			646			626		
Null model Log-Likelihood:	-1516			-3402			-3261		
Final model Log-Likelihood:	-952			-1598			-1910		
Pseudo R^2:	0.37			0.53			0.41		

Table 4 Results of the SuperPAZ destination choice model (HBO, NHBW, NHBO)

<table>
<tr><th></th><th colspan="3">HBO</th><th colspan="3">NHBW</th><th colspan="3">NHBO</th></tr>
<tr><th></th><th>Estimate</th><th>Pr(>|z|)</th><th></th><th>Estimate</th><th>Pr(>|z|)</th><th></th><th>Estimate</th><th>Pr(>|z|)</th><th></th></tr>
<tr><td>Distance (Km)</td><td>-2.217</td><td>0.000</td><td>***</td><td>-1.883</td><td>0.000</td><td>***</td><td>-2.141</td><td>0.000</td><td>***</td></tr>
<tr><td>x Auto (Yes)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>x Auto (No)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>x Child (Yes)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>x Child (No)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Network density (Km)</td><td>0.214</td><td>0.000</td><td>***</td><td>0.185</td><td>0.000</td><td>***</td><td>0.184</td><td>0.000</td><td>***</td></tr>
<tr><td>Size term (ln)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Service</td><td rowspan="5">0.389</td><td rowspan="5">0.000</td><td rowspan="5">***</td><td rowspan="5">0.667</td><td rowspan="5">0.000</td><td rowspan="5">***</td><td rowspan="5">0.516</td><td rowspan="5">0.000</td><td rowspan="5">***</td></tr>
<tr><td>Retail</td></tr>
<tr><td>Finance</td></tr>
<tr><td>Government</td></tr>
<tr><td>All other non-industrial</td></tr>
<tr><td>Household</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Industrial prop.</td><td></td><td></td><td></td><td>-0.749</td><td>0.117</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Slope (mean)</td><td>-0.381</td><td>0.000</td><td>***</td><td>-0.157</td><td>0.006</td><td>**</td><td>-0.060</td><td>0.220</td><td></td></tr>
<tr><td>Crossing Motorway</td><td>-0.828</td><td>0.000</td><td>***</td><td>-0.718</td><td>0.000</td><td>***</td><td>-1.361</td><td>0.000</td><td>***</td></tr>
<tr><td>Park (Yes)</td><td>0.510</td><td>0.000</td><td>***</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Sample size</td><td>1042</td><td></td><td></td><td>723</td><td></td><td></td><td>697</td><td></td><td></td></tr>
<tr><td>Null model Log-Likelihood:</td><td>-5438</td><td></td><td></td><td>-3728</td><td></td><td></td><td>-3621</td><td></td><td></td></tr>
<tr><td>Final model Log-Likelihood:</td><td>-2939</td><td></td><td></td><td>-1774</td><td></td><td></td><td>-1762</td><td></td><td></td></tr>
<tr><td>Pseudo R^2:</td><td>0.46</td><td></td><td></td><td>0.52</td><td></td><td></td><td>0.51</td><td></td><td></td></tr>
</table>

Table 5 and Table 6 present the estimation results of the PAZ-level destination choice models. In general, they had low goodness of fit. This could be due to fewer variations across small-scale destination zones or the lack of important factors in the model. Future studies need to investigate more factors, such as micro-level or street-level built environment variables (e.g., pavement condition and the number of trees).

Table 5 Results of the PAZ destination choice model (HBW, HBS, HBR)

	HBW			HBS			HBR		
	Estimate	Pr(>|z|)		Estimate	Pr(>|z|)		Estimate	Pr(>|z|)	
OriginPAZ	2.068	0.000	***	0.623	0.132		2.704	0	***
Distance (Km)	-1.335	0.000	***	-2.120	0.000	***	-1.974	0.000	***
Size term (ln)									
Retail	0.541	0.000	***	0.820	0.000	***	0.123	0.028	*
Service				0.188	0.000	***			
Finance									
Government									
Household	-0.433	0.000	***	-0.169	0.000	***	-0.560	0.000	***
Industrial prop.	1.629	0.000	***				-1.602	0.003	**
Park (acre)				-0.651	0.285		1.473	0.013	*
Sample size	289			646			626		
Null model Log-Likelihood:	-939			-2070			-1428		
Final model Log-Likelihood:	-813			-1617			-813		
Pseudo R^2:	0.13			0.22			0.14		

Table 6 Results of the PAZ destination mode choice model (HBO, NHBW, NHBO)

	HBO			NHBW			NHBO		
	Esti-mate	Pr(>|z|)		Esti-mate	Pr(>|z|)		Esti-mate	Pr(>|z|)	
OriginPAZ	3.162	0.000	***	0.654	0.000	***	1.628	0.000	***
Distance (Km)	-2.348	0.000	***	-2.894	0.000	***	-2.163	0.000	***
Size term (ln)									
Retail	0.145	0.000	***	0.316	0.000	***	0.355	0.000	***
Service									
Finance	0.559	0.000	***	0.062	0.016	***	0.137	0.001	***
Government									
Household	-0.507	0.000	***	-0.051	0.097	.	-0.082	0.022	*
Industrial prop.	-0.477	0.086	.				-0.694	0.058	.
Park (Yes)									
Sample size	1042			723			697		
Null model Log-Likeli-hood:	-3362			-2332			-2255		
Final model Log-Likeli-hood:	-2617			-2116			-1982		
Pseudo R^2:	0.22			0.09			0.12		

The parameters are calibrated to match the trip length frequency distribution and the average trip length reported for HBW trips in the OHAS data. After calibration, the estimated average trip length is 1.14 km (observed: 1.11km) and the cumulative trip length distribution of both estimated and observed results are similar (Figure 17).

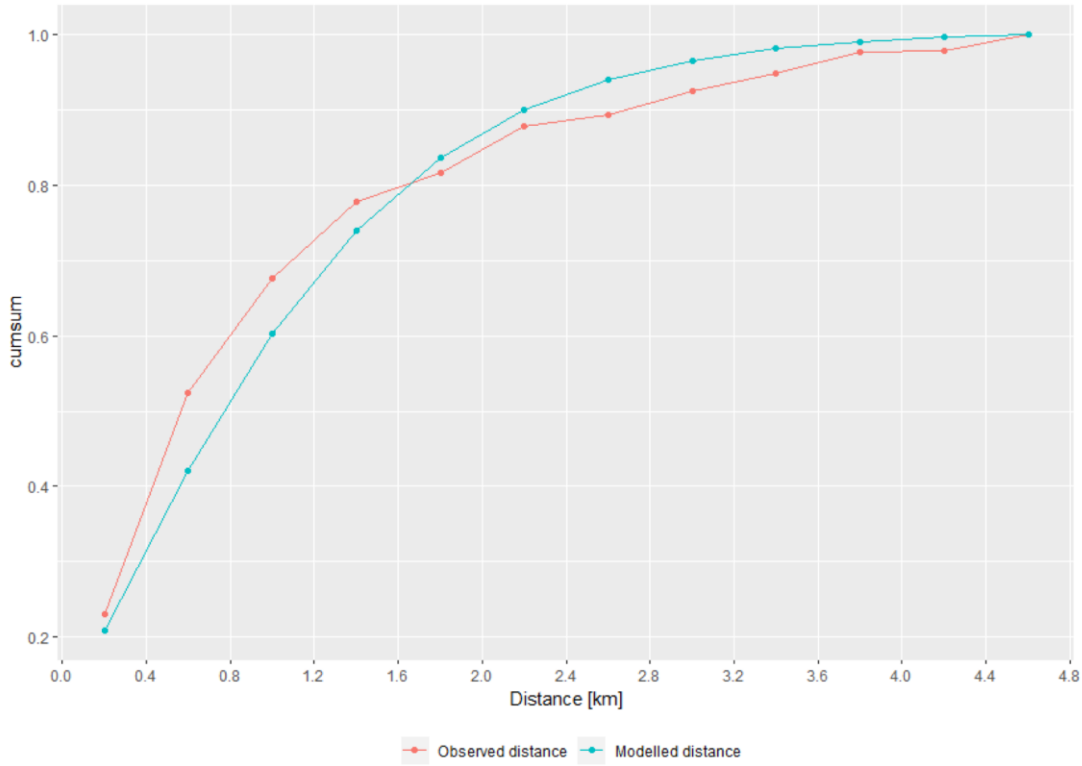


Figure 17 Cumulative trip length frequency distribution of modeled and observed trips

4.3. Discussion on MoPeD 2.0

While MoPeD was a well-established pedestrian planning tool, the enhancement made in this chapter upgrades the performance of MoPeD, particularly for model transferability and efficiency. The new pedestrian accessibility measurement replaced PIE in walk mode choice models. It is easy to construct, and it showed positive and significant influences on walk mode choice. However, the walk mode choice models in MoPeD 2.0 still had low model goodness of fit. This was mainly because travel distance was not incorporated in the mode choice models. The modeling sequence used in MoPeD 2.0 allowed us to better estimate the destination choice of walk activities at fine spatial resolution. However, it also eliminated the strong effect of travel distance in the mode choice model. Further research can focus on testing different modeling sequences or adding proxy parameters to represent distance in the mode choice stage. For example, adding habitual travel behavior, such as average trip length by purpose, may help improve the mode choice performance.

The walk destination choice models in MoPeD 2.0 better captured the relationship between built environment variables and the destination utility. The destination choice models at the superPAZ level had good performances and showed intuitive associations with the built environment. Network density and the accessibility to shops and retail stores showed strong positive effects on destination selection, whereas the proportion of industrial jobs, crossing the motorway, and slopes were barriers. However, the goodness-of-fit of PAZ-level discrete choice models was generally poor. It is challenging to estimate the destination choice at a very fine spatial unit. The PAZ-level selection may be related to the micro-level/street-level built environment attributes such as pavement condition and greenness. Since this information was rarely available, they were not considered in this study.

The enhanced MoPeD model will be applied in the following chapters. The application of policy and scenarios is presented in the next chapter, and the integration with the agent-based transport model MITO is introduced in Chapter 6.

5. Application of MoPeD 2.0 to scenarios in Portland Central City

To check the model sensitivity and explainability of MoPeD 2.0, the enhanced model is applied to the Portland Central City area for the assessment of urban and transportation scenarios. The impacts of scenarios are evaluated in terms of changes in walk shares, average trip length, and pedestrian flows. Section 5.1 presents the background and scenario setups. The case study areas are introduced in Section 5.2. Scenario results are discussed in Section 5.3.

The results presented here have been published in Zhang et al. (2022).

5.1. Background and Scenario Setups

With an increase in urbanization, Portland central city area (as shown in Figure 6) will continue to experience population and employment growth. The city projects it is going to gain approximately 38,000 new households and about 51,000 new jobs by 2035 (City of Portland, 2018). To meet this challenge, the city issued a set of goals and policies called the Central City 2035 (City of Portland, 2018). This new plan affirms that promoting walking is one of the solutions to build an efficient urban network and that the plan should put pedestrians at the forefront of city policies. The city will encourage investments in pedestrian facilities, such as pedestrian crossings, aiming to keep people walking safely and comfortably through the city. Combined with increases in the density and mix of use, these infrastructure and land use investments will support more travel by walking and other sustainable modes of transportation.

The benefits of urban pedestrian travel are well documented in the literature (Sallis et al., 2016; Saunders et al., 2013). Therefore, it is no surprise that cities like Portland incorporate these principles in their future policies. The city issued a set of goals and policies called the Central City 2035 (City of Portland, 2018). This new plan affirms that promoting walking is one of the solutions to build an efficient urban network and that the plan should put pedestrians at the forefront of city policies. However, planners and policymakers do not often have the appropriate tools to address planning questions and assess the impact of their policies on meeting their pedestrian-related goals. Regional travel demand models have been more focused on issues of moving vehicles and planning for their infrastructure needs and less on serving pedestrian behavior. Few practical applications focus on how these tools can be used to estimate future pedestrian demand in response to land use and transportation changes.

To illustrate the potential of pedestrian planning tools to assess the impact of future scenarios on pedestrian demand, we use the MoPeD 2.0 to assess urban and transportation scenarios for the Portland Central City area.

For the base year, we rely on population and employment conditions from the 2010 U.S. Census for these neighborhoods. For future conditions, we analyzed the impact of a planned pedestrian bridge and a newly-built car-free crossing. Aligned with these facilities, neighborhood land use scenarios have been created based on the Portland Central City 2035 Plan (City of Portland, 2018). New households and different types of employment are allocated to PAZs based on this plan for all ten districts in the Central City area. The case study locations are described in more detailed way in Section 5.2.

We implemented the MoPeD 2.0 model to test urban development scenarios. These scenarios serve to model various land use and transportation policies to assess to which degree the built environment supports an increase in the share of walk trips. For each of the case study locations, the following scenarios will be modeled:

A. **2010 Base year:** A 2010 base year scenario based on the census 2010 population and employment data.
B. **2035 with average growth:** A 2035 future year scenario with an average population and employment growth across all locations.
C. **Scenario B + Infrastructure:** Scenario B with pedestrian bridges completed and a denser street network.
D. **2035 with Central City Plan:** A 2035 future year scenario with population and employment growth corresponding to the Central City Plan.
E. **Scenario D + Infrastructure:** Scenario D with pedestrian bridges completed and a denser street network.

Scenario A is the baseline scenario. It employs the population and employment distribution in 2010. Scenario B is a business-as-usual scenario with an average of 1.5% increase in population and employment across all locations. In scenario D, future population, and job growth for 2035 are applied corresponding to the Central City Plan, which is described in the previous section. Pedestrian accessibility measures are recalculated with the new population and new jobs. In scenarios C and E, pedestrian facilities are tested with different population and job growth strategies. As a result of the new bridges and new pedestrian links, the pedestrian catchment area is enlarged, and the accessibility measures also increase.

5.2. Case Study Areas

The study area for this modeling exercise is the Portland Central City shown in Figure 6. The Central City area consists of ten different neighborhoods and stretches from the West Hills to East 12th Avenue, and from the Pearl and Lower Albina to the South Waterfront area and Powell Boulevard (City of Portland, 2018). The Willamette River divides this area and is spanned by several bridges, including the non-automobile bridge Tilikum Crossing, completed in 2015.

Table 7 shows the population and employment for each of these neighborhoods in 2010, which is the base year for the models, and projected growth in each for 2035, which is the modeled horizon year. In the base year, Downtown is the economic center with most of the office employment, retails, and services. The Pearl District is a mixed-use district with commerce and retails and the largest number of households. The Central Eastside and Lloyd Districts are characterized as an industrial center and an office core, respectively, and are less populated. South Waterfront is not yet developed and has the lowest density of population and jobs. The base year 2010 was chosen due to the availability of detailed land use and infrastructure at the PAZ level for that year.

In the Portland Central City 2035 Plan (City of Portland, 2018), scenarios are outlined for each of these neighborhoods in terms of projected residential and employment growth. In addition, there is a planned pedestrian crossing (the Congressman Earl Blumenauer Bicycle and Pedestrian Bridge) that will connect the Central Eastside with the Lloyd District and the Tilikum Crossing. It was built in 2015, which is after our base year, and is a car-free facility that links the Central Eastside with the South Waterfront.

By 2035, the Central City will gain approximately 2-times more households compared to the base year and a roughly 40 percent growth in jobs. The Central Eastside, Lloyd District, and South Waterfront will be the focus of demographic growth in the future decades, with an increase of households by 800%, 778%, and 364% respectively. The emphasis of the employment development is expected to be on the South Waterfront.

Table 7 Total household and employment in 2010 and projected for 2035 by districts.

District	Total households			Total employment		
	2010	2035	Change in %	2010	2035	Change in %
Central Eastside	900	7900	+778%	17000	25000	+47%
Downtown	1600	4600	+188%	48200	55200	+15%
Goose Hollow	3900	4900	+26%	5300	7300	+38%
Lloyd	1000	9000	+800%	16800	25800	+54%
Lower Albina	100	300	+200%	2100	2300	+10%
Old Town	1900	3900	+105%	5200	8200	+58%
Pearl	5600	11600	+107%	10700	14700	+37%
South Waterfront	1100	5100	+364%	1200	11200	+833%
University District	3200	6200	+94%	10400	14400	+38%
West End	3800	6800	+79%	6900	9900	+43%
Sum	*23100*	*60300*	*+161%*	*123800*	*174000*	*+41%*

To understand the impacts of planned infrastructures and growth on pedestrian travel, we have incorporated these realistic scenarios and allocated the projected population and employment growth to our PAZ structure based on the details provided in the Central City 2035 Plan. Each of the future scenarios outlined below is for the planning horizon of 2035 and compared against the base year 2010. Besides the locations described in more detail below, we also distributed housing and employment growth to other neighboring districts based on the Portland Central City 2035 Plan (City of Portland, 2018). Each district has a different distribution of population and employment growth based on the corresponding future development vision. The table in the Appendix 1 summarizes the detailed housing and jobs growth plan across ten districts in Central City.

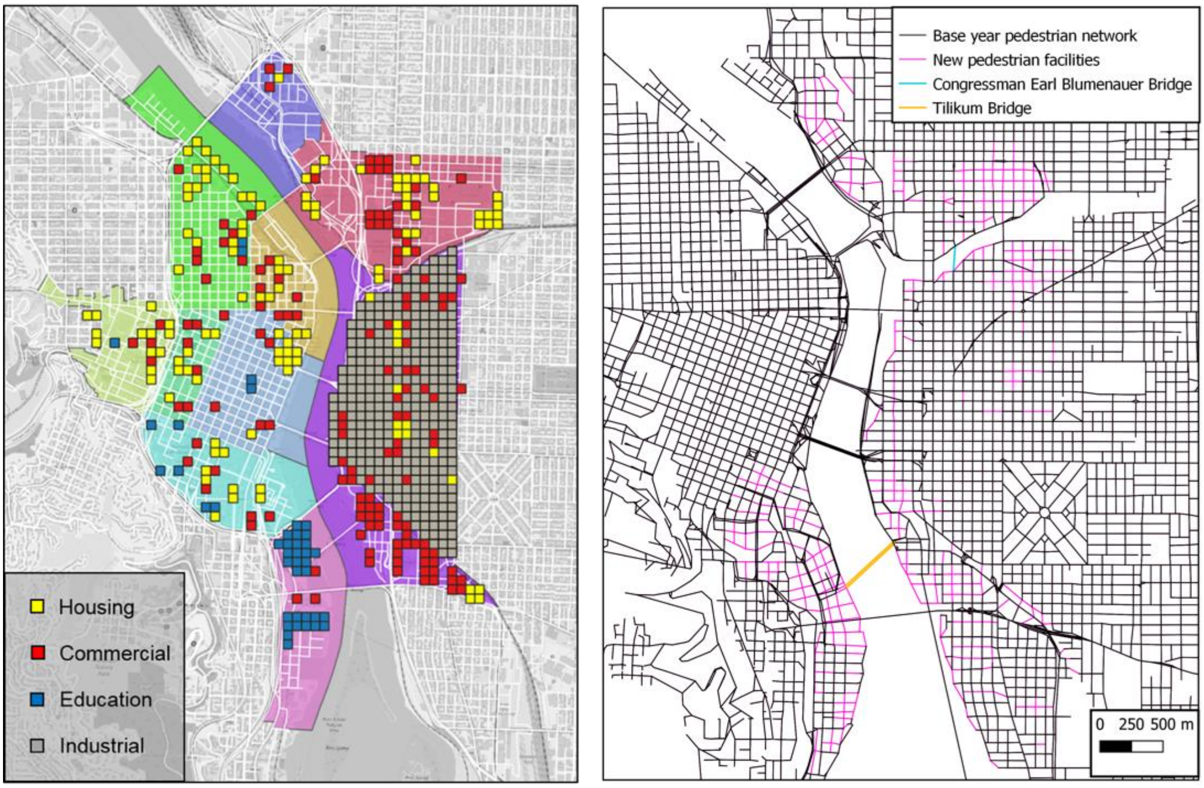


Figure 18 An overview of land use plans (left) and pedestrian facilities (right) for 2035 in the Central City.

There are two specific pedestrian bridge scenarios that we compared to the 2010 base year conditions. One is a new pedestrian bridge and the growth projected for the Lloyd Center and the Central Eastside (Case Study 1). The other is the Tilikum Bridge and the growth projected for the South Waterfront and the Central Eastside (Case Study 2). Besides two pedestrian bridges, more detailed pedestrian networks are added to several areas including the south triangle of Central Eastside, Lloyd District, and South Waterfront. In total, 41.2 kilometers of new pedestrian links were added in 2035 by extending the current grid network. These are described in more detail below and an overview is shown in Figure 18.

5.2.1. Case Study 1 Lloyd Center-Blumenauer Bridge-Central Eastside

A new pedestrian and bicycle facility – the Congressman Earl Blumenauer Bridge – is currently being constructed to link the Lloyd District and Central Eastside (Portland Bureau of Transportation, 2019). We aim to examine the implications of this increased connectivity and the anticipated growth described below.

The Lloyd District has been identified as an “eco-district” with a focus on equitable, sustainable, and resilient development. Between 2010 and 2035, Lloyd is expected to grow by 8,000 households and 9,000 jobs to a total of 9,000 households and 25,800 jobs. In this study, 8,000 households are distributed evenly over 33 PAZs identified as housing in Figure 19 (in yellow). Each PAZ added 242 households, and the same demographic attributes as in the 2010 distribution were assumed. For 9,000 jobs, we assume 25% distributed each to retail, finance, services, and government. In this scenario,14 PAZs around Convention and Lloyd are considered as office cores (in red). Accordingly, they will get all new finance employment (2,250), all government (2,250) jobs, 25% of service jobs (562), and 10% of retail (225). The remaining 9 PAZs will get 65% of service (1,462) and 70% of retail (1,575) employment. The remaining 10% of service (225) and 20% of retail (450) will be distributed across 33 housing PAZs.

Over the same period, the Central Eastside is expected to grow by 7,000 households and 8,000 jobs, for a total of 7,900 households and 25,000 jobs. In the growth scenario, 7,000 households are distributed evenly over the 15 PAZs identified as housing in Figure 20 (in yellow). Each PAZ added 467 households and the same demographic attributes as in the 2010 distribution were assumed. In the lower triangle, 41 PAZs will get 75% of the employment growth (shown in red). The remaining 13 PAZs targeted for commercial will get 1,500 jobs distributed by service, retail and financial. The remaining areas will realize a total growth of 500 jobs in industrial employment, distributed evenly over all the PAZs.

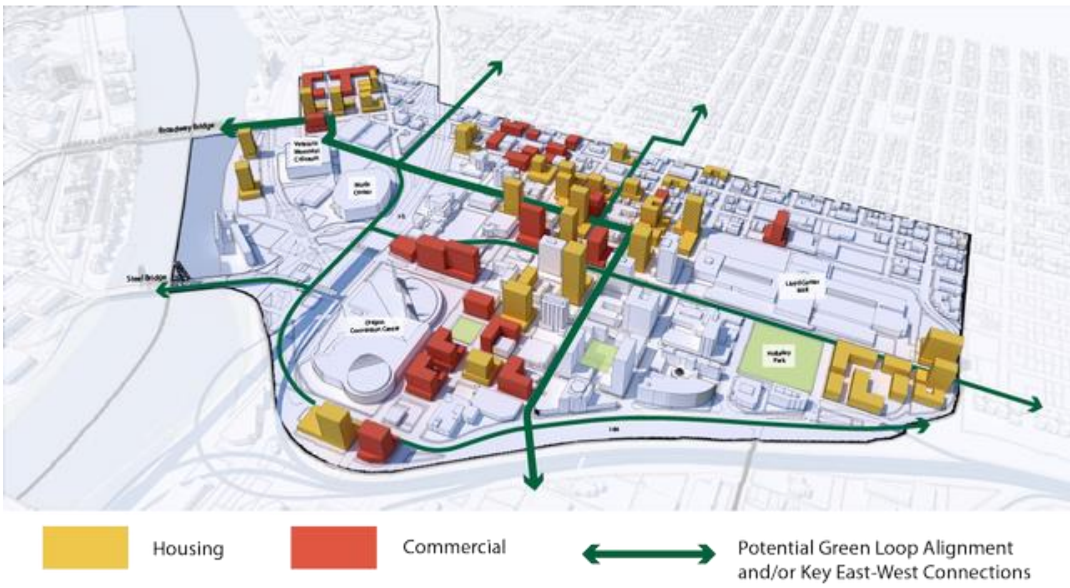


Figure 19 Land use development plan 2035 of Lloyd District (City of Portland 2018)

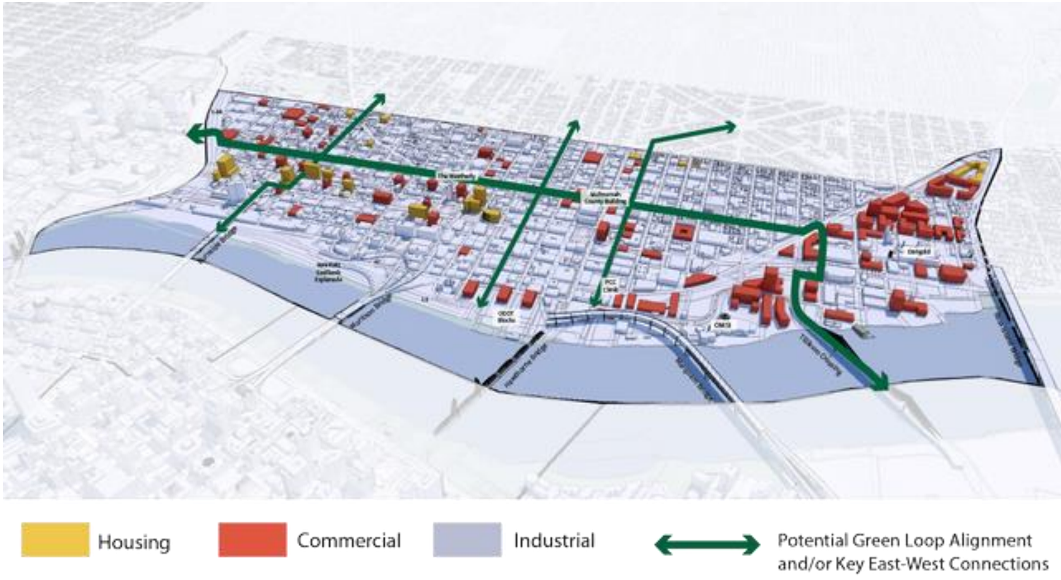


Figure 20 Land use development plan 2035 of Central Eastside (City of Portland 2018)

5.2.2. Case Study 2: Central Eastside-Tilikum Crossing-South Waterfront

The Tilikum Crossing was completed in 2015 and is the longest car-free bridge in the United States. It spans the Willamette River, linking Portland's South Waterfront to the Central Eastside Industrial District, described above.

The South Waterfront is a dense, walkable, mixed-use community and is expected to grow by 4,000 households and 10,000 jobs, for a total of 5,100 households and 11,200 jobs from 2010 to 2035. The bridge directly links the South Waterfront to the development planned near the Oregon Museum of Science and Industry (OMSI) on the Central Eastside.

In the South Waterfront plan for 2035, 4,000 households are distributed evenly over the 22 PAZs identified with housing. Each of these PAZs added 182 households and assumed the same attributes as the 2010 distribution. In this scenario, 8,000 government jobs are allocated to institutional PAZs, while 1,000 service jobs are allocated to commercial PAZs. To account for the mixing of land uses, 1,000 retail jobs are evenly distributed to all PAZs.

5.3. Scenario Discussion

5.3.1. Impacts on Network Connectivity

Network connectivity for pedestrians can be measured by pedestrian catchment ratio (PCR). Here the PCR is the ratio of the pedestrian catchment area to the theoretical circle area with a radius of 800 meters around the centroid of the same PAZ. The higher the PCR, the better the network connectivity. Figure 21 shows that the distribution of PCR under the bridge scenario is generally shifted to the right, which indicates that new bridges and pedestrian streets can help improve pedestrian network connectivity.

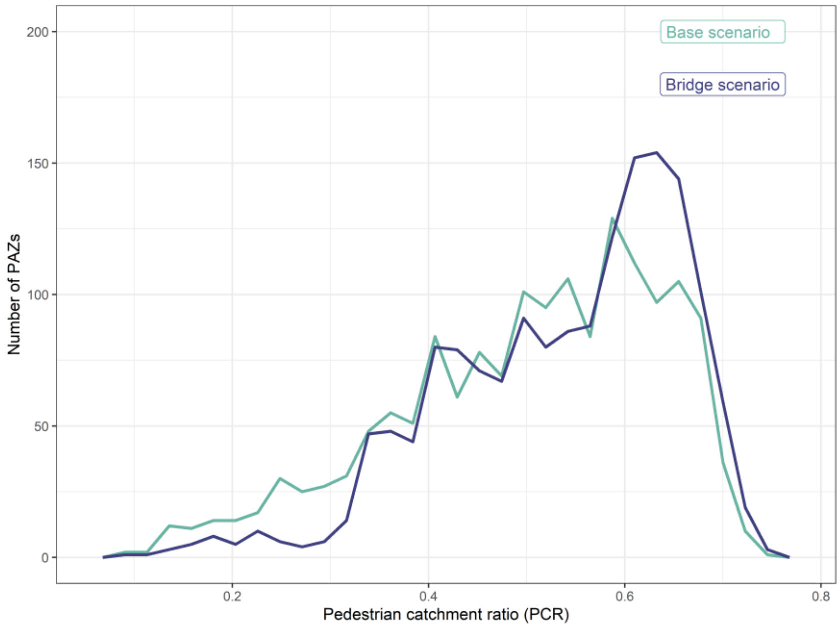


Figure 21 Frequency distribution of pedestrian catchment ratios under different scenarios

Figure 22 shows the spatial distribution of the increase in pedestrian catchment ratio (PCR) comparing between with and without the new bridges and new pedestrian streets. Most PAZs close to the new bridges experience an improvement in network connectivity. The newly built pedestrian streets also play important roles in improving network connectivity, which leads to dramatic increases of PCR on the left side of the Tilikum bridge.

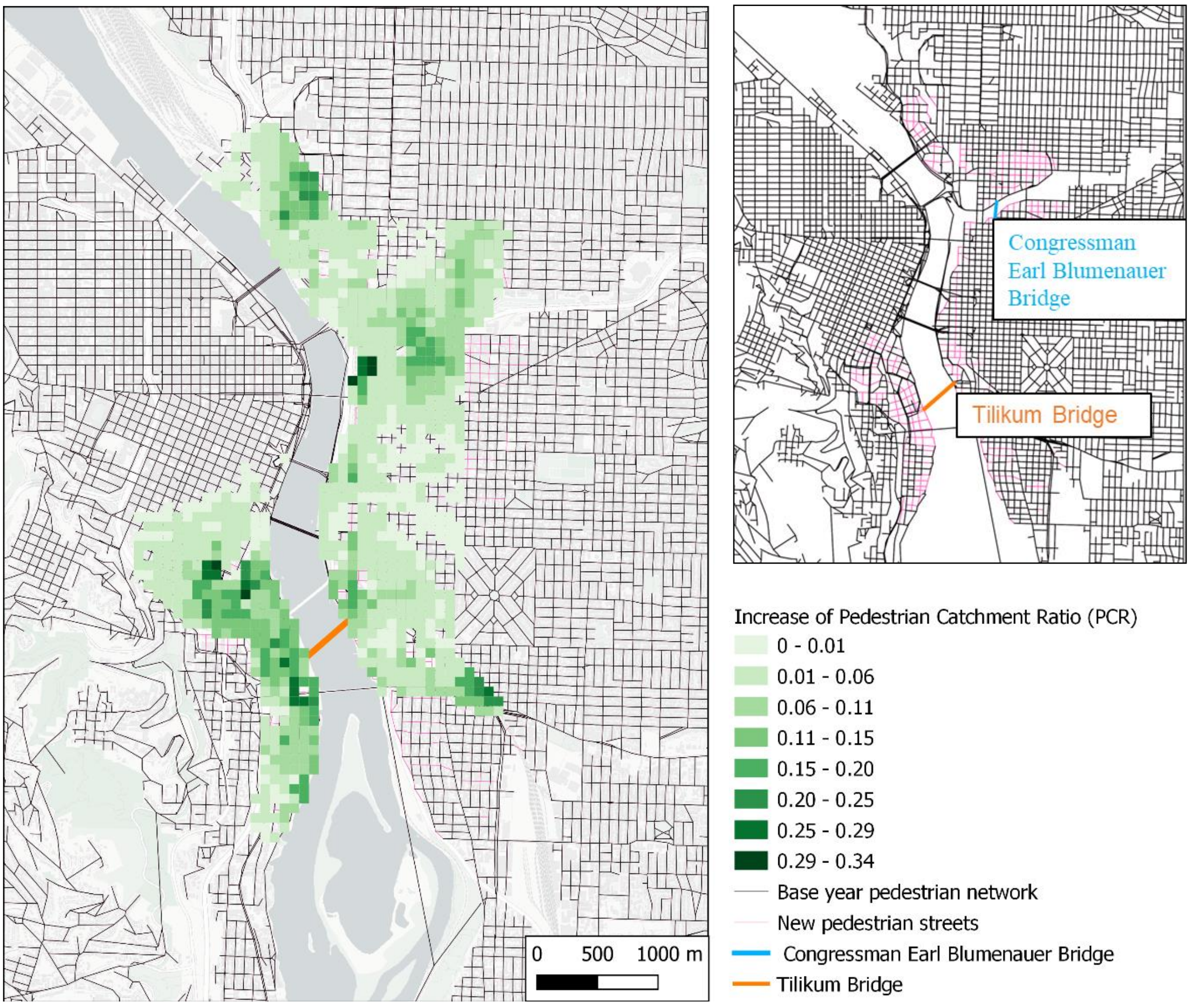


Figure 22 Increases of pedestrian catchment ratio of each PAZ when comparing between with and without new bridges and new pedestrian streets.

5.3.2. Impacts on Walk Share

The policy scenarios evaluated in this study have varying degrees of impact on the share of walk trips. Two scenarios with average growth (Scenario B and C) have little impact on increasing the walk share, while two Central City Plan scenarios (Scenario D and E) significantly influence the walk share. Figure 23 compares the distributions of PAZ walk shares based on different scenarios. Table 8 provides an overview of the number of walk trips and walk shares under five scenarios as well as their relative changes compared to the baseline.

Table 8 Total trips, walk trips and walk share across the whole Portland Central City area of five scenarios.

Scenario	A: 2010 Base year	B: 2035 with average growth	C: Scenario B + infrastructure	D: 2035 with Central City Plan	E: Scenario D + infrastructure
Households	23,100	23,446	23,446	60,300	60,300
…% change compared to base		*+1.5%*	*+1.5%*	*+161.0%*	*+161.0%*
Number of trips (all modes)	282,948	287,170	287,170	533,367	533,367
…% change compared to base		*+1.5%*	*+1.5%*	*+88.5%*	*+88.5%*
Number of walk trips	84,452	86,255	87,624	184,370	189,174
…% change compared to base		*+2.1%*	*+3.8%*	*+118.3%*	*+124.0%*
Share of walk trips	29.8%	30.0%	30.5%	34.6%	35.5%
…% change compared to base		*+0.6%*	*+2.2%*	*+15.8%*	*+18.8%*
Total trips/household	12.25	12.25	12.25	8.85	8.85
…% change compared to base		*0.0%*	*0.0%*	*-27.8%*	*-27.8%*
Walk trips/household	3.66	3.68	3.74	3.06	3.14
…% change compared to base		*+0.6%*	*+2.2%*	*-16.4%*	*-14.2%*

In the two scenarios with average growth (Scenario B and C), the population evenly grows by 1.5% across all PAZs with the assumption that household compositions remain unchanged. The distribution of walk shares by PAZ are very similar to the baseline. As shown in Table 8, the overall walk shares of the two average growth scenarios are fairly close to the baseline walk share.

In the Central City Plan scenario, the distribution of PAZ walk shares shifts towards the right, which indicates moderate-high walk shares. Most PAZs experience an increase in the share of walk trips. The same shift is observed in the Central City Plan scenario with the pedestrian facility development. The shift is even slightly larger than in the Central City Plan scenario without infrastructure (Scenario D). The pedestrian facility development appears to only show an impact for zones in the catchment area of the bridges. Overall, under two Central City Plan scenarios, the whole Central City area will produce roughly 100,000 more walk trips in 2035 (an increase of about 120%).

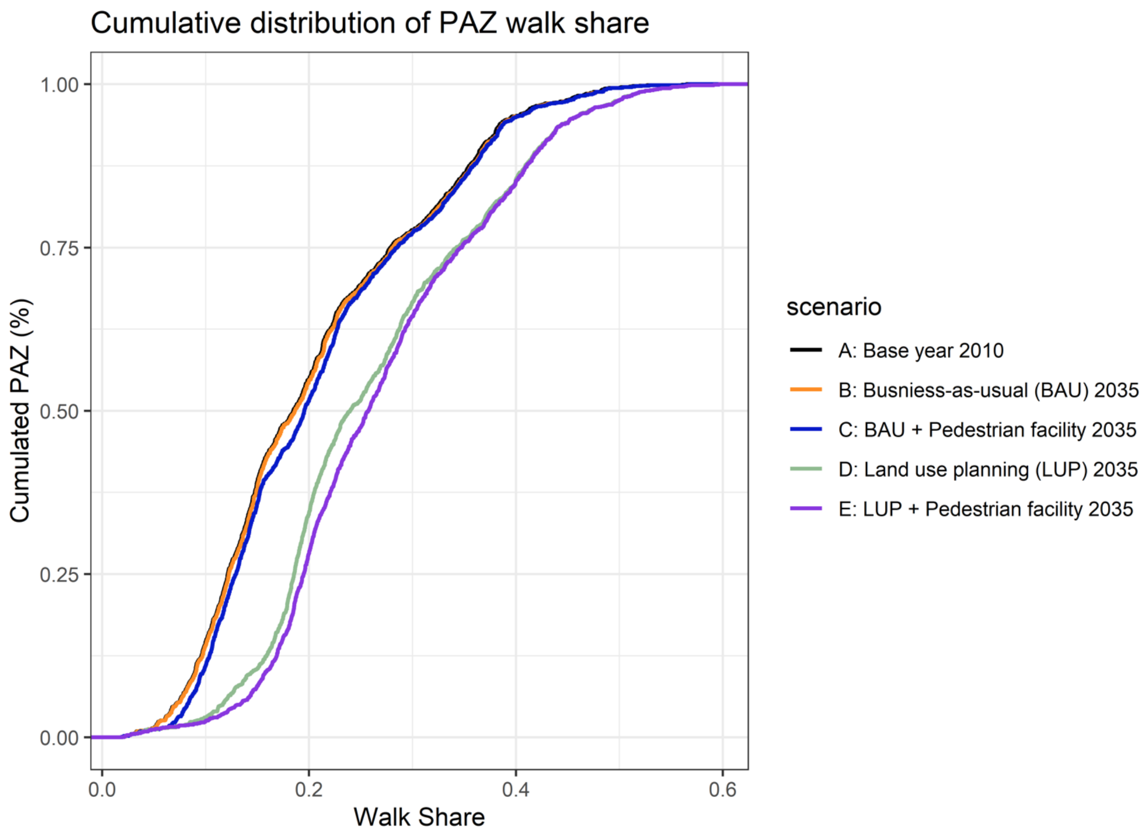


Figure 23 Cumulative distribution walk shares by PAZ.

Although the Central City Plan scenarios have notable effects on encouraging higher walk shares, it is observed that the value of average walk trips per household decreases by 16.4% and 14.2% separately in two Central City Plan scenarios (shown in Table 8). One reason for this notable decrease is the assumption that the household composition of PAZs with no household in the base year follows the average distribution of households in the Central City Plan scenarios, where 49% are assumed to be single-person households. Those households tend to generate fewer trips than larger households (Zhang, Clifton, Moeckel, et al., 2019), reducing the number of walk trips per household. The decrease may also imply the limitation of the trip generation model used in MoPeD. In this study, trip generation models can only reflect the demographic changes but are insensitive to the changes in land use development. The effects of pedestrian accessibility are currently not considered.

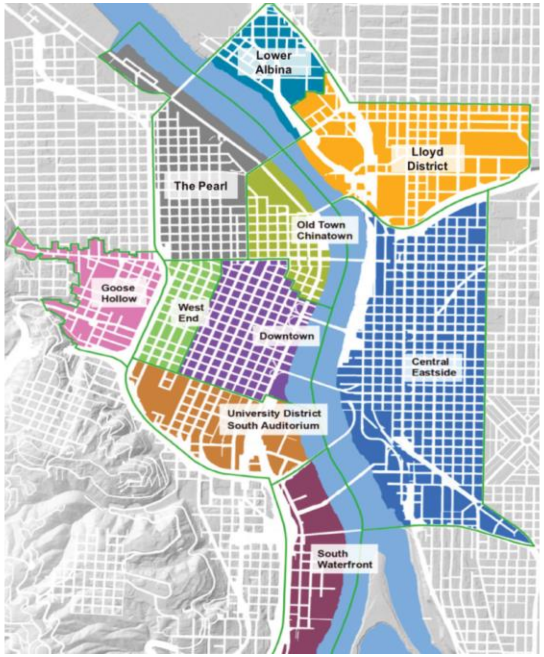
The following sections will illustrate the impacts of policy scenarios on individual districts. Table 9 provides an overview of the resulting walk shares of each district under different scenarios and the comparison to the reference scenario.

In the base year 2010, the West End district has the highest walk share, followed by its two neighboring districts, Downtown and Old Town. A higher density of households and jobs, as well as good street connectivity, create an attractive built environment to support walking in those districts. South Waterfront is the least walkable district because in the base year it was not yet developed with vacant brownfield sites and buildings were underutilized. Central Eastside and Lower Albina also have relatively low walk shares in the base scenario. This might be because they were characterized by an industrial core with a high share of manufacturing buildings and a low share of residential and commercial land use. Under the average growth scenario, the characteristics of each district are retained, and the growth is evenly distributed. Thus, walk shares are slightly increased.

As expected, the Central City Plan scenario without infrastructure leads to an increase in the walk share in all districts. Their increases in walk shares range from 12% to 119%. In particular, the walk share in South Waterfront is more than doubled. The change is caused by a large amount of development in housing and employment with a total of 4,000 new households and 10,000 jobs. Similar to South Waterfront, Lloyd District and Central Eastside also gain substantial increases in walk shares due to the rapid and large-scale development in residential and commercial uses. It reveals that the increased rates of walk shares largely depend on the number of new households and new jobs in the district and in the neighboring districts. However, the walk share does not necessarily grow proportionately with population and jobs. The walk mode choice model is designed with a logarithmic relationship between walk shares and pedestrian accessibility. Thus the marginal impact of additional population and jobs is a decreasing function. Figure 24 shows the nature of the logarithmic relation. It indicates that the magnitude of rates of change in walk shares highly depends on the baseline pedestrian accessibility. For example, although a large number of new households and jobs are placed in the Downtown and West End districts (purple and light green dots in Figure 24), the walk shares of these two districts grow only moderately compared to other districts. It suggests that Portland downtown is already very dense and has already reached a certain level of pedestrian accessibility. More activity density does not encourage many more walk trips.

In the scenarios with pedestrian facility development (Scenario C and E), the increase in walk shares is much more pronounced. When comparing the infrastructure scenarios with their corresponding growth-only scenarios (Scenario B and D), the change in walk shares of the infrastructure scenarios only occur in the districts associated with the new pedestrian facilities. Those are Central Eastside, Llyod District, Lower Albina, South Waterfront, and University District. Almost no changes are noticeable in the remaining districts.

Table 9 Share of walk trips across different districts in five scenarios



Scenario/District	**A: 2010 Base year**	**B: 2035 with average growth** % change compared to A	**C: Scenario B + infrastructure** % change compared to B	**D: 2035 with Central City Plan** % change compared to A	**E: Scenario D + infrastructure** % change compared to D
CENTRAL EASTSIDE	14.1%	+0.8%	+4.9%	+55.4%	+4.1%
DOWNTOWN	38.5%	+0.6%	+0.1%	+19.0%	+0.1%
GOOSE HOLLOW	27.9%	+0.7%	+0.0%	+17.3%	+0.0%
LLOYD	20.6%	+0.8%	+3.6%	+43.8%	+2.7%
LOWER ALBINA	9.7%	+0.9%	+6.1%	+62.8%	+1.3%
OLD TOWN	32.3%	+0.6%	+0.0%	+16.6%	+0.0%
PEARL	28.8%	+0.7%	+0.0%	+12.0%	+0.0%
SOUTH WATERFRONT	8.3%	+1.0%	+13.5%	+118.9%	+24.1%
UNIVERSITY DISTRICT	28.7%	+0.7%	+9.5%	+12.3%	+9.0%
WEST END	41.8%	+0.5%	+0.1%	+13.0%	+0.0%

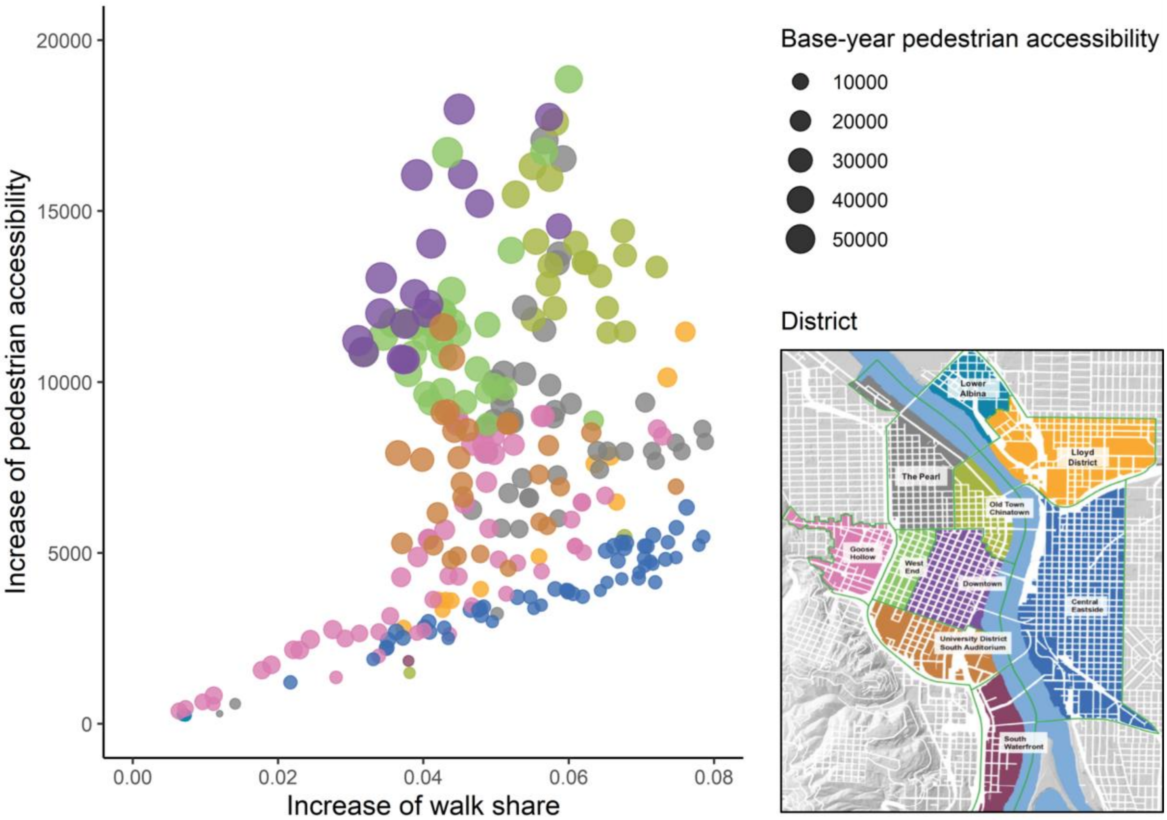


Figure 24 The change in walk shares and pedestrian accessibility (defined as population + employment within 800 meters) in the baseline scenario and the Central City Plan scenario without infrastructure.

To have a closer look into the pedestrian facility scenario, we compared the walk share of each PAZs between scenario D and scenario E (shown in Figure 25). New bridges and links enlarge the pedestrian catchment area because of good and direct connectivity. Most PAZs close to the new facilities experience an increase in walk share. PAZs located on the north side of Congressman Earl Blumenauer Bridge have smaller increases in walk shares than those located on the south side. According to the land use development plan shown in Figure 18, the south side of the bridge generally has less diverse growth than the north side of the bridge. While people living in the north of the Lloyd District could easily visit locations on the other side of the bridge, a lack of diversity on the south side of the bridge limits the growth in walk trips. The same situation of unbalanced growth is also found near the Tilikum Bridge. The bridge offers good connectivity to the west side of the river. Nevertheless, the land use growth on the west side lacks diversity and focuses on education. Thus, the Tilikum Bridge is not as attractive for people working and living on the east side of the bridge.

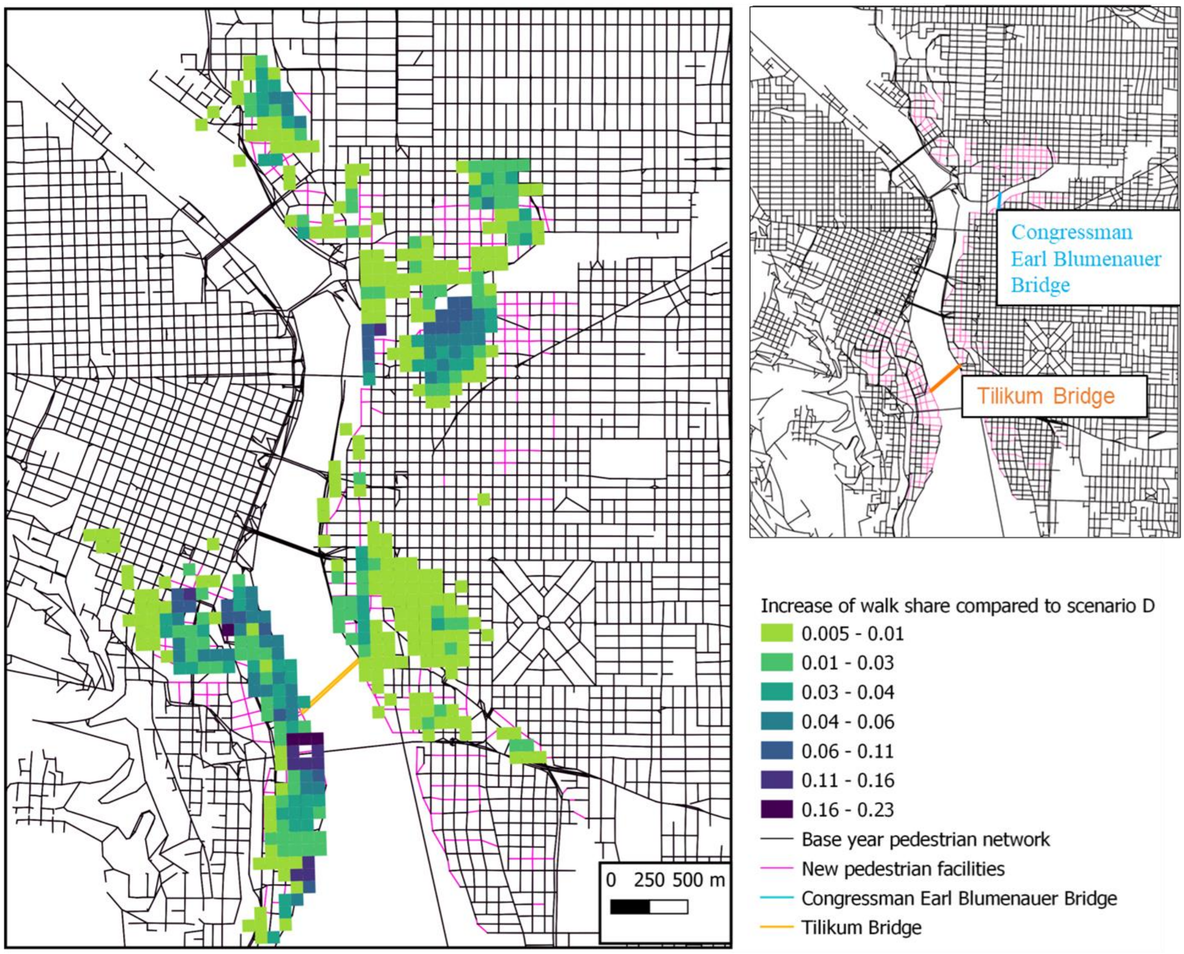


Figure 25 Comparison of walk shares under Central City Plan scenarios with/without pedestrian facilities (scenarios D and E)

5.3.3. Impacts on Average Trip Length

The average trip distance is largely dependent on the attractiveness of the surrounding district and the district itself. When the district itself has a good street network to access many opportunities, people tend to travel within the district. Otherwise, people will be attracted more strongly to neighboring districts. Figure 26 shows the cumulative distribution of the average trip length by PAZ under the five scenarios. In general, the impact of land use and infrastructure on average trip length is not particularly large. Three typical districts that represent three different distribution patterns are selected for discussion.

Scenario B with average growth has almost no effect on average trip length, and the curves of baseline and scenario B mostly overlap. Under the average growth scenario with pedestrian facility development in the Llyod district (blue lines in Figure 26), the distribution noteworthily shifts to the left, indicating an increase in walk shares. However, the distribution in the South Waterfront shows rather moderate changes. This is because the South Waterfront is still undeveloped in the average growth scenarios. Although the denser pedestrian network improves connectivity in the district, densities are relatively low. This suggests that the development of a denser pedestrian network without corresponding population and employment densities has a limited impact on walk shares.

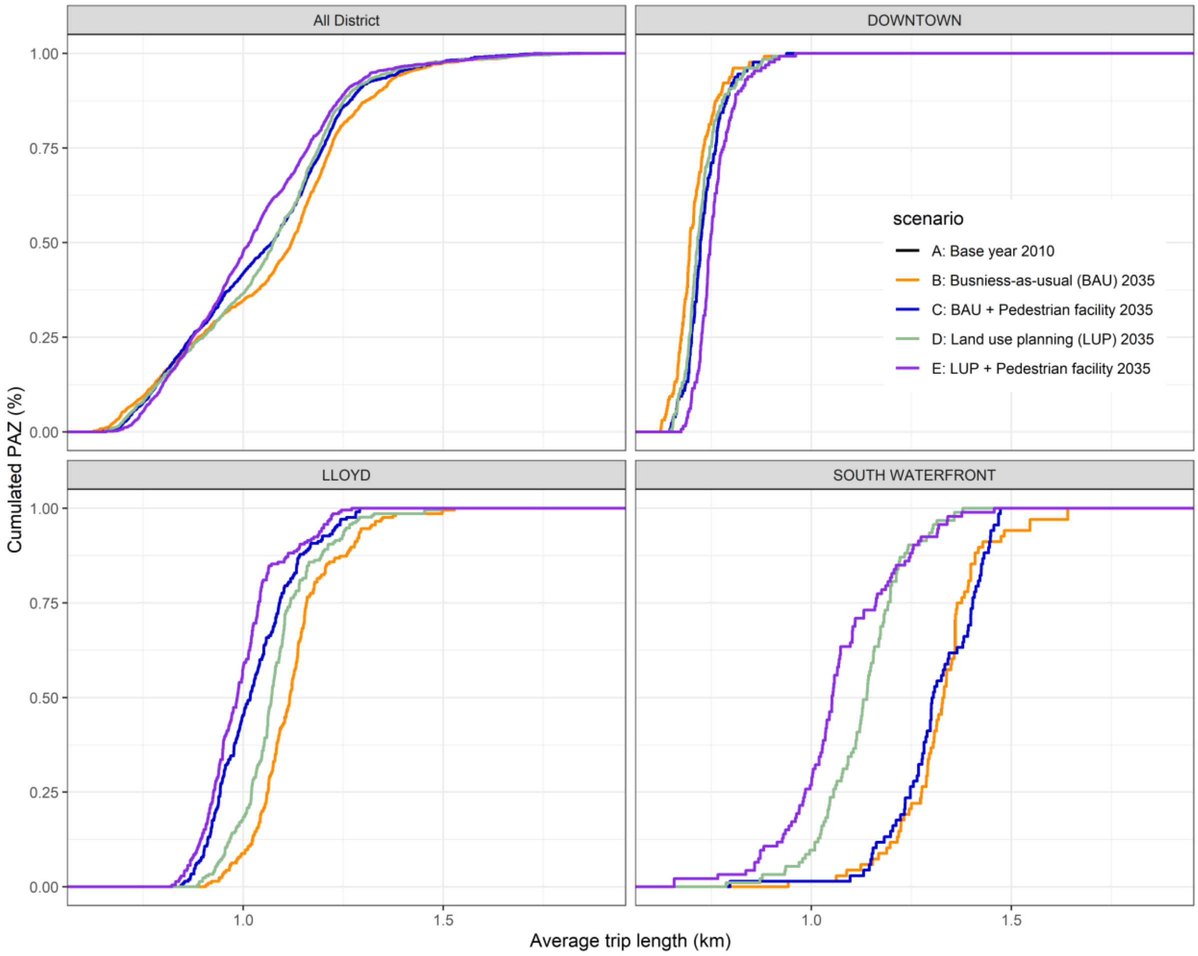


Figure 26 Cumulative distribution of average trip length by PAZ in selected districts

5.3.4. Impacts on Pedestrian Flows

Pedestrian route choices were implemented to assign walk trips to the pedestrian network. Figure 27 demonstrates how the pedestrian flows are distributed in the pedestrian network in different scenarios.

In the base year scenario, most of the pedestrian flows are populated in the downtown areas, with the highest segment usage of 2596 pedestrian trips. Many streets on the east side of the river had daily pedestrian trips under 150. Due to the moderate growth strategy, the pattern of pedestrian flow distribution in scenario B is unchanged. In scenario C, the Congressman Earl Blumenauer bridge is used by 320 pedestrians, and it also slightly impacts the volumes of the surrounding links. However, the usage of the Tilikum bridge is relatively low with an average pedestrian volume of 111 in both directions. Scenario D with Central City Plan significantly influences the pedestrian volumes in the entire study area. On the one hand, the pedestrian flows in the downtown area boosted to a higher level. The busiest street is crossed by 7321 pedestrians. On the other hand, there are more pedestrian flows occurring on the east side of the river, especially in the Lloyd District. It also increases the usage of the existing bridges connecting the two sides of the river. In scenario E, the pedestrian flows are sprawled across the east side of the study area. The usage of the minor roads is increased. The usage of the Tilikum bridge is increased to an average pedestrian volume of 797 in both directions. The variation plots in Figure 27 give us a better understanding of how the new pedestrian infrastructure impacts daily pedestrian flow. The Tilikum Bridge facilitates the number of pedestrians along the west side of the river, while the Congressman Earl Blumenauer bridge in the Lloyd District enhances the pedestrian volume along the freeway. Also, the new pedestrian roads in the Lloyd District attract a large number of walkers as well as relieve the burden of pedestrian traffic on the surrounding main roads.

Figure 27 Daily pedestrian volumes on the pedestrian network in different scenarios.

5.4. Discussion on Portland Scenarios

Overall, the scenario application presented in this chapter proved that the pedestrian planning tool MoPeD 2.0 is sensitive to the small-scale variations in local land use and transport development. It can help the policymakers to have a better understanding of the effects of various demographic policies and infrastructure planning on the walk probability. Most importantly, it can address planning issues. It can assess how the Central City 2035 policies support increased pedestrian activities. Moreover, it can support planning effective pedestrian networks based on maximizing accessibility and connectivity

On the other hand, this application exercise also reveals the limitations of MoPeD 2.0, which should be addressed in future work. The pedestrian model appears to be only sensitive to the level of street connectivity rather than the quality of street connectivity. The new bridge with the wide and dedicated pedestrian lane is treated the same as the old bridge with an unpaved pedestrian lane. This could be one of the reasons that there are no big changes in walk share in the bridge applications. The mode choice model lacks the attributes that can reflect the quality of connectivity. Therefore, in future research, characteristics of the pedestrian facility need to be added to the choice model. Furthermore, when we allocated households and jobs into the PAZ structure, we assumed the average demographic attributes based on the 2010 distribution. Realistic demographic changes such as aging or car ownership changes are not considered in this scenario application. MoPeD's aggregated modeling approach prevents us from setting up demographic changes in a more realistic way. This is one of the key motivations for integrating MoPeD into the agent-based environment of MITO. The MoPeD and MITO integration will be presented in the next chapter.

6. MoPeD Meets MITO: an Integrated Agent-Based Model for Pedestrian Travel Demand

In the previous two chapters, we introduced the enhanced MoPeD and showed its ability to capture small-scale variations in land use and transport policy. However, MoPeD is limited to its nature of aggregated modeling. Average social-demographic attributes are simulated at the zonal level, which makes it difficult to consider demographic changes such as aging or car ownership in scenario application. Moreover, it also has a limited representation of non-walk modes. Agent-based transport model MITO can complement the limitations of MoPeD. By simulating at the individual level, MITO can help researchers identify transport issues at a fine resolution. The outcomes could be used to calculate an individual's units of physical activity (metabolic equivalents, METs) which are critical inputs to health impact assessment models.

Therefore, in this chapter, we propose an integrated modeling framework (referred to as MITO/MoPeD) that incorporates the fine-grained resolution model of pedestrian demand (MoPeD) into a sparser spatial resolution of an agent-based transport model (hereafter referred to as the Munich Model). The author attempts to explore whether the integrated modeling framework (MITO/MoPeD) has a better performance than the Munich Model in terms of simulating walk-related metrics. The central hypothesis is that the Munich Model would not be suitable for pedestrian modeling and health impact assessment, but the following adaptations from MoPeD could improve their accuracy:

- Separate walk trip decisions from the multimodal models.
- Apply a finer zone system for walk trip decisions.
- Apply more built environment factors for walk trip decisions.

The approach to integrate MoPeD and MITO is presented in Section 6.1. After that, the input data containers are prepared for the Munich context (Section 6.2). Since the mode choice and destination choice models were developed in the Portland context and then married into the Munich study area, a model transfer/calibration process is presented in Section 6.3. To check the plausibility, the Munich model and the MITO/MoPeD model are applied to the Munich study area. In Section 6.4, model performances are discussed based on the prediction of walk shares, walk trip length distribution, the spatial distribution of walk trips, and physical activity volumes.

6.1. Model Integration

Figure 28 shows the working process of the MITO/MoPeD model. First, MITO provides trip generation at the individual level to MoPeD. In the MoPeD module, walk trips are generated and processed, and then fed back to MITO.

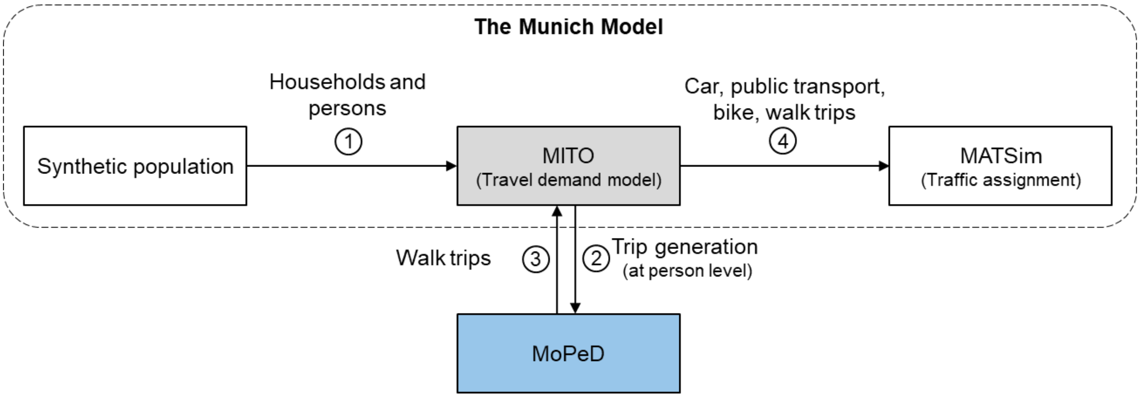


Figure 28 The framework of the integrated modeling framework – the MITO/MoPeD model

The incorporation of MoPeD and the Munich Model is not trivial. MoPeD is carried out with a fine-grained resolution, whereas the Munich Model uses a much sparser TAZ as its spatial unit. The difference in spatial resolution should be adapted. In addition, the walk trip decision sequences are not aligned in MITO and MoPeD. As shown in Figure 4, individuals in MITO first choose the trip destination and then the trip mode, while in MoPeD individuals choose to walk or not first, then select a destination if they walk. An integrated decision process needs to be defined in the MITO/MoPeD model. Moreover, the mode choice and destination choice factors of MITO and MoPeD need to be integrated to meet the needs for pedestrian modeling. Therefore, to develop an integrated model suite, the following adaptions were implemented:

- appropriate zone systems for measuring built environment that is relevant to pedestrian and modeling pedestrian behaviors.
- built environment factors in the mode choice models and the destination choice models for different trip purposes; and
- integrated trip decision processes to establish the interfaces between MITO and MoPeD.

Figure 29 gives an overview of the integrated trip decision process linking MITO and MoPeD. The framework essentially follows the paradigm of the traditional four-step model.

Given that the locations of work and education trips were predefined in the synthetic population, those trips are modeled first. The number of work and education trips is modeled in MITO. Then, work and education trips are fed into MoPeD to make the walk mode decisions. Work and school locations are stored in the synthetic population. Hence, no destination choice is necessary here for work and education trips. After that, the modes of work and education trips that are not made by walking are modeled in MITO.

Following the work and education trips, other trips are generated, which takes the number of work and education trips as one of the independent variables into account. HBS, HBR and HBO trips are simulated before the non-home-based ones (NHBW and NHBO). This is because we assume that the trip origins of non-home-based trips are influenced by the destination end of home-based trips. Similar to the work and education trips, those trips are first sent to MoPeD to select walking or not walking. The destination of walk trips is subsequently selected in MoPeD. Then, non-walk trips are fed back to MITO for TAZ destination choice and non-walk mode choice.

In the end, the selection of the preferred arrival time is modeled in MITO for all trips. The resulting trips are passed on to MATSim for trip assignment on the networks of different modes. Pedestrian flows are simulated in MATSim using the shortest path algorithm. We recognize that pedestrians may use other decision criteria besides trip distance (e.g., safety, comfort, scenery); however, data about network qualities were not available.

Note that this modeling suite does not intend to simulate detailed paths of pedestrians. While the assignment selects links chosen by the pedestrians, the location on a link is not modeled here. This level of detail is represented by a large amount of research work on pedestrian microsimulation models (Borrmann et al., 2012; Erdmann & Krajzewicz, 2015; Kielar & Borrmann, 2016). At the level of a metropolitan area, we simplify this step and focus on mode choice and destination choice instead.

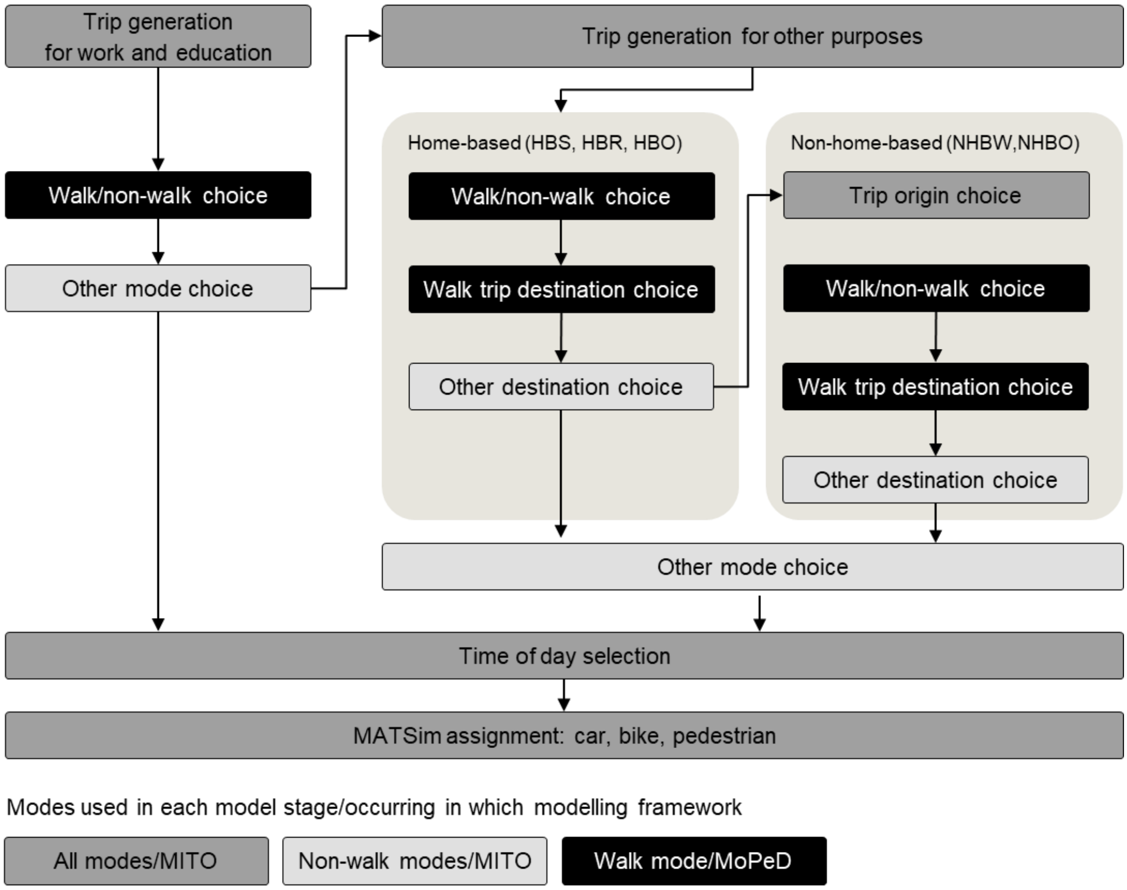


Figure 29 The travel demand modeling process of the MITO/MoPeD model

6.2. Model Setup for Munich

6.2.1. Define Zonal System

The Munich Model uses 4,953 gradually sized zones as its spatial unit (as shown in Figure 30). The sizes of zones vary from 200 meters to several kilometers. Coarser scales that are larger than 800 m might cause high errors in predicting walk behavior (Zhang, Clifton, & Moeckel, 2019). Thus, a finer and uniform zone system is needed for the MITO/MoPeD model.

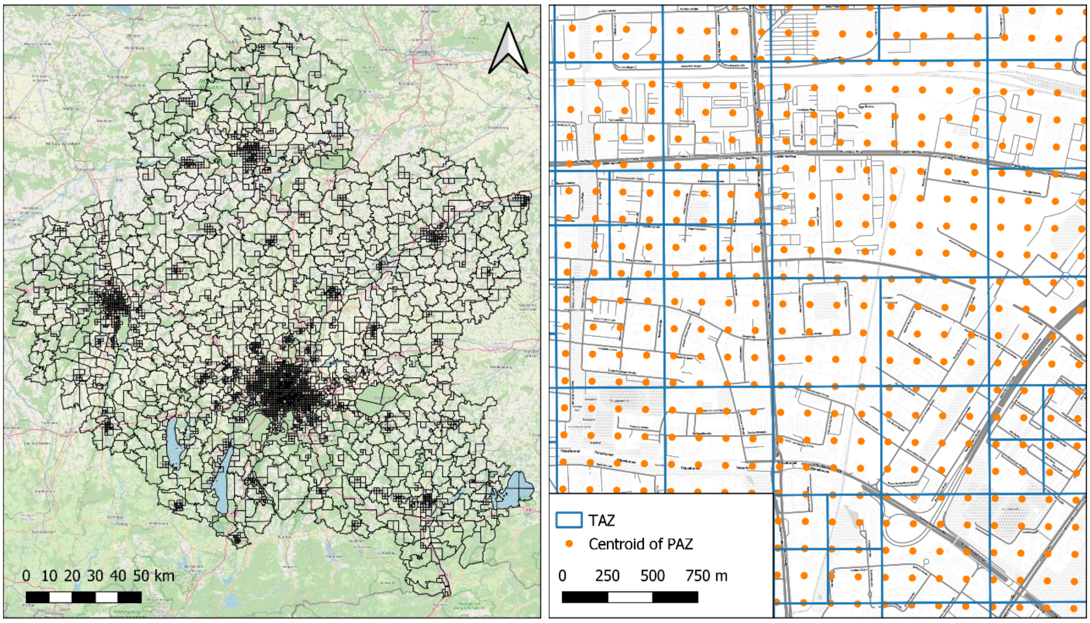


Figure 30 The Munich Metropolitan Area with TAZ zone systems (left) and the comparison of TAZs and PAZs (right)

Previous studies have pointed out that the finest spatial resolution may not be the appropriate scale and the spatial resolution selection is a trade-off among model performance, efficiency, and the availability of spatial data needed to implement (Zhang, Clifton, & Moeckel, 2019). In this study, we define the PAZ scale for the Munich region as 100m×100m raster. There are two main reasons for choosing this PAZ scale. First, the spatial units chosen to model the pedestrian behavior for this research depend on the relevant data available for Munich. Land cover data for the Munich region is provided as a 100m×100m raster. More importantly, the appropriate spatial resolution highly depends on the type of application. In this research, we aim to assess the unit of physical activity (PA). Health studies concluded that there is a non-linear dose-response relationship for PA (J. Woodcock et al., 2011). This means short walk trips might have large impacts on individuals with low PA. Because of this non-linearity, short walks need to be better captured in the MITO/MoPeD model, which further determines the necessity of using a finer spatial resolution. With such a small PAZ scale, the study area of the Munich region is covered by approximately 2,000,000 PAZ equivalents. A comparison of TAZs and PAZs can be found in Figure 30.

As the trip origin and destination locations in MITO are simulated at the micro-location level (in x/y coordinate), it is flexible to aggregate locations to any spatial resolution. The PAZ zone system is utilized in MoPeD for walk mode choice. In the walk trip distribution stage, destination choice is first conducted at SuperPAZs (400-meter grid cells). Then, trips are allocated from the selected SuperPAZ to the constituent PAZs. Trips that are not made by walking are handled in MITO at the TAZ zonal structure.

6.2.2. Create Pedestrian Street Network

The pedestrian street network in this study is obtained from OpenStreetMap. The links in OpenStreetMap are categorized into several road types with a certain hierarchy order as shown in Table 10. In the Munich Model, the network links are built until the level of residential road, which covers the important ways for vehicles. For the integrated model, pedestrian-centric roads such as the category “living street” and “pedestrian” are considered. This helps to boost the completeness of the road networks for pedestrian simulation. Also, we can get a better measurement of pedestrian network connectivity, which is an important feature for pedestrian mode and destination choice. Although service and track roads are also mostly used by non-motorized users, they are not included for this study. There are some reasons for dropping service and track roads. First, service roads are mainly connection links between real estate buildings and road accesses/exits. In MATSim, the access/exit road distance from activity coordinates to road node is measured using Euclidean distance. Because most of the service roads are very short, the Euclidean distance is close to the real network distance. Then, track links defined in OpenStreetMap are roads mostly located in rural areas for agricultural or forestry uses. The Munich metropolitan area is mainly covered by urban land use, so the amount of track roads is not significant in the study area. Lastly, dropping service and track links can help to avoid a boost in computational burden at trip assignment stage.

Table 10 Link types defined in OpenStreetMap (OpenStreetMap Wiki, 2020)

Hierarchy	Road type
1	Motorway
2	Trunk
3	Primary/Primary link
4	Secondary/Secondary link
5	Tertiary/Tertiary link
6	Residential road
7	Living street
8	Pedestrian
9	Service
10	Track

6.2.3. Calculate Built Environment Factors

Walking behavior is highly correlated to built environment variables. The Munich Model considers a few built environment variables in the walk mode choice and walk destination choice model. For example, area type dummies are used primarily to differentiate land use at an aggregated level. To model the effects of the built environment on pedestrian travel behavior, the measurement of pedestrian accessibility is included in the MITO/MoPeD model. Pedestrian accessibility is defined as population and non-industrial jobs within an 800-meter network distance. Pedestrian accessibility data in the Munich region has been measured at the PAZ level. First, the isochrones, also known as pedestrian catchment areas, are generated based on the pedestrian network from OpenStreetMap. Afterward, we calculated the total number of non-industrial jobs and population that locate within each isochrone. Results of pedestrian accessibility for the Munich region are shown in Figure 31.

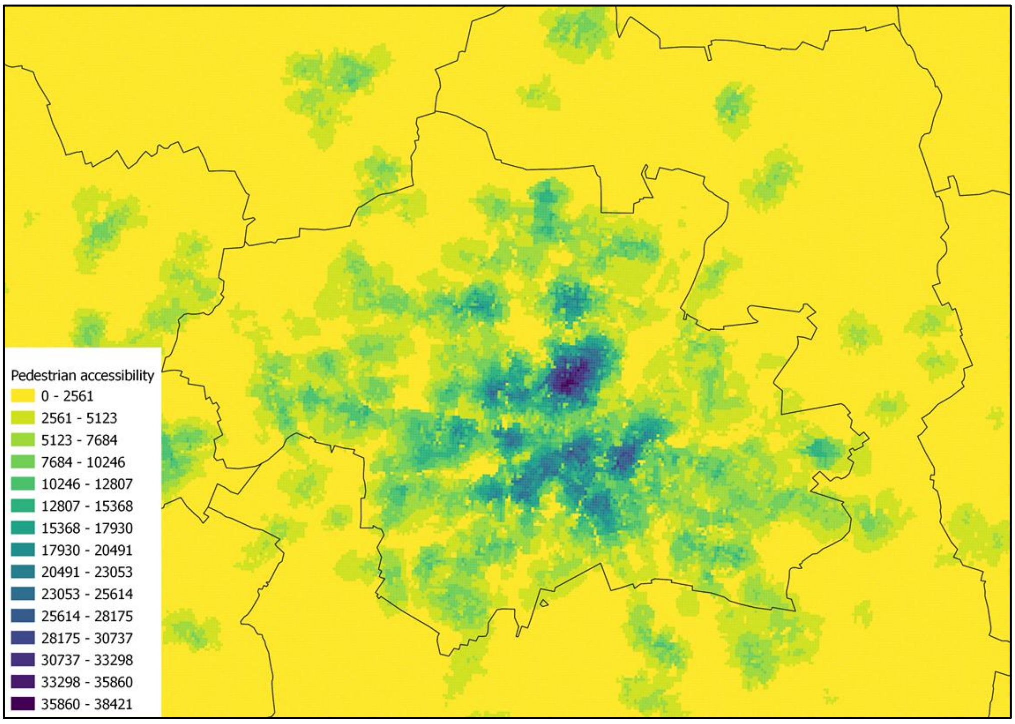


Figure 31 Pedestrian accessibility in the Munich city area

6.3. Mode Choice Model Calibration

Due to the lack of geographic information about the trips in the German national household travel survey, model parameters cannot be re-estimated using local surveys. Therefore, the model estimates used in the MITO/MoPeD model were developed in the context of the Portland, Oregon metropolitan area in the United States while they are applied to the Munich study area in Germany.

First, the binary walk mode choice models of home-based purposes were re-estimated by using the OHAS data. The estimation models in MoPeD 2.0 did not include distance as a predictor for HBW and HBE trips because the mode choice was made before the destination choice. Benefits from agent-based nature of MITO, work and education location was known as an input information from synthetic population. As a result, the models were re-estimated by separating HBW and HBE purposes and adding distance factors to each. Table 11 presents the estimation results of HBW, HBE and other HB purposes. Distance is a strong predictor in mode choice. By adding distance factor in the models, the model goodness-of-fit improved compared to those in Table 2. The rest of the predictors remained the similar magnitude as shown in Table 2.

Table 11 Binary walk mode choice model estimation of home-based purposes in the integrated model

	HBW			HBE			Other HB purposes		
	Estimate	Pr(>|z|)		Estimate	Pr(>|z|)		Estimate	Pr(>|z|)	
(intercept)	-7.425	0.000	***	-2.178	0.000	***	-7.276	0.000	***
Distance to work/school (km)	-0.422	0.000	***	-1.371	0.000	***			
Number of vehicle (0)							0.945	0.000	***
Number of vehicle (2)							-0.239	0.002	**
Number of vehicle (2+)							-0.408	0.000	***
Vehicle (Yes)	-0.491	0.127		-0.445	0.147				
Number of children (1)				1.637	0.000	***			
Number of children (2)				1.626	0.000	***			
Number of children (2+)				1.268	0.000	***			
Child (Yes)	0.745	0.000	***				0.184	0.000	***
log(pedestrian accessibility)	0.778	0.000	***	0.208	0.000	***	0.727	0.000	***
HBRecreation							base		
HBShop							-0.585	0.000	***
HBOther							-0.555	0.000	***
Log-Likelihood:	-334			-758			-4406		
McFadden R^2:	0.270			0.322			0.129		

Then, the binary mode choice models for home-based purposes (shown in Table 11) and the models for non-home-based purposes (shown in Table 2) were transferred from the Portland context to the Munich context. Previous works have focused on the spatial transferability of travel forecasting models (Agyemang-Duah & Hall, 1997; Cotrus et al., 2005; Everett, 2009; Huntsinger & Rouphail, 2013; Sikder et al., 2013). They showed the ability to transfer models between different regions and suggested the methods used to enhance model transferability. Updating constants/adding new constants is a widely used method in practice to enhance the model suitability.

To make a fair comparison between the two modeling frameworks, a mode choice calibration process was implemented in the MITO/MoPeD model. In this research, we assumed that the parameters other than the constants are transferable in two contexts. Additional constants were introduced into the mode choice models to scale up the average walk shares to match the observed shares. Table 12 shows the final calibration factors implemented in the models by trip purposes. These calibration factors represent the difference between the influences of unobserved factors (e.g., mobility culture, geography, and weather) in Portland and Munich.

Table 12 Calibration factors of walk mode choice models by trip purposes

Purpose	Calibration factor
HBW	0.919
HBE	1.369
HBS	0.784
HBR	0.708
HBO	0.612
NHBW	0.781
NHBO	0.776

Figure 32 compares the simulated walk shares by trip purposes to the observed walk shares taken from Germany household survey data 2017 - Mobilität in Deutschland (MiD). Before calibration, the walk mode choice models in the MITO/MoPeD model underestimated the shares of walk trips across all purposes. Although there are deviations in the absolute values, the relative relationships among purposes are consistent with the observed data. For example, recreational trips (HBR) have the highest shares of walking while work-related trips (HBW and NHBW) have lower walk shares.

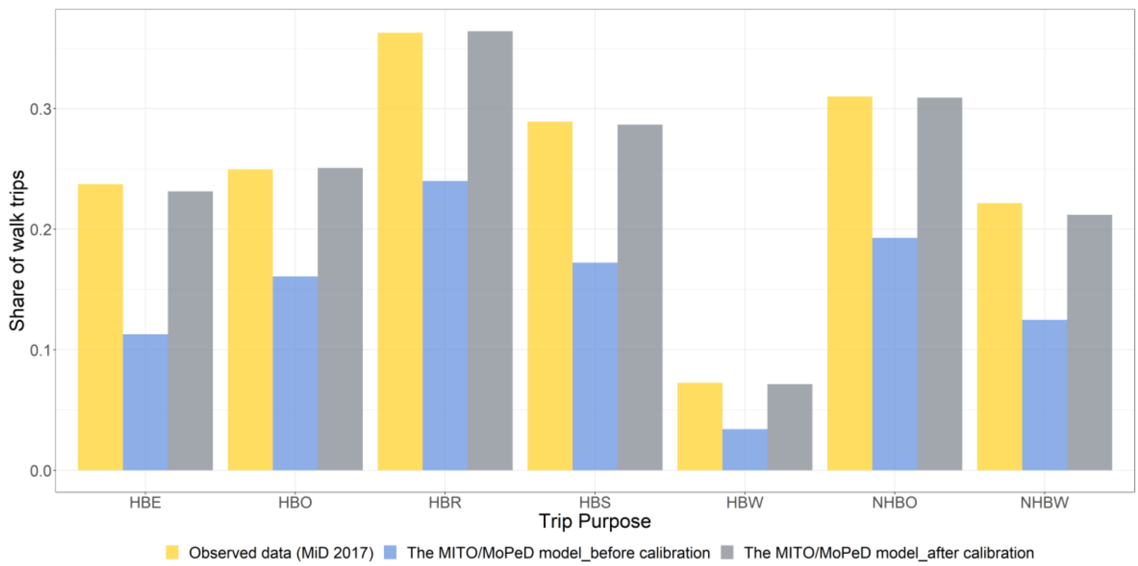


Figure 32 Observed walk shares of MiD and predicted walk shares resulting from the MITO/MoPeD model by trip purposes.

6.4. Model Results

To answer the research questions mentioned above and to test the plausibility of the MITO/MoPeD model, this section evaluates how effectively the MITO/MoPeD model structure improves the estimation of transport outcomes and physical activity volumes. The Munich model and the MITO/MoPeD model are applied to the Munich metropolitan area. Model performances are analyzed based on average walk shares, the spatial distribution of walk shares, mean trip length, trip length distribution, pedestrian flows, and physical activity volumes.

6.4.1. Walk Mode Shares

The share of walk trips is one of the key outcomes for investigating policies and strategies to encourage walking.

After calibration, both models can simulate the average walk shares accurately. Besides the mean value, the spatial distribution of walk trips is also an important outcome to evaluate the model performance. Figure 33 presents the spatial distribution of walk shares across the Munich metropolitan area resulting from two modeling frameworks. From an overall perspective, both models demonstrate similar patterns that walk shares in the urban areas are higher than those in the suburban or rural areas. Nevertheless, we note that the distribution in the Munich Model (Figure 33a) has a boundary issue. The border areas sometimes show high walk shares which are at the same level as the urban centers.

There are three major reasons why these border areas have higher walk shares than we expected. First, trip distance is a key factor of the mode selection in the Munich Model. Shorter trips tend to use the walk mode. The border areas have limited destination alternatives and their surrounding areas are mostly rural and less attractive. Thus, trips generated in those areas are more likely to select a closer destination with a short distance which leads to a high share of walk trips. The second reason may be due to the coarse zone system used in the Munich Model. Larger zones result in a greater share of intrazonal trips. As intrazonal trips have the same trip length, which is relatively short, having larger zones causes an overestimation of walk trips. Lastly, built environment variables like population and employment density are not considered when selecting modes in the Munich Model. This means short trips generated in the less-populated areas have the same likelihood of choosing walking as those in the urban areas.

Benefit from the fine spatial resolution used in the MITO/MoPeD model, trips are modeled with more precise network distances rather than being considered as same length intrazonal trips. In addition, by introducing the activity density into the walk mode choice models, the MITO/MoPeD model can better capture the differences in walk shares between urban areas and rural areas. As shown in Figure 33b, the MITO/MoPeD model mitigates the issue of walk share overestimation at the border areas.

To have a closer look into the walk share distributions in the Munich city area. The results of the Munich Model (Figure 33c) show that walk trips are sprawled in the entire Munich city area though we can see a decreasing trend in the outer areas. Given the coarse spatial resolution, it is difficult to observe the areas for pedestrian demand in the Munich city area. Nevertheless, the MITO/MoPeD model can give us a clearer picture of the hotspots for walk trips as shown in Figure 33d.

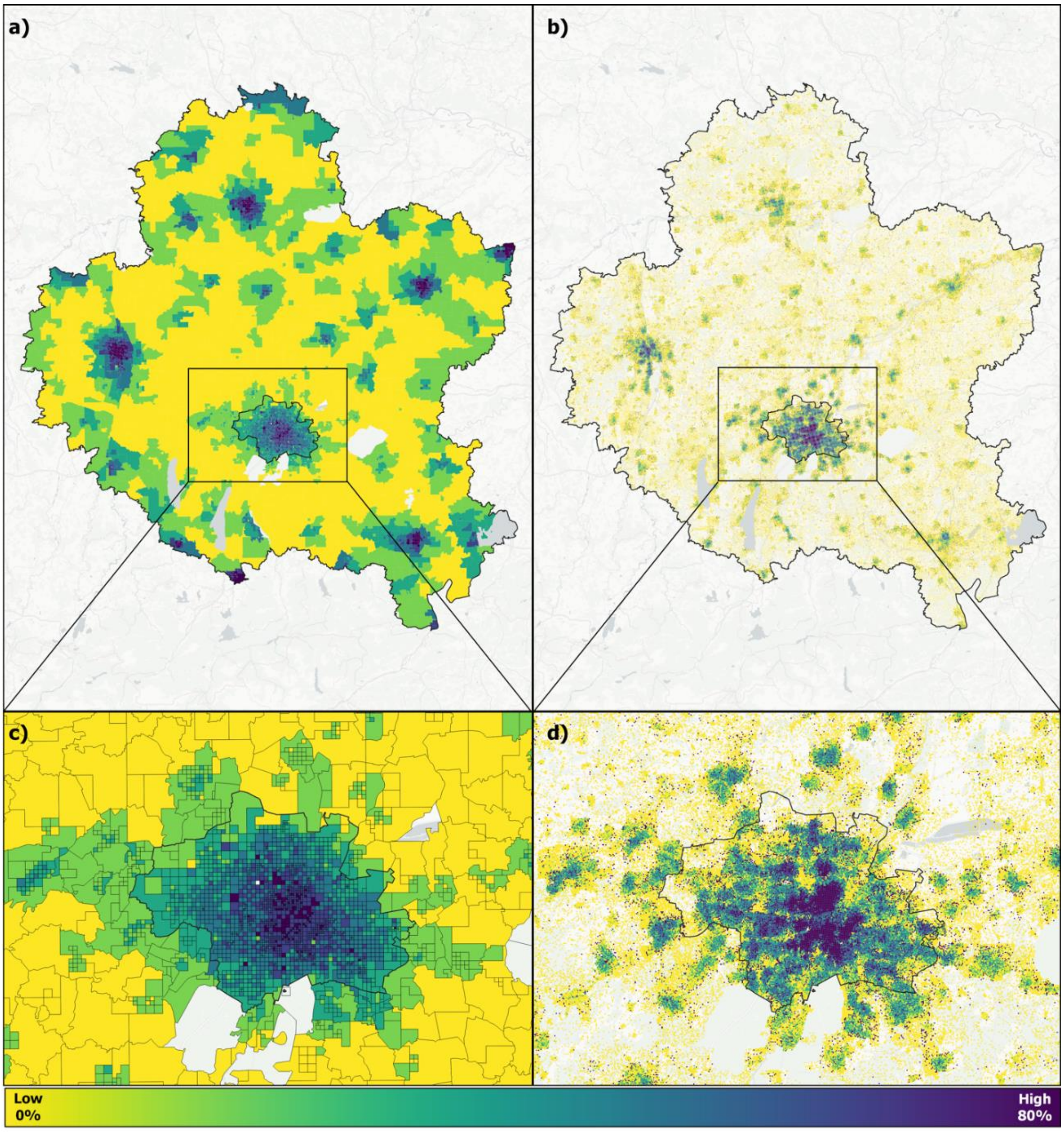


Figure 33 Shares of walk trips resulting from the Munich Model across Munich metropolitan area (a) and in Munich city (c); Shares of walk trips resulting from the MITO/MoPeD model across Munich metropolitan area (b) and in Munich city (d)

6.4.2. Trip Length Distribution

Previous studies have found that destination choice models are the largest source of errors in travel demand models (Zhao & Kockelman, 2002). Trip length distributions can be used to evaluate the performance of destination choice models. Figure 34a shows the trip length distribution of all trips. Compared to the observed data, the Munich Model has a significant deviation for the short distance trips. Figure 34b presents the distribution of walk trips. It indicates that the Munich Model overestimated the trip lengths for walk trips and then further confirms that the Munich Model is poor at capturing pedestrian travel behaviors. Nevertheless, both figures indicate that the MITO/MoPeD model performed better in modeling walk trip lengths. Without any calibration, the shape of the walk trip distribution of the MITO/MoPeD model is almost matched to the observed data. For the MITO/MoPeD model, we find that the left side of the distribution and the peak are close to the observed data, but there is a large difference on the right side of the distribution. This means that the MITO/MoPeD model underestimates long-distance walk trips.

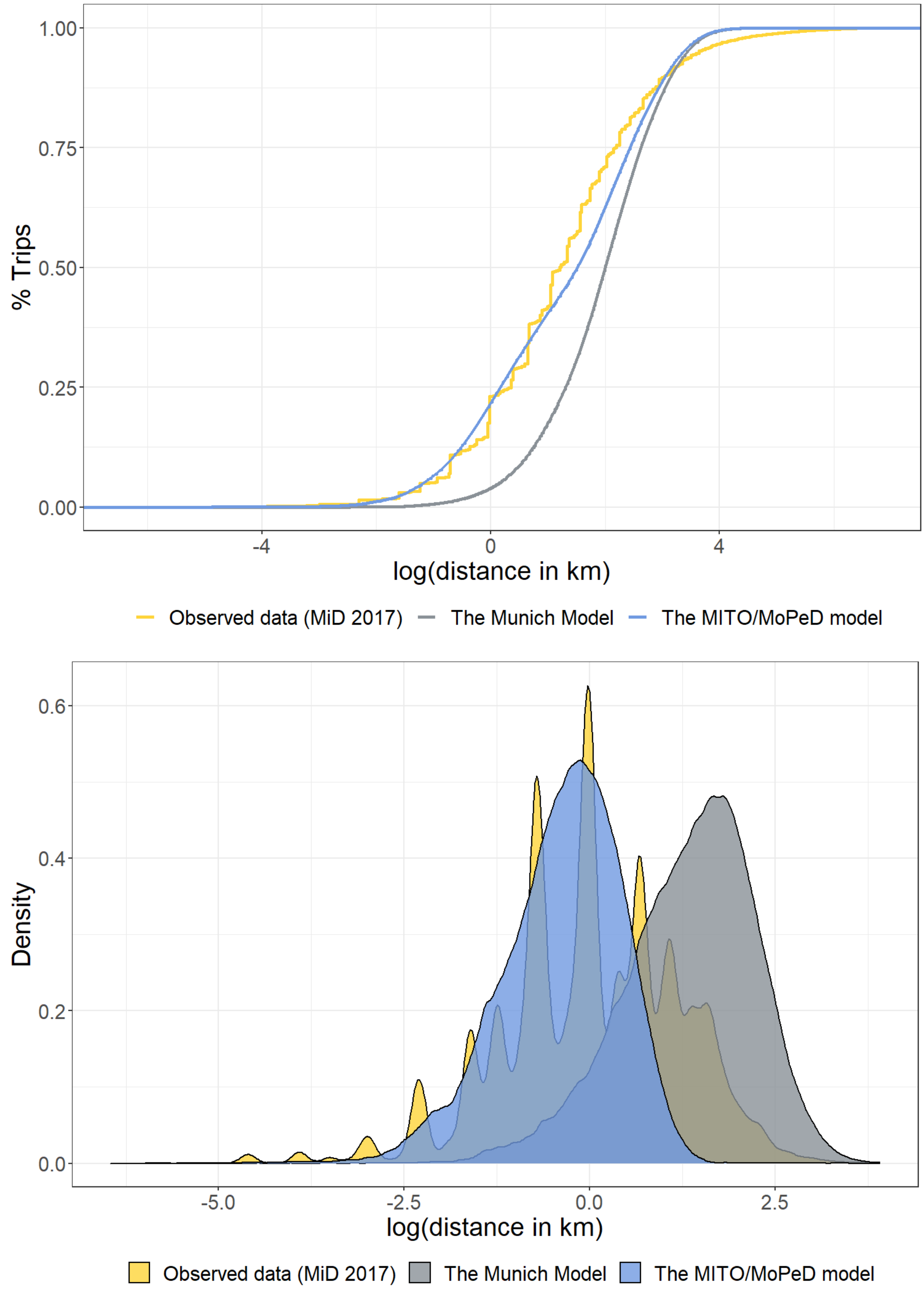


Figure 34 Comparison of trip length distribution of all trips (a) and comparison of trip length distribution of walk trips (b)

6.4.3. Pedestrian Flows

After walk trips were generated, they were then passed on to MATSim for route assignment. In this study, the shortest path algorithm is used for assigning walk trips. In future studies, pedestrian route choice behavior needs to be further investigated and more advanced routing techniques need to be applied to get more precise route estimates.

Figure 35 compares pedestrian volumes of network links resulting from two models. We can find that pedestrian flows are sprawled in the Munich Model and pedestrian volumes are much higher than those in the MITO/MoPeD model. The lack of count data is a barrier to model validation. Thus, in this study, we cannot make a conclusion about the model performance based on pedestrian flow maps. However, we hypothesize that the Munich Model overestimated the pedestrian volumes on the links. As discussed above, the Munich model was poor at simulating walk trip lengths. The overestimation of walk trip lengths results in an overestimation of pedestrian kilometers traveled in the trip assignment stage. The total pedestrian kilometer traveled is 239,539 km in the MITO/MoPeD model while the number is 738,599 km in the Munich Model which is almost tripled. The high value of pedestrian kilometers traveled in the Munich Model may be the reason for high pedestrian volumes on network links. The results from the MITO/MoPeD model show a more reasonable pattern in that the network in the city center has higher pedestrian volumes with a diminishing trend outward.

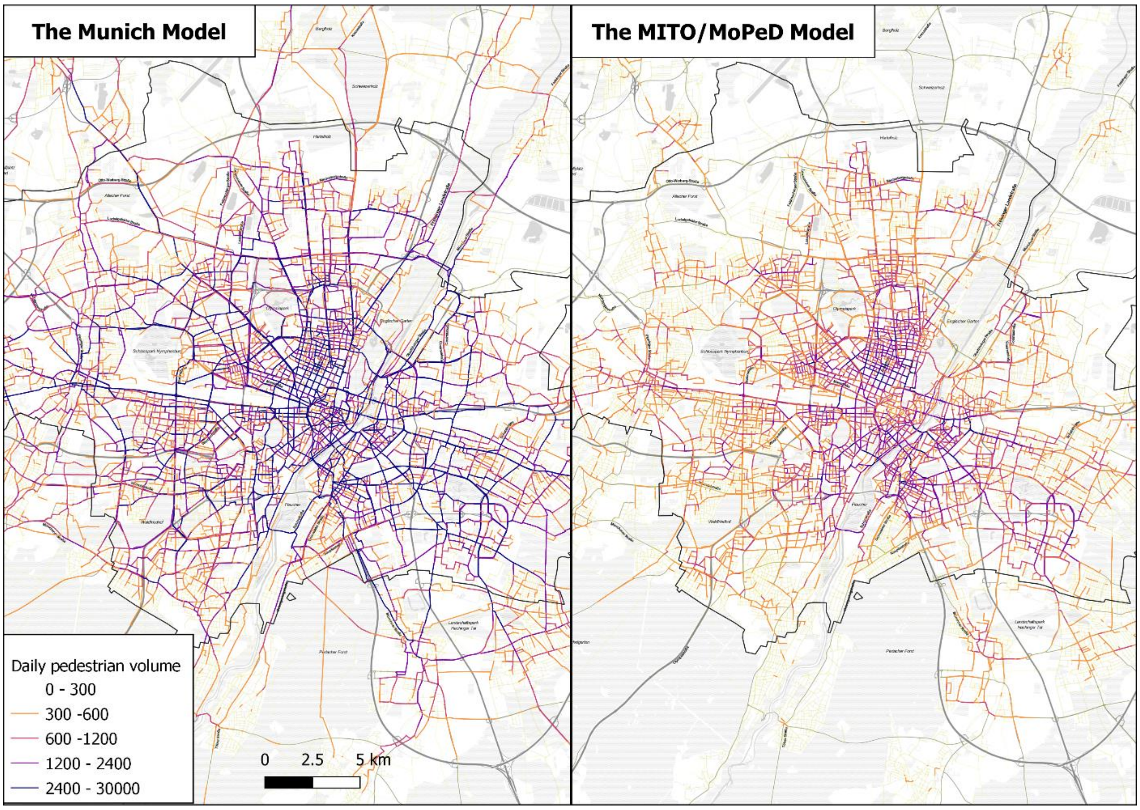


Figure 35 Comparison of pedestrian volumes assigned to the network links

6.4.4. Volume of Physical Activity

PA volume is one of the critical inputs to measure an individual's health indicators. PA commonly has four domains including occupational, domestic, leisure, and transportation. In this study, we focus on PA in the transportation domain. For this, information on all walking and cycling activities is usually included to measure PA volume (James Woodcock et al., 2013). In this study, modeling components of bicycle trips are unchanged so that the cycling activities retain the same in the base model and the MITO/MoPeD model. To avoid confounding comparisons, this study only measures the walk trips for physical activity volume.

Physical activity volume is the product of frequency, duration, and intensity. Commonly, it is measured over a week-long period with the unit of mMET-hours per week (James Woodcock et al., 2013). In this study, the transport models only simulate the travel demand on a typical day, so the walk trips are multiplied by a factor of 7. Walk trip durations are measured using the walk trip distances resulting from MATSim. Trip distances are converted to durations using an average speed. Ideally, speeds could be differentiated by age and sex to measure PA volume more precisely. However, for this study, we try to skim off confounding factors and investigate a pure comparison between the base model and the MITO/MoPeD model. Therefore, an average speed of 4.8 km/h is chosen for walk trips (Kahlmeier et al., 2014). Similarly, a single value of intensity is used for all walk trips. This is 3.61 mMETs for walking which is the median intensities obtained from a recent objective study on commuters by Costa et. al. (2015). Thus, the PA volume of individual i is measured by the following equation:

$$PA_i = 7 * \sum_{n=1}^{walk\ trips} \frac{walk\ distance}{4.8} * 3.61$$

Figure 36 compares the distribution of PA volume resulting from two modeling frameworks. The observed PA distribution based on MiD 2017 is considered as the reference. In this plot, the proportions of inactive individuals (the values at x = 0) are very high. This is because we only measured walk trips and it is incomplete for assessing total PA. The results first confirm the hypothesis that the Munich Model is poor at assessing PA volumes. The fundamental issue is that the Munich Model overestimates the length of walk trips. As a result, the Munich Model overestimates the PA volumes of each individual.

As discussed before, the MITO/MoPeD model has a better performance in capturing the distribution of walk trip lengths, so we didn't see the issue of overestimating PA volumes in the MITO/MoPeD model. However, the plot reveals that the MITO/MoPeD model generally underestimates PA volumes.

There are three major reasons for the underestimation of PA volumes in the MITO/MoPeD model. First, as discussed above, the MITO/MoPeD model is poor at modeling long and uncommon walk trips while those long walk trips are recorded in MiD data. Missing these long walk trips causes the underestimation of PA volumes. Another reason is that access and egress trips to public transport are not modeled in the transport model, resulting in the elimination of a small share of PA. Lastly, the observed data may be biased by walk distances since the distances in the survey were self-reported.

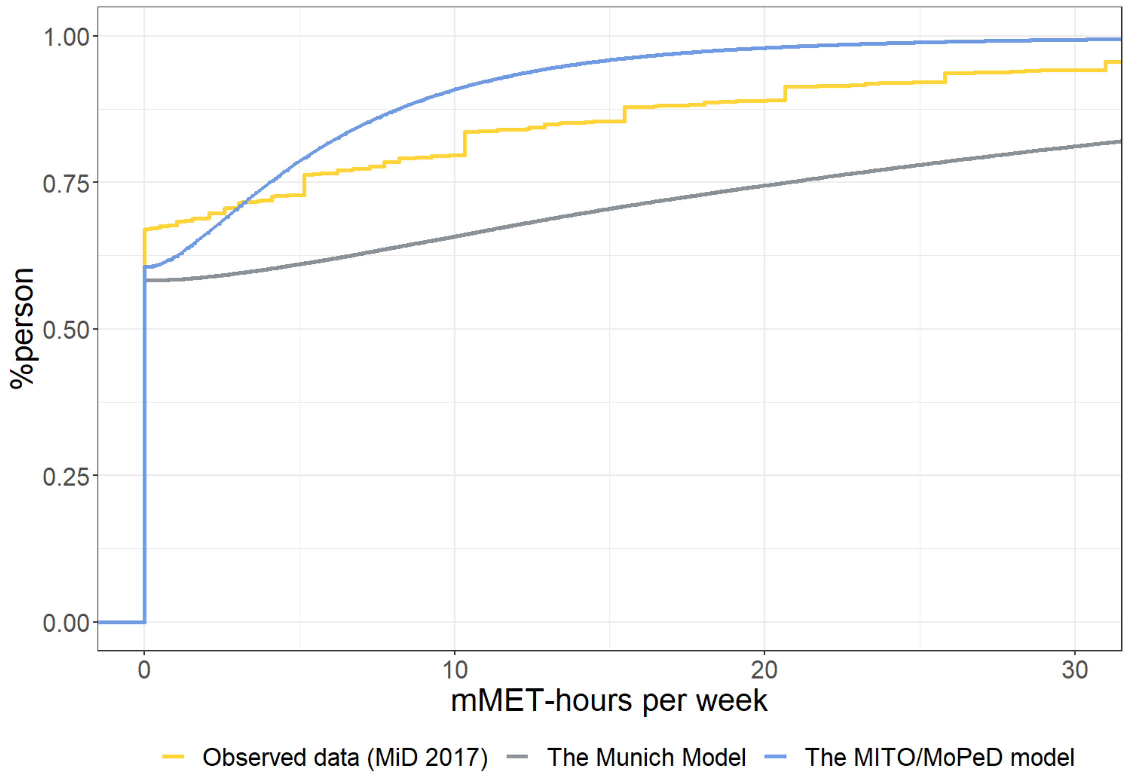


Figure 36 Comparison of physical activity volume distribution

6.5. Discussion on the Integrated Model

In this chapter, we advanced the state of modeling pedestrian behaviors by integrating MoPeD into the agent-based transport model MITO. This integrated modeling framework builds a link between advanced pedestrian modeling techniques with the agent-based modeling approach. More importantly, the results proved that the integrated model can better simulate travel outcomes than the Munich Model.

However, the integrated model still faces some limitations that should be addressed in future work. First, the access and egress walk trips are neglected in the integrated model. To simulate physical activity volumes more precisely, those trips navigating to public transport need to be considered in the model.

Second, there are limitations in the physical activity volumes calculation methodology because it has been heavily simplified. For comparison of model results, a simplified calculation methodology can help us to eliminate confounding factors. However, for a more realistic health assessment, future research could introduce more precise PA intensities using available information on the gradient, speed, age, and sex.

Furthermore, though having pedestrian route choices in the MITO/MoPeD model is an improvement, more advanced algorithms need to be considered in future studies. For example, route choice decision factors such as safety, comfort, and scenery need to be investigated and applied.

Moreover, to transfer model estimates from Portland to the Munich context, we applied the simple transfer method by updating constants. This approach helps us to match the mode shares at an aggregate level, but it does not capture the differences in the magnitude of variation in observed and unobserved factors. Therefore, more advanced approaches can be applied in the future to enhance model transferability.

Lastly, the integrated model is sensitive to spatial attributes but not sensitive to temporal attributes. The pedestrian travel behavior variability is not captured in the model. This can lead to some issues in model accuracy and model sensitivity. For example, the physical activity volume is simply measured by multiplying walk time/distances by 7 to represent the weekly physical activity. For the comparison purpose in this chapter, the simplified measurement is not an issue. However, imputing weekly walk trips from single-day behavior will cause some issues when using travel outcomes for health benefits assessment. Therefore, the next crucial task for pedestrian modeling is to understand weekly travel behavior and the variability of walk trips.

7. Application of the Integrated Model to scenarios in the Munich Metropolitan Area

In this chapter, three policy scenarios were developed to assess the impact of radical strategies on walking behavior. The scenarios include 1) teleworking, which may become more common post-Covid; 2) excess commute, which can minimize people's commuting time; and 3) car-free city, which discourage personal vehicle trips. It is expected that the scenarios will encourage more walking over private car use, resulting in better health outcomes. All three scenarios were applied to the Munich Metropolitan area.

7.1. Teleworking Scenario

The scenario investigates the travel demand changes and health impacts of a scenario in which 80% of the workers in the administration, financial, and service sectors are required to work from home. The integrated model randomly selected 80% of teleworkers in affected job sectors and assigned them zero work trips in the work trip frequency model. For this scenario, travel demand changes only for workers in affected job sectors. Therefore, this section discusses only individuals in that group.

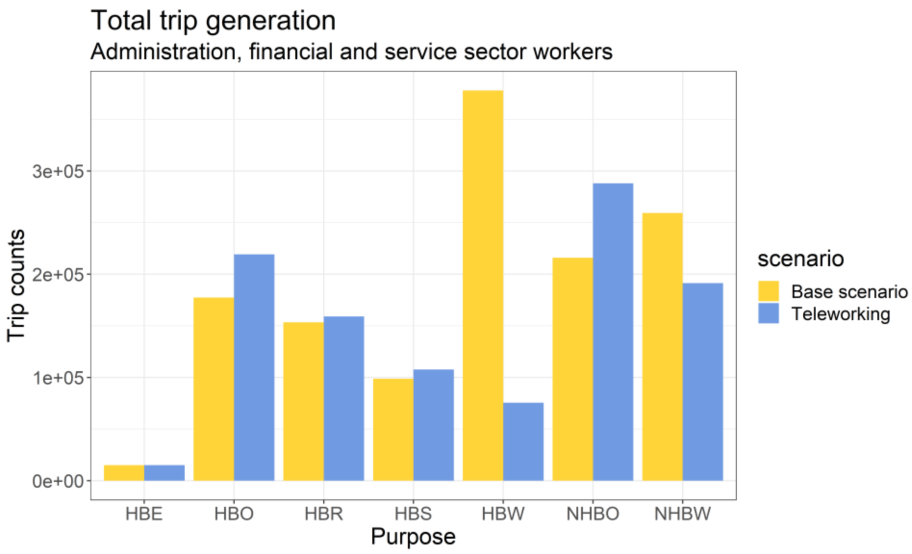


Figure 37 Total trip generation by purposes of the workers in affected job sectors

Figure 37 shows the number of trips generated by workers in affected job sectors. For mandatory purposes, there are exactly 80% fewer work trips as defined and no change in education trips. MITO incorporated travel time budgets (TTB) into demand generation, so the integrated model can simulate the compensatory behavior between mandatory and discretionary travel. Therefore, in this scenario, the reduction in mandatory trips is compensated by an increase in discretionary trips for all purposes except non-home-based work. NHBW trips are usually lunch trips during working time or leisure trips before or after work, so the teleworking scenario also cuts down the number of NHBW trips. Overall, there is an 18.6% decrease in the total number of trips made by workers in the affected sectors.

Figure 38 presents the number of trips by different modes generated by workers in affected sectors. There is a decrease in auto drivers and public transport modes while an increase in auto passengers and almost no changes in walk mode. These changes are the result of different mode preferences for mandatory purposes versus discretionary purposes. First, motorized and public transport modes are commonly used for commuting trips. Therefore, the number of trips of these modes had a dramatic decrease in HBW and NHBW purposes. Active modes are more likely to be used for discretionary purposes. On the one hand, there was a drop in active commuting trips due to teleworking. On the other hand, there was an increase in walking and cycling trips for discretionary purposes. Therefore, the total number of walking trips showed no changes, and the total number of cycling trips showed a slight decrease. Auto-passenger mode is usually used for educational purposes. The telework scenario had no influence on education trips. The small increase in auto passenger trips was gained from the increase in discretionary trips.

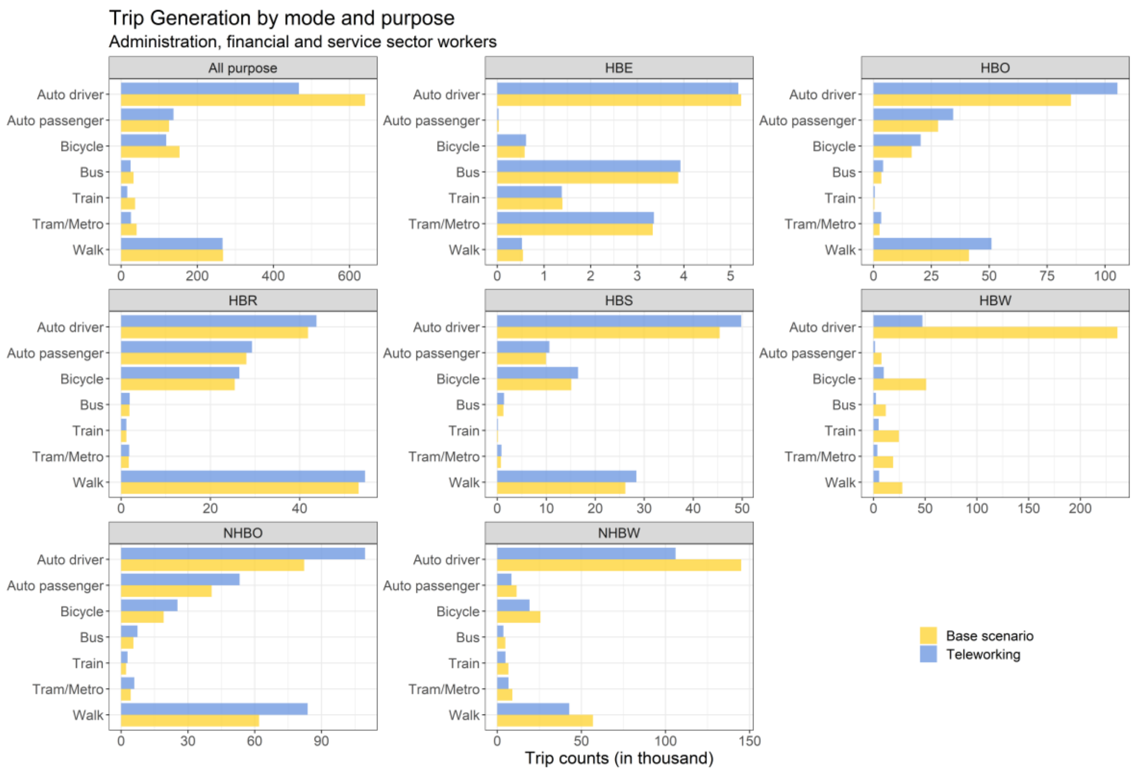


Figure 38 Trip generation by modes and purposes of the workers in affected job sectors

Figure 39 shows the mean daily distances travelled by workers in affected sectors, differentiated by mode. There is a decrease in kilometers travelled by car, train, tram/metro, and bicycle, but an increase in kilometers walked. These changes are related to different trip length preferences for mandatory purposes versus discretionary purposes. Teleworking resulted in more walk discretionary trips as shown in Figure 38. Those active trips for leisure purposes are more likely to have longer trip lengths, which leads to an increase in the distance walked.

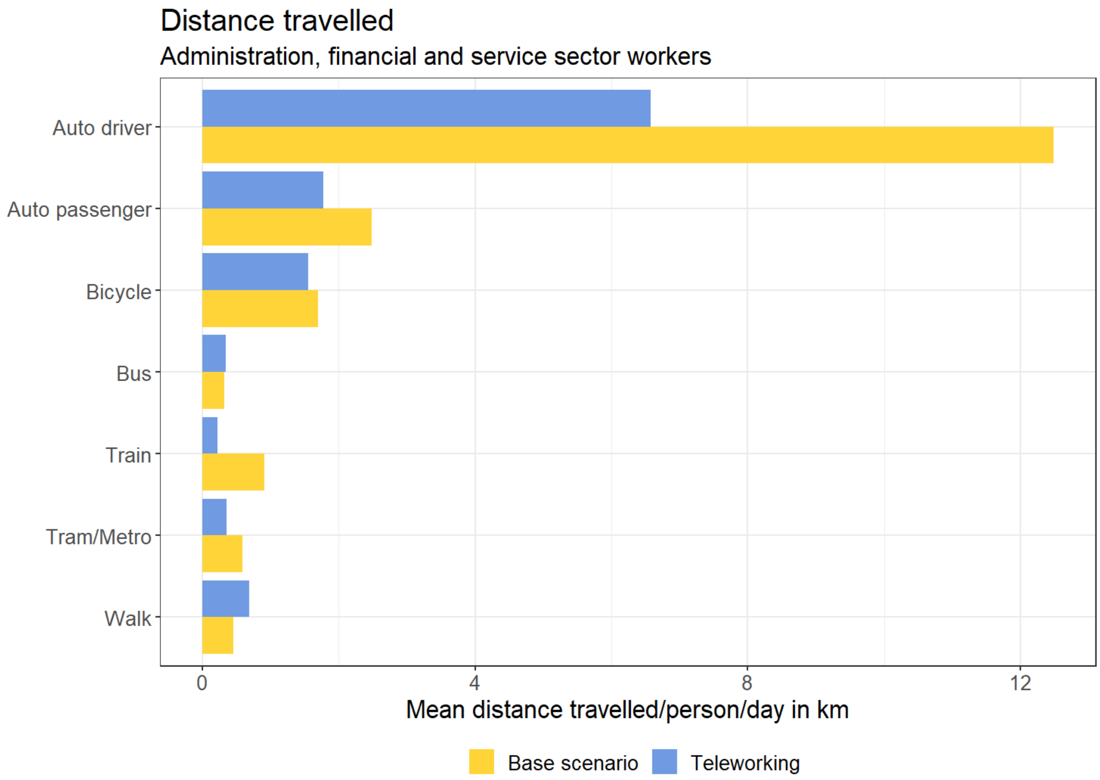


Figure 39 Distance travelled for each mode of the workers in affected job sectors

Finally, there is an overall increase in PA for workers in affected sectors. The PA volume calculation for walking trips is the same as the one presented in Section 6.4.4. The same function was used for cycling trips but with different parameters for average speed and PA intensity. An average speed of 13.9 km/h and a PA intensity of 5.44 mMETs is chosen for cycling (Kahlmeier et al., 2014, Costa et. al. 2015). For these individuals, there is an average 8% decrease in cycling PA in the teleworking scenario. This is offset by a 31% increase in walking. As shown in Figure 40, the increase in walking PA mainly happened to the individuals who have a high level of PA, which is mainly due to increased discretionary (particularly recreational) travel.

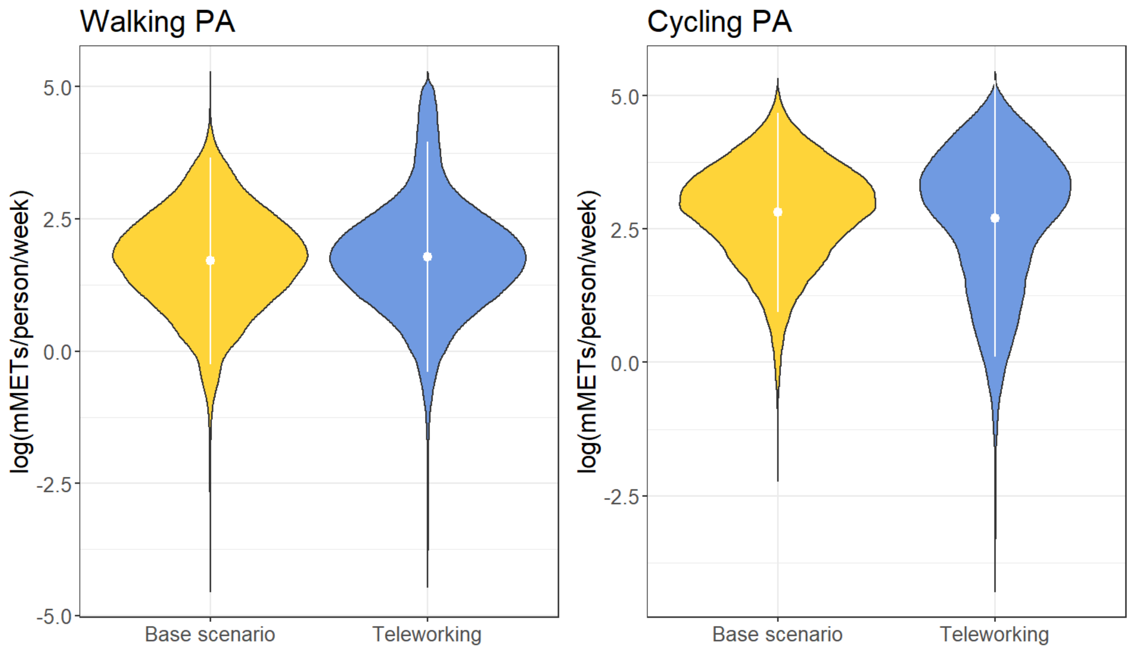


Figure 40 Walking and cycling physical activity volumes of the workers in affected job sectors

7.2. Excess Commute Scenario

Excess commute (EC) refers to the sub-optimal allocation of workers to jobs in an urban region in terms of minimal commute distances. This application investigates the changes of walking behavior under a radical scenario where jobs and workers are perfectly assigned to achieve minimal commuting time. This scenario analysis is based on the work done by the author and other colleagues. The method of estimating the optimized worker-job allocations can be found in the paper (Kuehnel et al., 2022). The actual mean commute time was about 12.4 minutes, and the mean commute time for the optimized job-worker allocation was about 3.3 minutes. In this scenario, travel demand changes only for commuters. The integrated model did not consider the intra-household interaction in travel behavior, so the travel behavior of non-commuters has no change in this scenario. Therefore, the following paragraphs discuss the changes in commuting trips.

Figure 41 compares the number of trips generated by commuters. The overall trip count increased by about 0.6% because of the increase in discretionary travel. This can be explained by the TTB theory. Commuters in this scenario had shorter commute time, allowing them to devote more time to other activities. Besides that, NHBW trips were reduced significantly. According to Staves (2020), this is because active commuters are more likely to take home-based discretionary trips rather than more complex non-home-based trip chains.

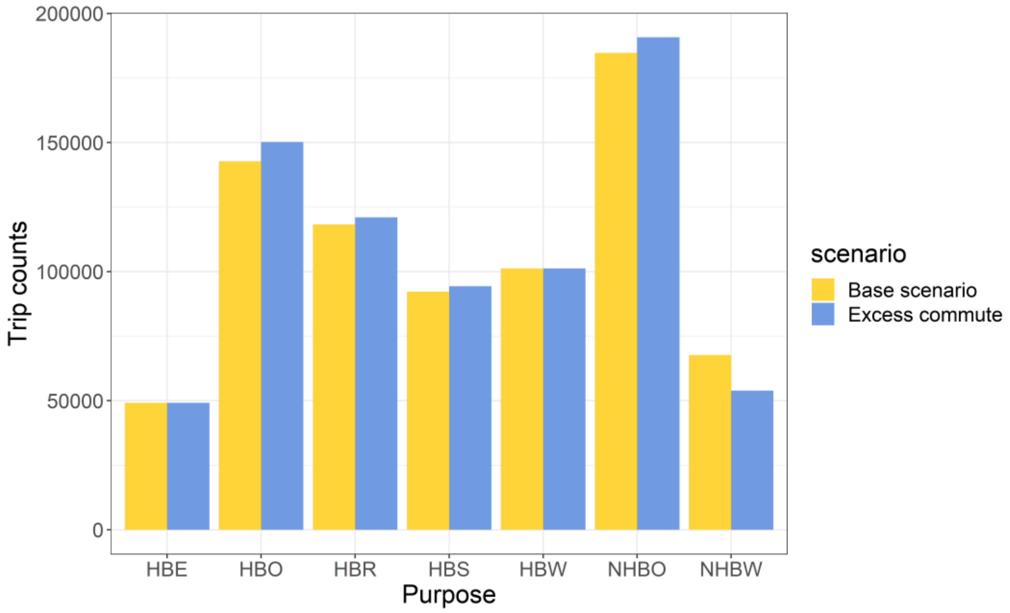


Figure 41 Total trip generation by purposes of the commuters

As shown in Figure 42, the modal split for commute trips did not change significantly, which is not intuitive. The hypothesis is that shorter commute distances lead to more use of active modes. However, the integrated model cannot capture the mode shift in the excess commute scenario. This is mainly due to the limitation of the walk mode choice model for HBW trips that the model poorly captures the relationship between distance and utility.

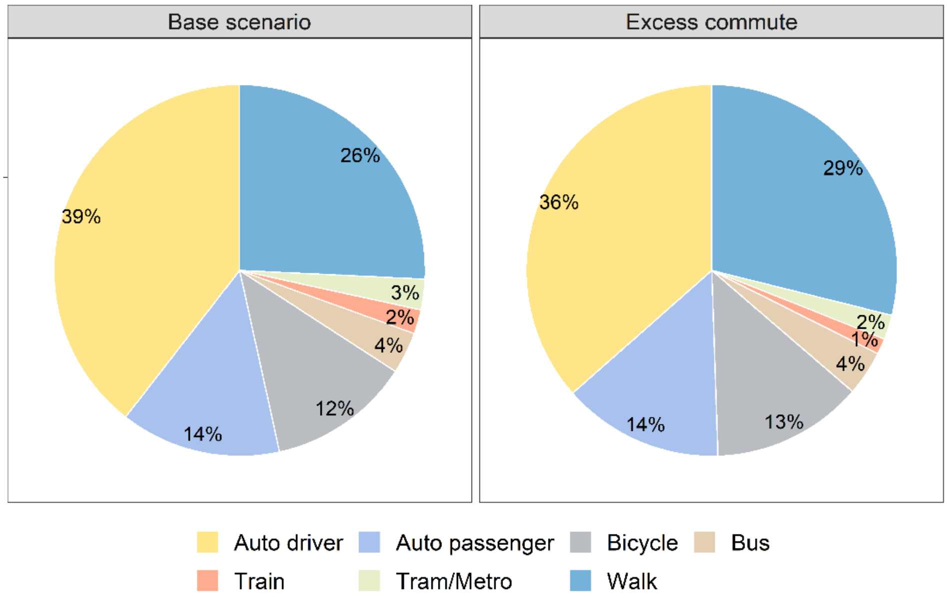


Figure 42 Mode split of commute trips.

7.3. Car-Free City Scenario

In this radical scenario, two policies were applied at the household level and the city level. First, all households with cars must reduce their number of registered vehicles by one. Then, car-free zones were implemented in the city center of Munich. Car-free zones are districts where motorized vehicles are prohibited, and people mainly rely on public transport, walking, or cycling. Since 2008, a low emission zone has been in effect in Munich (City of Munich, 2021). This scenario utilized the spatial context of the low-emission zone and fully converted it to a car-free zone (see Figure 43).

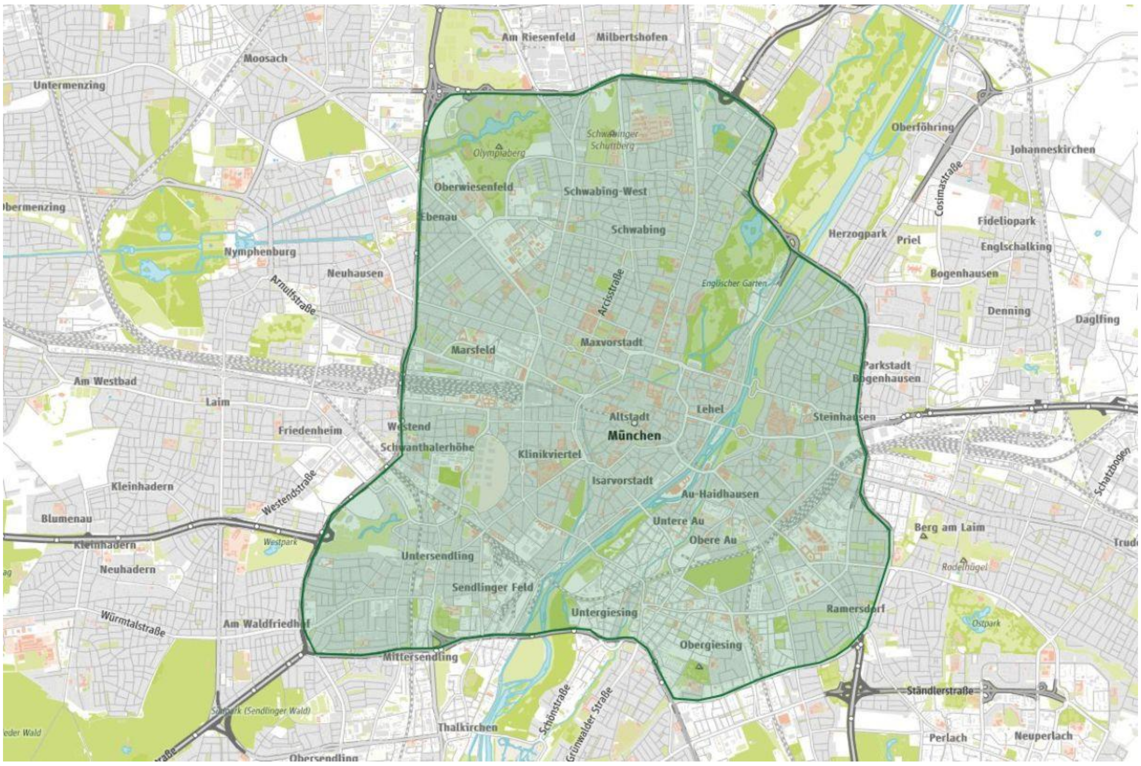


Figure 43 Munich Low Emission Zone (City of Munich, 2021)

Table 13 shows the number of trips generated by different modes. It is not surprising that there is a dramatic decrease in motorized vehicles, which is aligned with the scenario design. As car usage is restricted, those trips to/from the city center were shifted largely to active modes and public transport modes.

Table 13 Number of trips by modes in base and car-free city scenarios

Mode	Base scenario	Car-free city	Difference
Auto Driver	2,977,585	1,191,258	-60%
Auto Passenger	1,057,144	682,604	-35%
Bicycle	932,118	1,438,343	54%
Bus	270,377	561,672	108%
Train	158,793	388,008	144%
Tram/Metro	201,016	461,418	130%
Walk	1,949,765	2,768,980	42%

Figure 44 gives a clear picture of mode shifts for each purpose. For mandatory purposes (HBW and HBE), trips were shifted from motorized modes to bicycle and public transport, while there was no significant change in walk mode shares. For discretionary purposes, walking became the dominant mode for recreational trips. Both walking and cycling had a high share in shopping purposes.

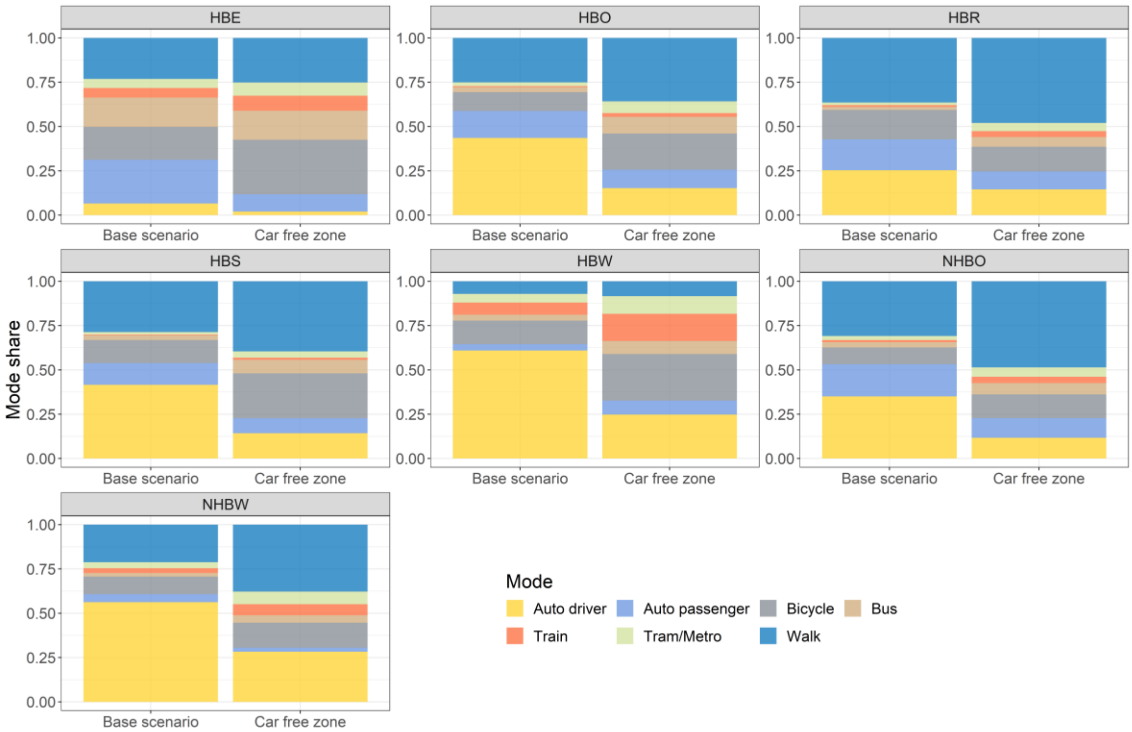


Figure 44 Mode shares in base and car-free city scenarios

Figure 45 illustrates the mean distances travelled by different modes. In the car-free city scenario, individuals relied on public transport, bicycle, and walk modes, resulting in a significant increase in distance travelled by these modes. The average daily walking distance increased from 445 meters to 622 meters, while the average daily cycling distance was almost doubled.

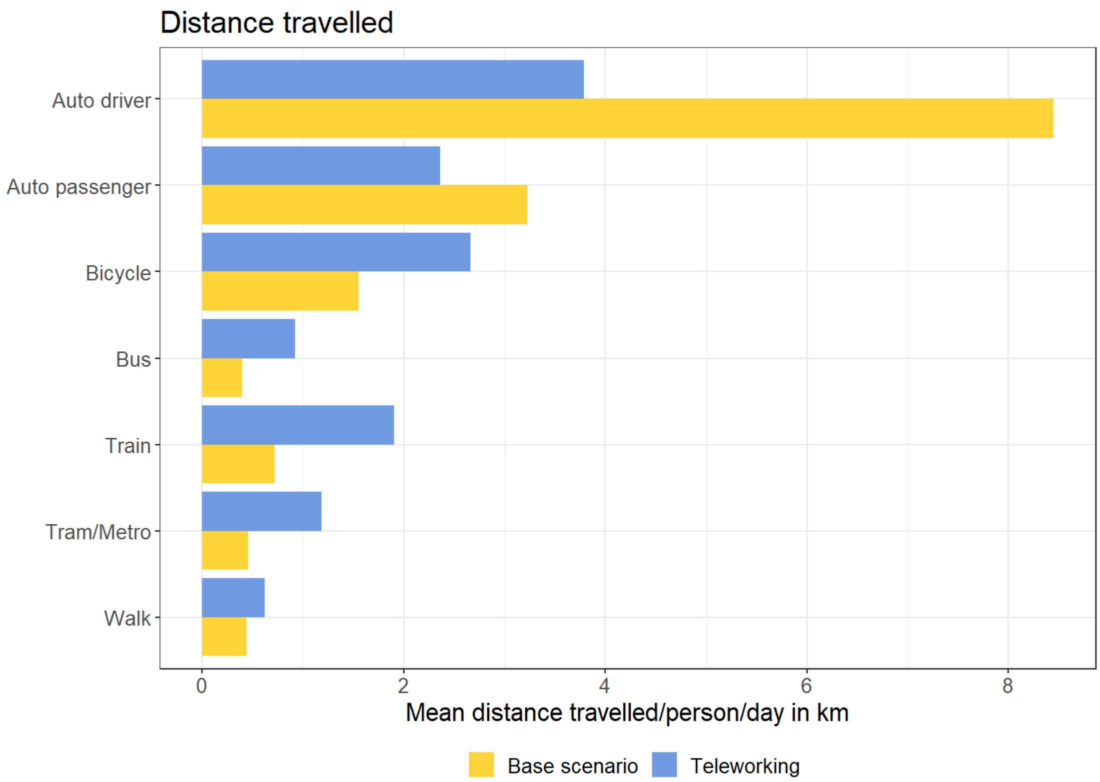


Figure 45 Distance traveled by modes in base and car-free city scenarios

As expected, there was an overall increase in physical activity volume in the car-free city scenario. It is notable that the shares of non-active individuals decreased. This indicates that some individuals who did not walk or cycle in the base scenario started using active modes in the car-free city scenario, perhaps these were car-dominant users.

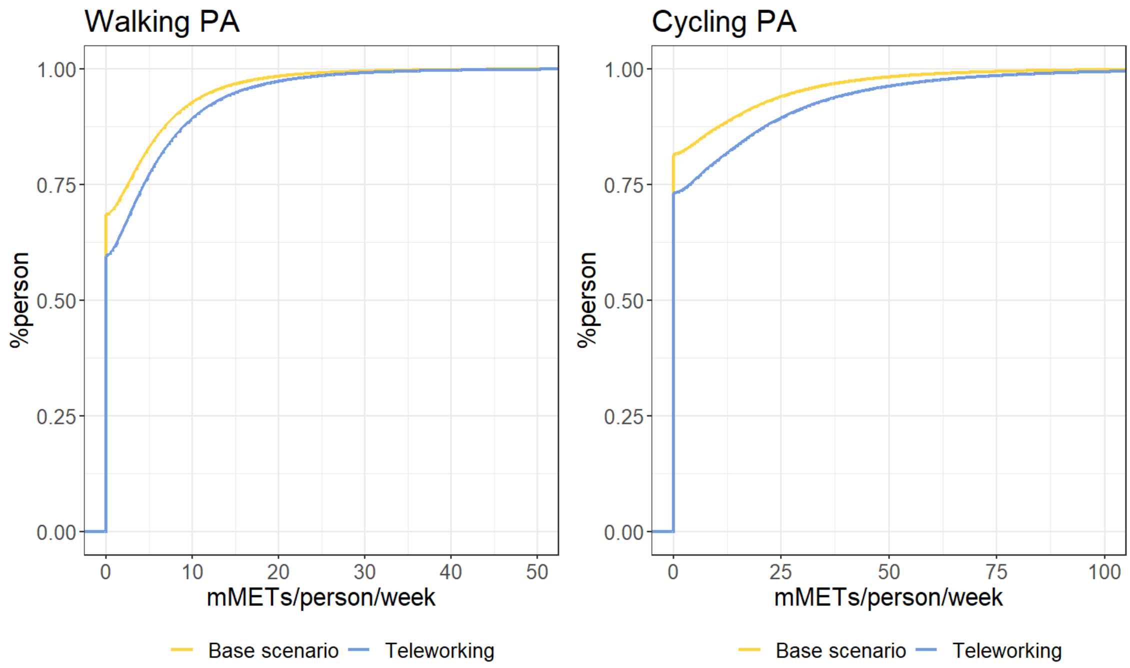


Figure 46 Cumulative distribution of walking and cycling physical activity volume in base and car-free city scenarios

7.4. Discussion on Munich Scenarios

In this chapter, three scenarios were applied to the integrated model for the Munich Metropolitan area. This helped us to check the capability of the integrated model, as well as its limitations.

First, the telework scenario revealed that individuals who work from home tend to have more and longer active discretionary trips. The restriction of car use in the car-free city scenario also encouraged a large shift from motorized vehicles to active modes. While both scenarios underlined an increase in physical activity volume for the affected population, it is important to note that the effects may not be that significant when applying more realistic assumptions. For example, the calculation of weekly PA volume was heavily simplified by using average PA intensity. The scenarios may have different effects on different population groups when using more precise PA intensities differentiated by gradient, speed, age, and sex.

Second, the significant increasing effect on physical activity volume may not be true. Kölbl & Helbing (2003) pointed out a law of constant average energy consumption for the physical activity of daily travel. This physical activity budget effect was not considered in the integrated model. Therefore, the increase in walking and cycling activity in the scenarios was not constrained. This study did not prove whether the physical activity budget would increase across populations in the scenario. With the consideration of the daily physical activity budget, the increase in walking and cycling modes may be moderate, and more car trips may be shifted to non-active modes such as public transport.

Third, the excess commute scenario revealed a major limitation of the integrated model. The walk mode choice model for HBW trips poorly captured the relationship between distance and utility, resulting in less sensitive to the excess commute scenario. On the one hand, there was an insufficient distance penalty for active modes. On the other hand, it could be because the binary walk mode choice model failed to account for the utility of competing modes. Further research is needed to have a closer investigation into the model estimation approach.

Finally, it is important to note that the scenarios presented here were defined by some radical strategies. They have served as a demonstration but not for an actual policy evaluation.

8. Google Timeline: A Novel Dataset for Exploring Pedestrian Travel Behavior

In the previous chapter, the integrated model revealed that the lack of representation of travel behavior variability could lead to some issues in model accuracy and model sensitivity, particularly when assessing travel outcomes (e.g., weekly walk and cycle time) for evaluating the health benefits of policies and scenarios.

Given the scarcity of longitudinal travel behavior data and the limitations of household travel surveys, this dissertation will conduct a survey to collect longitudinal and passive data – Google Location History (GLH) – that records individual trip diaries with coordinates and a board time horizon. Following that, the GLH data will be employed to have a closer investigation of travel behavior variability. This chapter demonstrates the survey design, empirical examination, and some approaches to analyze travel behavior variability.

First, Section 8.1 introduces what Google Timeline is and how trips and modes are identified in the data. Then, Section 8.2 documents the development and administration of the longitudinal travel behavior survey. Survey results, such as sample distribution, are demonstrated in Section 8.3. After that, the methods of data filtering and processing are employed in Section 8.4. Some descriptive statistics of respondents' travel behavior are presented in Section 8.5. Finally, the processed GLH data is employed in Section 8.6 to conduct four research tasks of exploring travel behavior variability. In general, all the research tasks aim at finding the regularity of travel behavior, further assessing week-long pedestrian activities, and investigating the potential determinants of week-long travel behavior by using theory-based or data-driven approaches.

8.1. Introduction of Google Timeline data

In the Google Maps smartphone application, the location information is passively collected if the user enables the built-in function called Google Timeline. It automatically records the geographic coordinates of user's daily trips across a wide time spans and spatial context (Figure 47), resulting in the longitudinal dataset called Google location history (GLH).

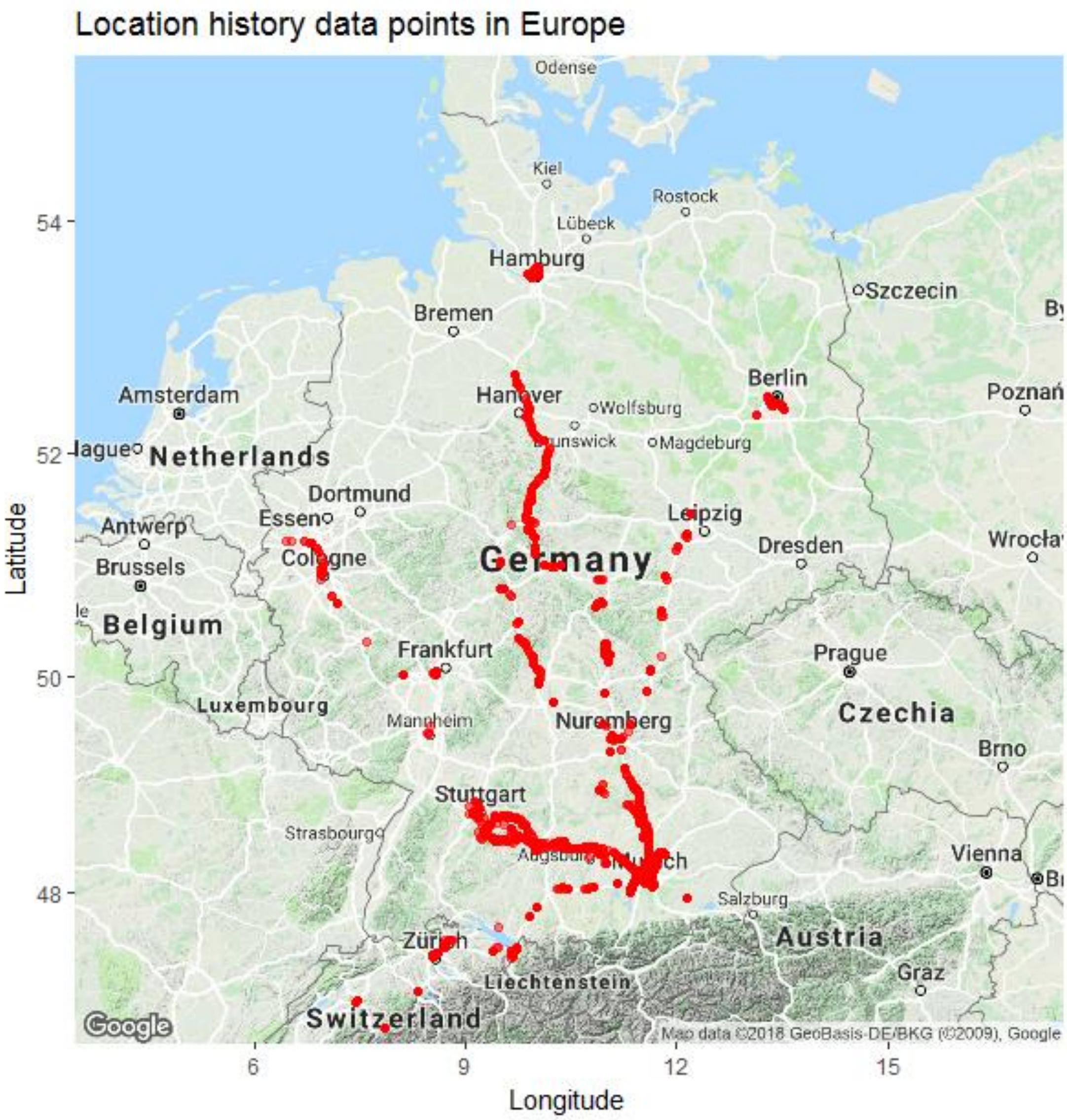


Figure 47 Example of GLH data points across Europe over five years

As shown in Figure 48, conventional household travel surveys are limited to coarse spatial units and the single-day time scale. Although there are efforts made in combining HTS with GPS tracker technologies, it is a relatively new approach with only a few examples and limited performance. The GLH data used in this dissertation covers a wide range of time spans that are not possible to be captured in other datasets. Because GLH data is passively collected, it avoids recall bias when compared to self-reported travel surveys. It also reduces the burden on respondents when compared to GPS trackers. More importantly, GLH data like the other smartphone data is identified using a combination of the phone's internal GPS, connected WiFi devices, and cell towers, so it is spatially as fine as GPS tracker data.

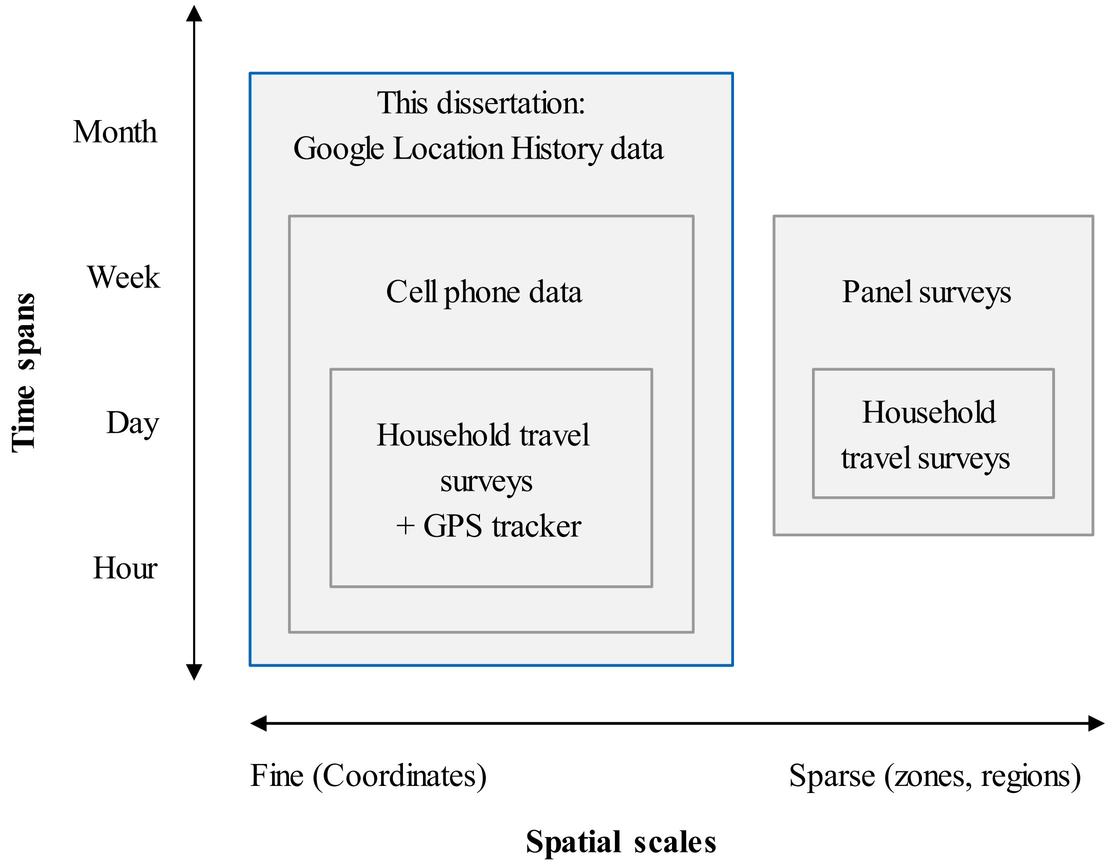


Figure 48 Spatial scales and time spans covered in different data sources

Once GLH data is recorded, each user can access their own GLH data and download the source data file on the Google Takeout website. The GLH data points are stored in JavaScript Object Notation (JSON) files. There are two types of downloadable files: raw location history JSON files and semantic location history JSON files. Raw GLH data consists of a list of timestamped location records arranged chronologically. Compared to the raw data, semantic GLH data contains more high-level and processed information. The semantic information is the same as those presented in the Timeline pages (seen in Figure 49). Besides the timestamped location records, the information is aggregated and summarized as a daily travel diary, including a sequence of inferred place visits, activity segments (as known as means of transport in travel surveys) between place visits, and all with a start time and an end time.

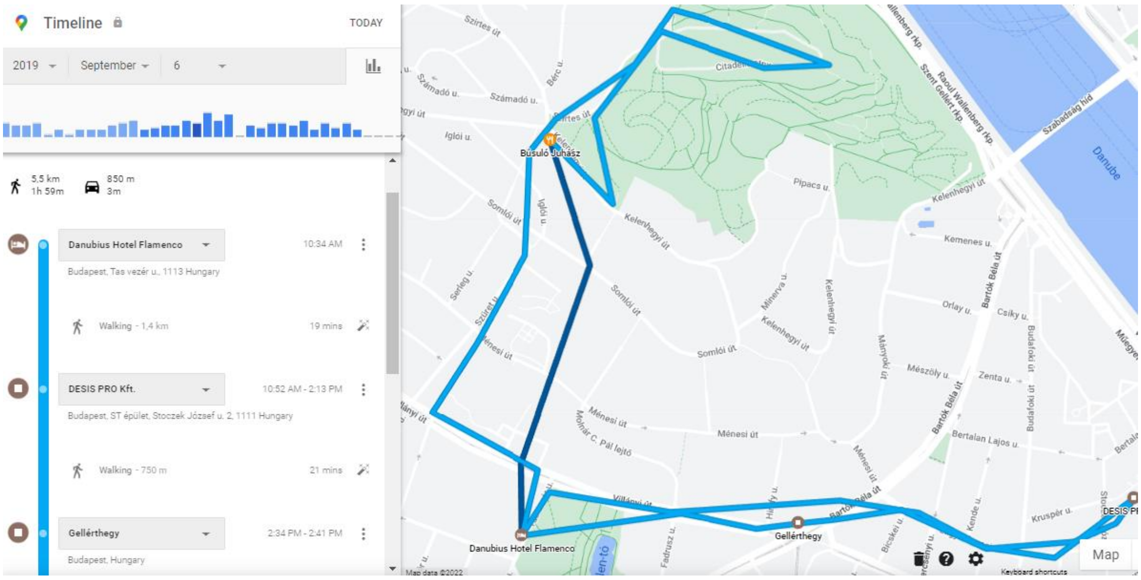


Figure 49 Example of semantic GLH information presented on the Google Timeline webpage

In this study, semantic GLH data will be employed for several reasons. First, common challenges of using location points data are inferring travel mode and purpose, as this information is usually not provided along with the location points. Many algorithms were developed to detect travel mode and identify activity purpose, but the performance of inference varies greatly, particularly when differentiating non-motorized modes. The inferred information provided in the semantic GLH data can help the author tackle this challenge so that more efforts can be made into utilizing the data for further investigation of travel behavior. Furthermore, benefits from some built-in infrastructures of smartphones, GLH data can make use of the accelerometer sensors to deduce travel mode. Therefore, the author makes the hypothesis that GLH data can better capture the information of travel mode.

However, the algorithms for processing GLH data are not accessible to the public. To test the hypothesis, a small validation test was conducted. Three volunteers were asked to enable their GLH services in Google Maps application for a week. In the meanwhile, they were asked to manually record their travel diary, including the arrival and departure time of each activity location, detailed address and name of activity location, and the means of transport used between locations. The trip information in GLH can be edited/corrected by users. However, in this validation study, participants were required not to make any correction to their GLH information. Overall, in the validation dataset, GLH captured 94% of the trips generated by samples across a week. The 6% missing trips were mostly short walking trips less than ten minutes. For the captured trips, the mode detection was totally correct for walk trips, while it had lower accuracy in detecting different kinds of public transport modes. Because the focus of this study is on analyzing pedestrian travel behavior, the bias in public transport detection will not have large impacts on the following analyses.

8.2. Longitudinal Travel Behavior Survey

In this study, a survey was carried out to better understand individual's travel behavior variability. The goal is to determine how variable or stable travel behavior is, as well as how various life events interact with transportation choices, including the use of active modes. The survey collected the GLH datasets along with information about personal characteristics, transportation resources and major life events. After the survey development and a small pilot study, the final survey was administered to residents in Munich metropolitan area during January, February, and March 2020. The following sections go into detail about the survey design process, pilot survey feedback, and recruitment procedures.

8.2.1. Questionnaire Design

To understand more about travel behavior variability and its potential determinants, an online questionnaire was designed and administered through Qualtrics. The questionnaire was developed both in English and German, so that the majority of the residents in Munich can have equal access to the survey. This survey consists of two parts. All participants were invited to take part in both stages. The first part is to fill out a short questionnaire that asks questions about their individual and household characteristics, and then a brief introduction about Google Timeline that is automatically collected by their smartphone. In the end of part one, the participants were asked to upload their GLH data for the period from 2018 and 2019. The second part is a series of questions about the occurrence of major life events that happened since January 2018. The logic/procedure of the questionnaire is shown in Figure 50.

Although some participants may have GLH data for longer time before 2018, the time period in this survey was limited to two years due to several reasons. First, technology was rapidly evolving. Different technologies may have impacts on the quality and the structure of GLH data. Data with larger time spans may involve multiple technologies, resulting in incomparable data quality. Second, asking people to recall life events that occurred a long time ago may be an unnecessary response burden that can lead to respondent fatigue. Finally, two-year travel behavior data is sufficient for the study purpose and is much richer than those used in the previous studies.

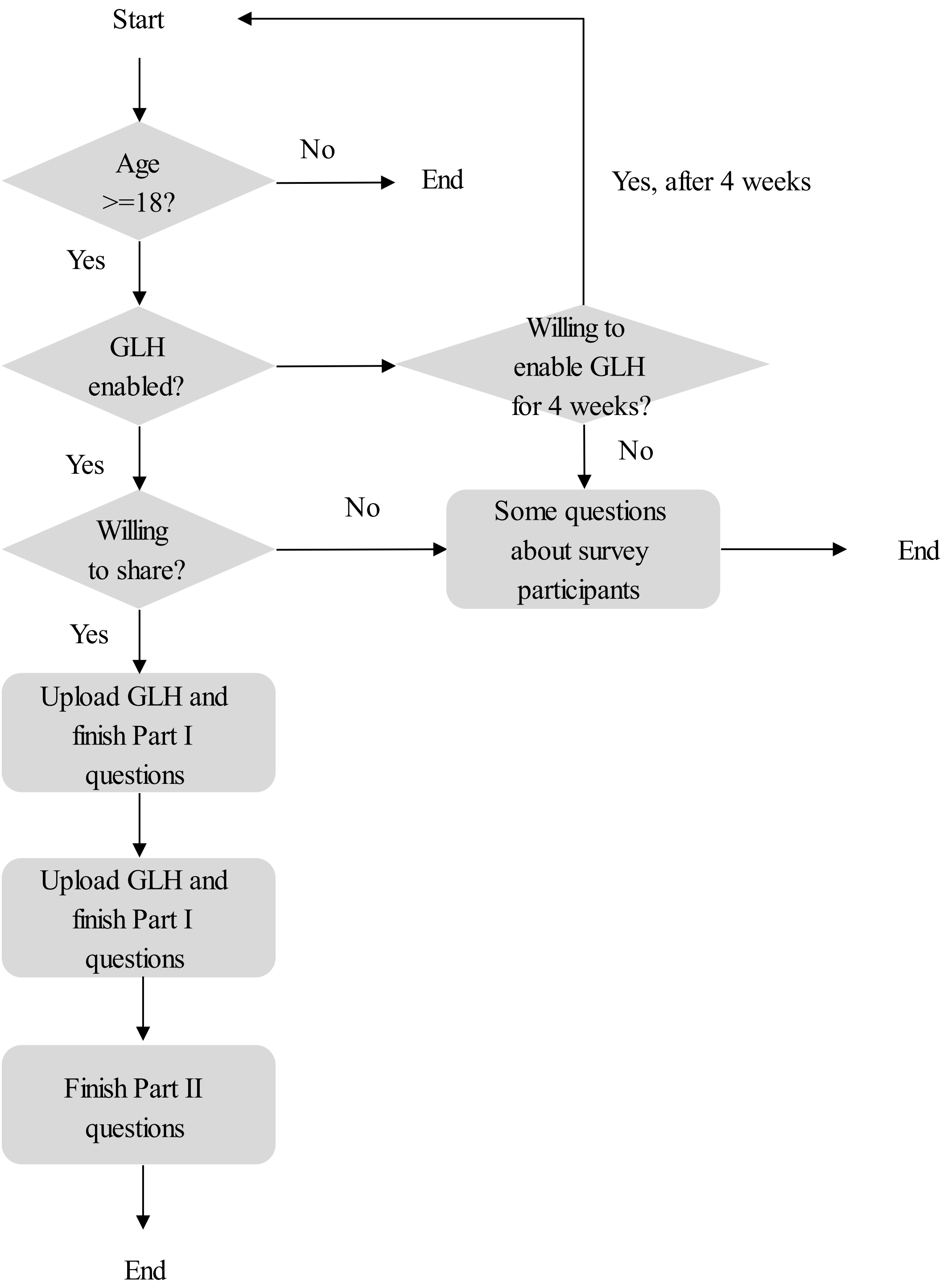


Figure 50 Flow chart of the questionnaire logic

8.2.1.1. Part I

Part 1 begins with a descriptive introduction page and indication of consent, followed by an eligibility question that restricted participants to those who were aged over eighteen.

After that, it shows a brief introduction of what is Google Timeline and asks the question to check if GLH service was enabled on participants' smartphone. Instructions in both English and German are provided on how to check whether GLH is enabled. For those who already have GLH enabled on their smartphones and are willing to share GLH data, they will follow the instructions to download their Google Timeline files and upload them to the survey platform. The instructions are provided in English and German for iPhone users and Android users.

For those who didn't have GLH enabled, they were asked if they were willing to participate and turn on GLH service for the next four weeks. Those who were willing to participate will be followed up and asked to conduct the survey again after four weeks.

After uploading the GLH data, a group of questions about transportation-related and personal characteristics are asked in order to link them with travel behavior variability in later studies. Transportation characteristics include:

- Driver license holding
- Transit pass holding
- Membership in car-share, ride-share, and bike-share services
- Personal vehicle access, including both automobiles and bicycles

Personal characteristics include:

- Age group
- Gender
- Employment status
- Household structure (e.g., number of people living in the household, number of children under six years old, number of children between six and eighteen years old)
- Number of employed people in the household
- Monthly income (before taxes), in categories

8.2.1.2. Part II

Part II of the survey is to ask about changes participants have experienced since January 1st 2018, including changes to the household structure, transportation resources, employment status or location, and student status. Participants are asked to select events that happened to them, and then correspondingly select in while month(s) these events happened. If the event happened more than once since January 2018, then they need to select all the months that apply. The change in household size could have major impacts on individual's lifestyle and travel behavior. The survey attempts to capture not only the changes in household size but also the structure of the household members. The following relevant events are summarized and asked about in the survey:

- A family member/partner/roommate moved out of your residence
- A family member/partner/roommate (who lived with you) passed away
- A family member/partner/roommate moved into your residence
- A new child/children arrived in your household (through birth, adoption, step child etc.)
- Other events that changed the number of adults or number of children in your household

Then, the transportation-related events include:

- Got/lost access to a car/motorcycle
- Gained/got rid of a driver's license
- Registered/got rid of car-sharing membership
- Got/got rid of transit pass
- Got/lost access to a bicycle
- Registered/got rid of bike-sharing membership

For employees or students who have an internship, the following events could happen to them:

- Started/lost a new job
- Increased/decreased working hours
- Changed job location
- Promoted to a position with more responsibilities

For students, the following events could happen to them:

- Started attending a school/university
- Graduated from school/university
- Dropped out of school/university
- Changed school/university location

8.2.2. Pilot Survey

Before deploying the survey for final data collection, a pilot deployment was undertaken. This piloting involved two steps: First, the initial survey was distributed to a small group of students and employees from Portland State University (N=6). An oral feedback session was conducted with the pilot survey respondents directly after they finished the survey. In the oral session, pilot participants shared their thoughts about the survey, any confusing sections, and suggestions for improvement.

Appendix 2 documents the full list of concerns and suggestions derived from the oral feedback session. Several changes in language, format and styles were made to the survey based on participants' suggestions, which makes the survey easier to read and the questions easier to answer. In addition, two major changes were made to the survey as a direct result of the piloting and participants' feedback. First, the participants pointed out the importance of having a more compelling welcome page/introduction, which should establish a relationship with the person and present the goal of the study by using simple and easy-to-understand language. Therefore, some facts and figures were added to the welcome page and visualized nicely in the final survey design (see Figure 51). It linked this study of active transport with an individual's physical well-being. The final survey that incorporated the feedbacks from pilot participants is shown in Appendix 3.

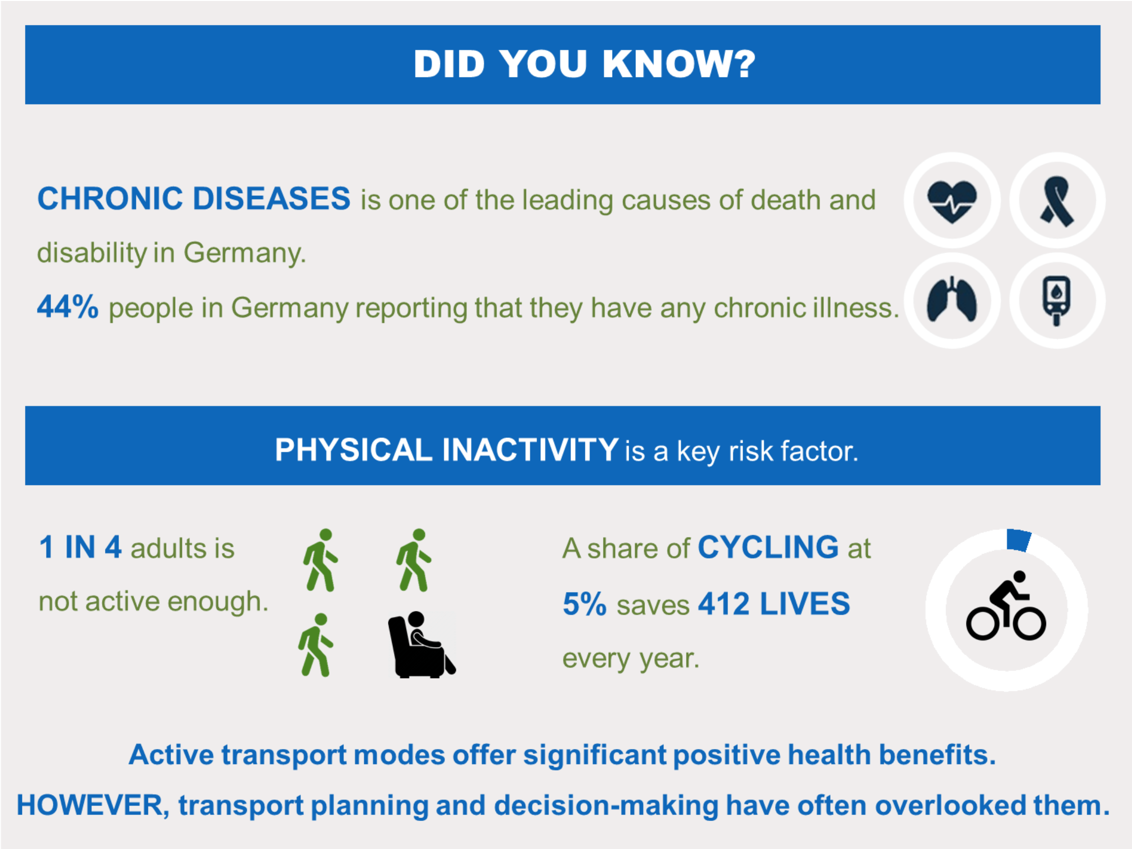


Figure 51 Welcome page in the final survey design

8.2.3. Recruitment

Because of the uniqueness of the GLH data, the recruitment process was deliberately planned. The most serious concern among the pilot participants is the issue of data privacy. The author acknowledged that the GLH data is extremely private. Therefore, the traditional recruitment methods of random sampling, such as flyers, press release, and postcard/letter, are not applicable in this study. Given the specialty of GLH data and limited time, budget and human resources, convenience sampling method was employed. The primary target groups are convenient samples, such as friends and colleagues of the researcher, students, and employees at the Technical University of Munich (TUM) because it is more possible to establish trust with people who know the researcher. Besides that, an existing email list maintained by TUM graduate school was used to access a large group of doctoral students.

Several actions were taken in the recruitment procedure to establish trust with people and minimize their concerns about data privacy.

First, in order to discuss privacy concerns about providing personal GLH data, the survey recruitment of convenient samples was all conducted in person. The researcher first contacted organizations via emails, including research groups at TUM and associated departments in companies, to express the research interest and the study purpose of the survey. Then, an in-person information session was arranged for all potential participants at each organization. During the session, a five-minute in-person pitch was given by the researcher. The presentation focused on giving information about the study purpose, what is Google Timeline, and how the researcher protects data privacy. Besides that, an information sheet was provided, which clearly stated the data privacy protection rules. (See Appendix 4). After the in-person information session, a recruitment email was sent to the potential participants. A sample email invitation is shown in Figure 52. Due to the effort required for in-person recruitment, the number of recruited individuals in-person is limited. In total, 31 friends and families of the author were contacted. 14 organizations were contacted including three departments in companies, six research groups at TUM and four university classes at TUM. Approximately 200 potential participants were recruited in person.

Portland State UNIVERSITY

Technische Universität München TUM

Long-Term Travel Behavior Survey

We kindly ask for your help with this research survey about your long-term transportation choices. The data that you share will be useful in understanding how people make changes to their transportation patterns after life events. The survey includes questions about:

- Your personal and household characteristics
- The occurrence of your major life events since January 2018.
- Your location data history is automatically collected by your smartphone.

The survey will take about 25 minutes to complete and can be answered **only via computer**. If you complete the survey, you will have the opportunity to enter a draw to win one of ten Amazon gift cards.

Click here to take the survey!
https://portlandstate.qualtrics.com/jfe/form/SV_9AF4NdIram3WtQF

Your responses will be confidential, and you will not be identified in any reports or analysis.

This research is being carried out by the doctoral candidate Qin Zhang, a research associate at the Technical University of Munich (TUM) advised by Dr. Rolf Moeckel of TUM and Dr. Kelly Clifton of Portland State University (PSU). If you have questions or concerns, you may contact us by email at qin.zhang@tum.de or by calling +49-89-289-22698.

Figure 52 Sample recruitment email for the Longitudinal Travel Behavior Survey

Second, for those who cannot attend the in-person information session and for people who were recruited via email. A URL that linked to the project website was provided in the email. On this website, the project goals were briefly described, and a link to the Qualtrics survey was highlighted. More importantly, a five-minute video was posted on this website to inform project goals and address data privacy concerns. Figure 53 shows a screenshot of the project website.

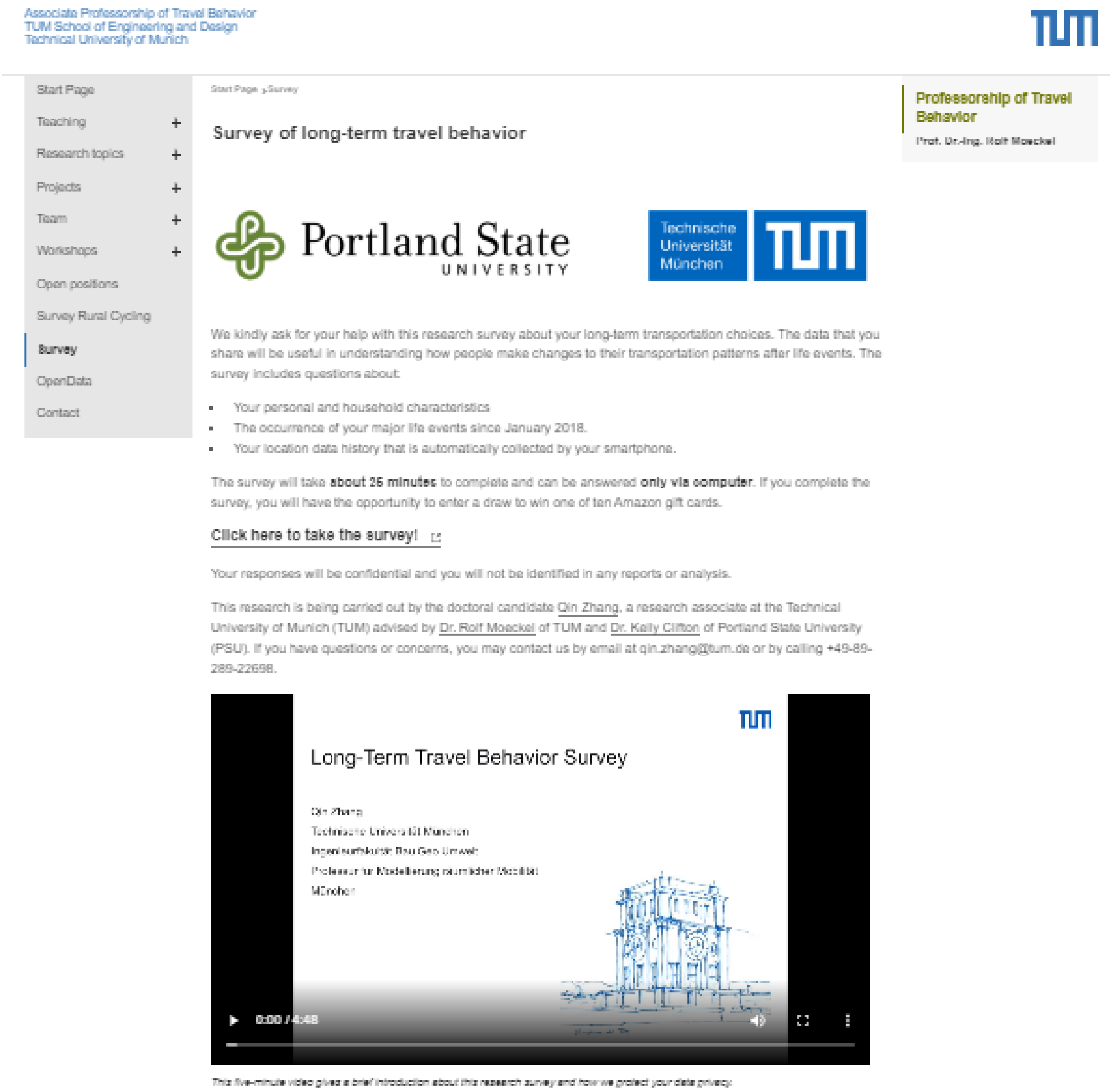


Figure 53 Project website for the Longitudinal Travel Behavior Survey

8.3. Survey Results

After in-person and email recruitment, 98 survey responses were collected. Almost all the respondents were recruited from the in-person recruitment approach, and the email-list approach did not provide any successful responses.

All participants were invited to take part in an online survey about their socio-demographics, transportation resources and major life events. Then, they were asked to upload their location history data (GLH) for the period from 2018 and 2019. Out of 98 responses, 53 respondents completed the online survey. Of those, only 29 respondents had GLH data collection previously enabled on their mobile phones and were willing to share them. 14 respondents did not have GLH enabled but were willing to turn it on and come back in four weeks. However, no additional samples got after four weeks. Table 14 shows the number of responses of each survey section.

Table 14 Number of responses completing each section of each survey

Survey section	**Count**	**Share**
Started	98	100%
Introduction of Google Timeline	98	100%
Check if Google Timeline enabled	94	96%
Check if Google Timeline enabled - Yes	56	57%
Transportation resources	53	54%
Personal characteristics	52	53%
Upload Google Timeline data	29	30%
Life events	29	30%
Submitted	29	30%
Check if Google Timeline enabled - No	38	39%
Keep Google Timeline enabled for the next 4 weeks - Yes	14	14%
Keep Google Timeline enabled for the next 4 weeks - No	16	16%
Personal characteristics	26	27%
Submitted	24	24%

Figure 54 summarizes the times respondents took to complete the survey. For respondents who did not have Google Timeline enabled had a mean completion time of 6.4 minutes, while for those who had Google Timeline enabled finished the survey with a mean duration of 24.9 minutes. As expected, downloading and uploading GLH data took most of the survey time. There were some outliers taking more than 40 minutes to finish the survey. It may be caused by the pause, interruption, or some technical difficulties in downloading and uploading GLH data.

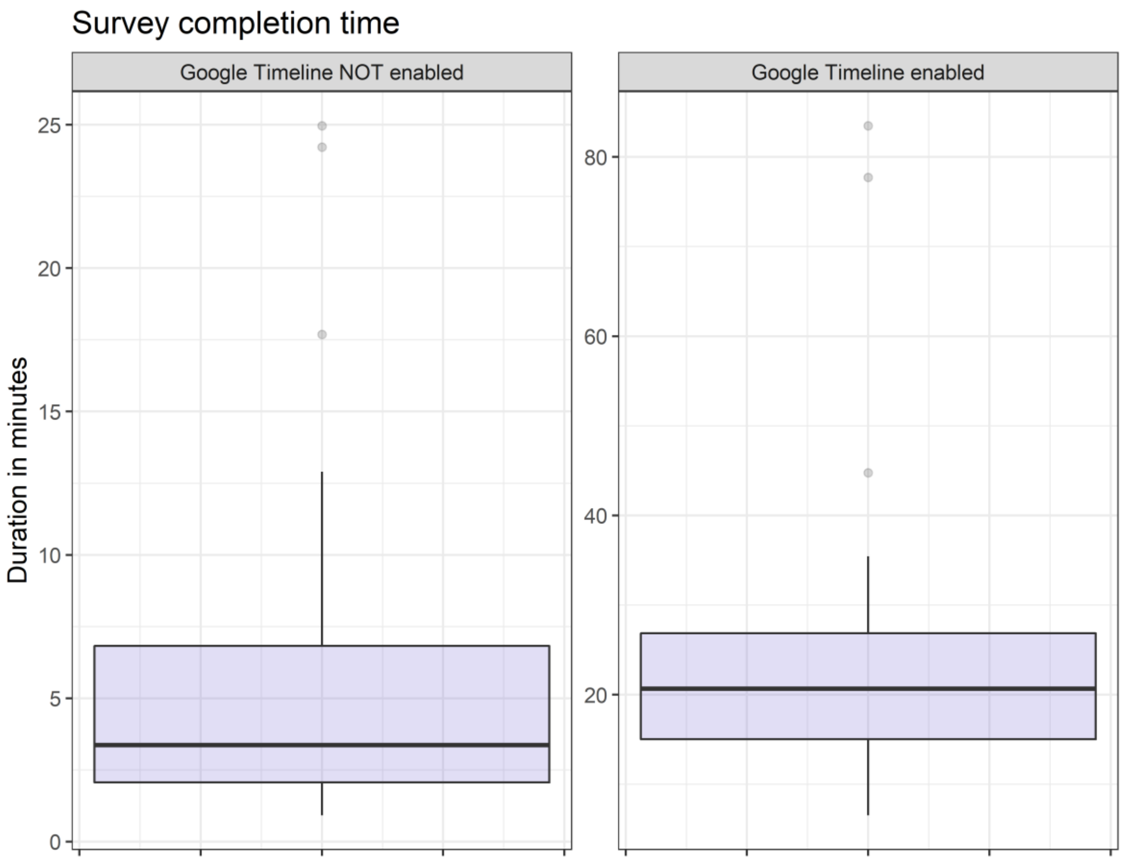


Figure 54 Boxplot of survey completion time

8.3.1. Sample Description

Figure 55 presents the demographic information of the 29 respondents who shared their GLH data and completed the survey. Figure 56 compares the sample distribution with the German census distribution in terms of some key socio-demographic attributes. The data are skewed towards participants who are aged from 25 to 34. This is not surprising because as anticipated, the convenience sampling recruitment resulted in university-centric respondents. The data are also skewed towards male participants. Most of the respondents (20 out of 29) are employed full time. medium-to-high income, living in two-person households without children.

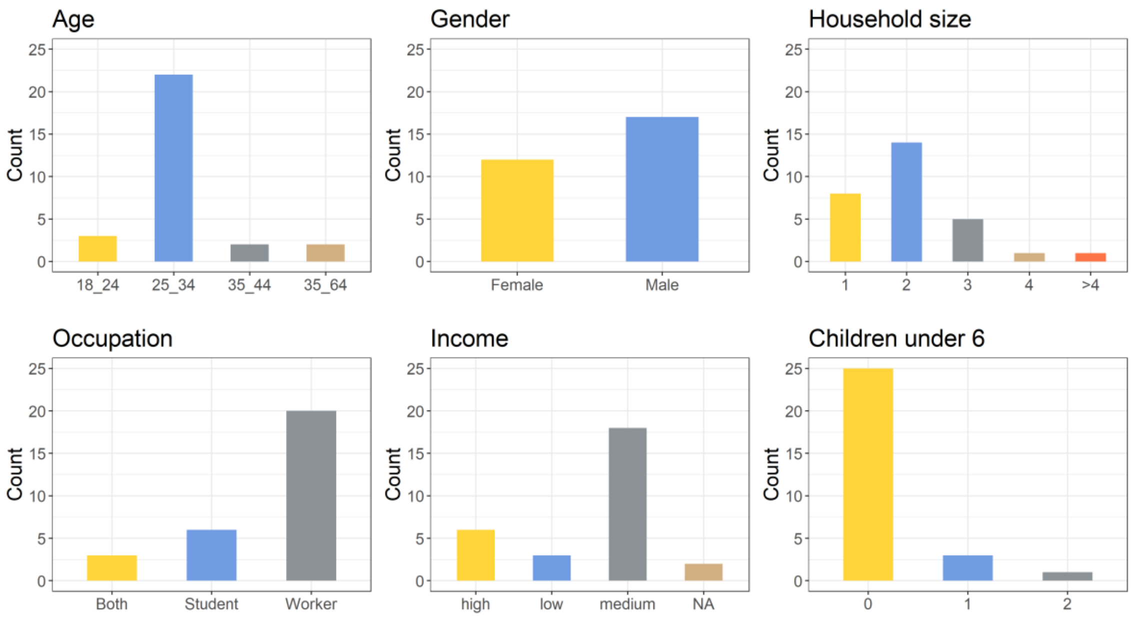


Figure 55 Survey respondents' socio-demographic attributes

Figure 56 Respondents' socio-demographic distribution compared to the German census distribution

Table 15 shows the number of respondents in each mobile phone type. The mobile phone types can have impacts on the quality of GLH data points. In this study, respondents' mobile phone types covered the most popular brands in the market and were evenly distributed.

Table 15 Number of respondents in each mobile phone type

Mobile phone type	Count
Apple (iPhone)	7
Huawei	8
Samsung	6
Other	8

Figure 57 demonstrates transportation resources of survey respondents. Transportation resources are important indicators of choosing travel mode. It also to some extent indicates an individual's travel behavior pattern. For example, people who hold transit pass are more likely to be a frequent transit user and have more stable behavior in terms of mode usages. Most of the respondents hold a driver license and/or a transit pass. Shared mobility subscription is another way of accessing a car and bike. Munich currently has several shared mobility service providers, including DriveNow, Sixt Share, and MVG Bike. Shared mobility is still not commonly used by the survey respondents. Most of the respondents have no access to a car but have one or more access to a bike.

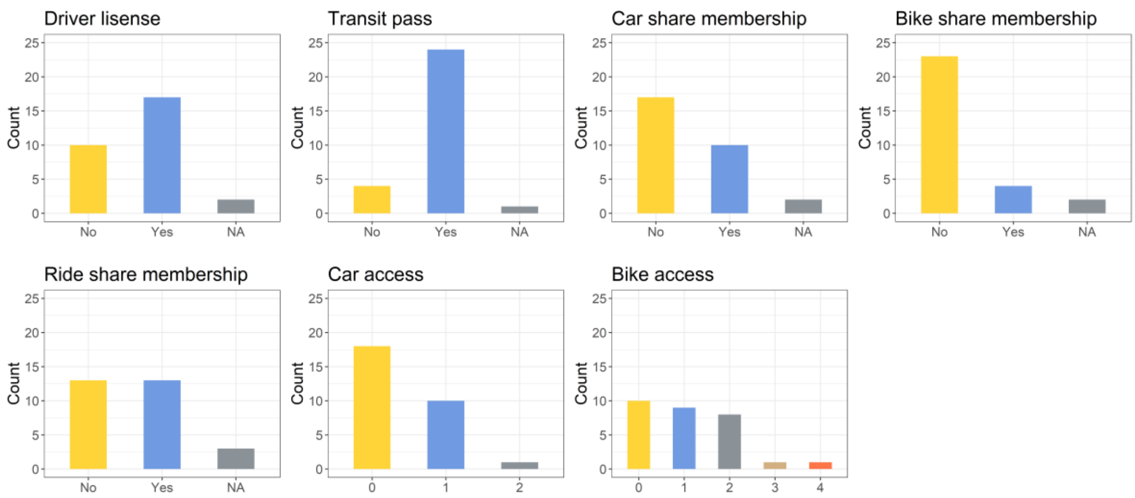
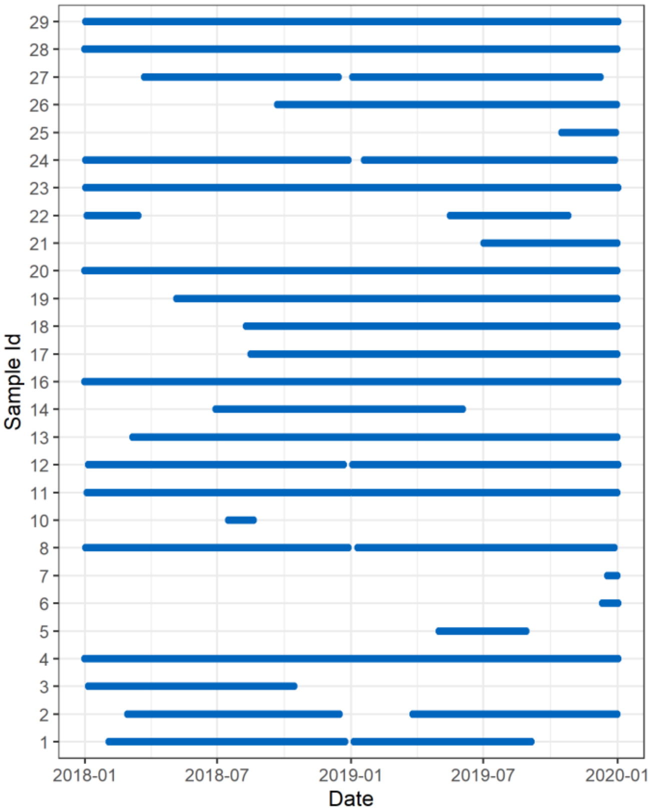


Figure 57 Survey respondents' transportation resources

8.3.2. Google Timeline Data

The time period of GLH data varies for each respondent and not all have two complete years of GLH data. This may be because respondents disabled the GLH at some point. Figure 58 shows the time period of GLH data and the number of recorded days for each respondent. Respondent 9 and Respondent 15 are excluded since their GLH datasets are empty. This may be because the GLH service was never enabled on their devices, but they answered "is enabled" in the survey. As a result, 27 valid GLH datasets were used in this study. Most of the GLH datasets cover a large number of successive days over the two-year period. The mean number of days recorded is 481 days. More descriptive analysis of GLH data will be demonstrated in Section 8.5 after GLH data processing.



SAMPLEID	NUMBER OF DAYS RECORDED
29	726
28	730
27	606
26	465
25	74
24	704
23	730
22	232
21	183
20	727
19	600
18	507
17	501
16	731
14	335
13	663
12	714
11	722
10	34
8	712
7	13
6	21
5	119
4	731
3	282
2	570
1	567

Figure 58 The time length of GLH data and number of recorded days of each respondent.

8.3.3. Life Events

The second part of the survey is a series of questions about the occurrence of major life events that happened in 2018 and 2019. In total, 90 life events were recorded among all respondents. As shown in Figure 59, four respondents had no life events occurrence. Most of the respondents had one to four events that happened in 2018 and 2019.

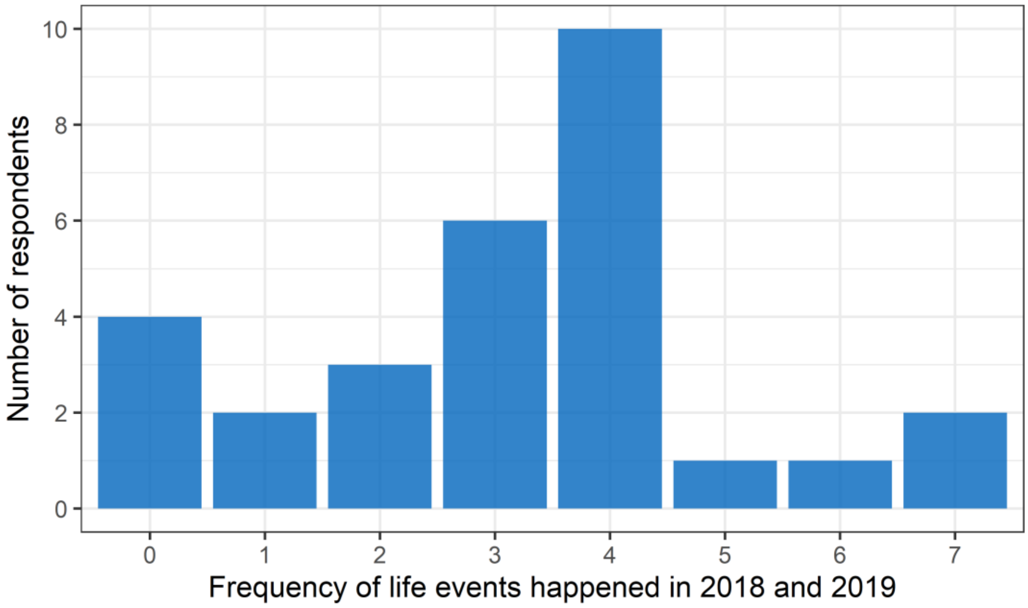


Figure 59 Number of respondents for frequency of life events happened since January 1st 2018

Table 16 demonstrates the number of cases of each event type. Getting a bike and starting a new job are the most common events. It would be worthwhile to investigate how cycling behavior changes after gaining bike access as well as whether starting a new job affects individual's habitual travel mode. The household-related events indicate the changes of household size and household structure. Besides that, there are three household relocation events. The change of neighborhood built environment may also change individual's travel behavior. These events will be employed in Section 8.6.4 to explore the association between life events and travel behavior changes.

Table 16 Number of cases of each life event type

Category	Life event type	Number of cases
Household events	Person moved out	4
	Person moved in	5
	New birth	1
	Moved home	3
Mobility events	Got car access	3
	Lost car access	4
	Got driver license	2
	Registered car sharing	1
	Registered e-scooter	1
	Got bike access	13
	Lost bike access	5
	Registered bike sharing	1
Job events	Started new job	15
	Lost job	2
	Changed job location	5
	Promotion	4
	Increased work hour	5
	Decreased work hour	2
School events	Started school/university	6
	Graduated from school/university	6
	Changed school location	2

8.4. Google Timeline Data Processing

Semantic GLH data contains more high-level and processed information. Besides the timestamped location records, the information is aggregated and summarized as a daily travel diary, including a sequence of inferred place visits, activity segments (as known as means of transport in travel surveys) between place visits, and a start time and an end time of each activity.

As seen in Figure 60, Semantic GLH data is stored in JSON file. The timeline object consists of all available semantic information in chronological order. Each item in the list is either an activity segment or a place visit.

Activity segment is an activity involving location changes from one place to another, which can be considered as a trip in the context of transport models. Each activity segment contains the start and end location, duration of the trip, distance traveled during this trip, and activity type. Here the activity type refers to the means of transport. The activity type is inferred with a certain probability. The most common activity types are walk, car drive, bus ride, subway ride, cycling or running.

Place visit means a visit or stay at a fixed location for a duration of time, for example stay at home, at work, or at a shopping mall or a restaurant. Place visit provides information includes location coordinates and length of stay at the location. It also provides confidence in demonstrating the accuracy of location inferences.


```
"timelineObjects" : [ {
  "placeVisit" : {
    "location" : {
      "latitudeE7" : 99577965,
      "longitudeE7" : 763198223,
      "placeId" : "hidden",
      "address" : "hidden",
      "name" : "hidden",
      "sourceInfo" : {
      "locationConfidence" : 19.28083
    },
    "duration" : {
      "startTimestampMs" : "1555171705633",
      "endTimestampMs" : "1555171995755"
    },
    "placeConfidence" : "LOW_CONFIDENCE",
    "centerLatE7" : 99584120,
    "centerLngE7" : 763186620,
    "visitConfidence" : 83,
    "otherCandidateLocations" : [ {
    "editConfirmationStatus" : "NOT_CONFIRMED"
  }
```

```
"activitySegment" : {
  "startLocation" : {
    "latitudeE7" : 99713231,
    "longitudeE7" : 763182966,
    "sourceInfo" : {
  },
  "endLocation" : {
    "latitudeE7" : 99586420,
    "longitudeE7" : 763242722,
    "sourceInfo" : {
  },
  "duration" : {
    "startTimestampMs" : "1555673426752",
    "endTimestampMs" : "1555674040532"
  },
  "distance" : 1628,
  "activityType" : "IN_PASSENGER_VEHICLE",
  "confidence" : "HIGH",
  "activities" : [ {
    "activityType" : "IN_PASSENGER_VEHICLE",
    "probability" : 95.86777827179893
  }, {
    "activityType" : "MOTORCYCLING",
    "probability" : 1.9259377296023605
  }, {
    "activityType" : "WALKING",
    "probability" : 1.2955728144128784
  } ],
  "waypointPath" : {
  "simplifiedRawPath" : {
}
```


Figure 60 An example of the GLH data structures stored in a JSOM file

To use GLH data for further travel behavior analysis, several procedures are needed to process the data. First, each activity segment object is converted to a trip, then combined with place visit, resulting in an activity chain of each day. The values of start and end location are defined as the origin and destination of the trip. The travel time is calculated as the difference between timestamps at activity ends and starts. Travel distance is directly obtained from the distance value. Trip mode is defined by the activity type with the highest probability. Overall, the data for all respondents include 45,050 trips on 12,999 person-days. 1,420 trips (3%) were removed due to unknown activity type and unknown travel distance.

Then, a validity check was conducted to filter out trips with unreasonable travel time or travel distance. Travel speed is calculated based on the observed travel time and travel distance. For motorized modes such as passenger vehicles and public transport (e.g., bus, tram, subway, and train), trips with travel speeds over 300 km/h are filtered out. For cycling and running trips, those with travel speeds over 30 km/h are sorted out. For walk trips, those with travel speeds over 7.5 km/h are excluded. As a result, 823 trips (1.8%) were filtered. Table 17 shows the trip counts and shares by activity types before and after filtering. Finally, there are 42,744 trips used in this analysis.

Table 17 Trip counts and shares by activity types

	Before filtering		After filtering	
Activity type	Count	Share	Count	Share
Passenger car	8,810	20.2%	8,765	20.5%
Public transport	12,101	27.7%	12,055	28.2%
Cycling	3,082	7.1%	3,052	7.1%
Running	251	0.6%	250	0.6%
Walking	19,323	44.3%	18,622	43.6%

8.5. Descriptive Analysis

Table 18 shows the overall descriptive characteristics of the samples analyzed from the GLH data in terms of travel behaviors. The average daily travel time across the respondents is 122 minutes (about 2 hours). It is higher than the average traveling time observed in the German national travel survey 2017 (MiD), which is 75 minutes per person per day (BMVI, 2019b). Similarly, the average number of trips per day is 4.6 in the GLH dataset while it is only 3.1 in the MiD 2017 dataset (BMVI, 2019b). Overall, the respondents traveled more than the average level observed in MiD data. First, this is because the respondents are mostly characterized as graduate students or middle-aged university staff, who usually engage in more daily activities. Another reason is that GLH data recorded not only the daily short-distance activities but also long-distance trips, such as tourist trips during holidays and business trips. More importantly, one limitation of traditional HTS is that respondents generally underreport their travel, resulting in a smaller number of trips recorded in HTS compared to the GLH data.

Regarding walking behavior, on average, the respondents have 2.1 walk trips per day and walk 28.9 minutes per day. This is also above the average level observed in MiD data, which is 0.8 walk trips and 20.1 minutes per day (BMVI, 2019b). On one hand, it suggests that people in the GLH dataset are more active. On the other hand, it may indicate that GLH data can capture more active trips since each short-distance trip is recorded precisely.

Regarding cycling behavior, the respondents have an average daily cycle trips of 0.2 and cycle time of 6.0 minutes. This is much lower than the average values observed in MiD data, which is 0.4 cycle trips and 9.5 minutes per day. This suggests that cycling trips are underrepresented among respondents.

Table 18 Descriptive statistics of respondents' travel behaviors

Travel behavior indicators	Time scale	Mean	SD.	Min	Max
Number of non-mobile days	Weekly	1.96	5.62	0.10	4.74
	Monthly	6.35	26.64	0.25	15.42
Number of trips	Daily	4.62	1.34	1.75	8.54
	Weekly	24.51	11.74	3.96	52.82
	Monthly	118.38	39.21	22.00	193.67
Total travel time (mins)	Daily	122.88	38.00	63.46	262.27
	Weekly	609.67	224.42	235.17	1273.87
	Monthly	2929.83	1026.67	1269.63	6713.08
Number of walk trips	Daily	2.07	0.89	0.61	4.87
	Weekly	11.13	6.61	1.38	30.09
	Monthly	53.28	24.31	2.00	110.33
Total walk time (mins)	Daily	28.93	13.83	7.03	66.17
	Weekly	149.21	89.30	19.39	383.80
	Monthly	709.90	371.03	232.74	1705.00
Total walk distance (km)	Daily	1.77	0.81	0.44	3.74
	Weekly	9.22	5.54	1.21	21.67
	Monthly	43.53	23.06	4.64	96.58
Number of cycle trips	Daily	0.24	0.29	0.00	0.88
	Weekly	1.30	1.70	0.00	5.50
	Monthly	6.62	7.93	0.00	24.04
Total cycle time (mins)	Daily	6.04	9.84	0.00	47.12
	Weekly	27.99	36.97	0.00	110.27
	Monthly	154.29	241.38	0.00	1059.26
Total cycle distance (km)	Daily	0.82	1.30	0.00	5.45
	Weekly	4.02	5.88	0.00	22.28
	Monthly	23.15	37.60	0.00	156.26

8.6. Analysis of Travel Behavior Variability

8.6.1. Investigation of Travel Behavior Variability

Many policy questions about travel behavior are predicated on the desire to understand behavioral change and travel variability in response to changes in an individual's economic changes, environmental conditions, and policy interventions. Until now, the lack of data has hampered the ability to examine these relationships. The emergence of longitudinal, location-based data offers the potential to understand the variability of individual activity and travel patterns. Recent studies have found that people have large variability in daily travel behavior (Hanson & Huff, 1982, 1988; Pas & Koppelman, 1986; Pas & Sundar, 1995; Raux et al., 2016; Schlich & Axhausen, 2003; Susilo & Axhausen, 2014). The level of variability could be quite different when analyzed using different temporal scales, travel metrics, and measures of variation. Most of the recent studies have focused on the day-to-day variability, and few have had the opportunity to observe variability over longer time periods, such as week-to-week and month-to-month.

In this section, we employed the collected GLH data to explore travel behavior variability. Different from previous studies which mainly focused on day-to-day variability, this study observed the variability across various temporal scales (day-to-day, week-to-week, and month-to-month). Furthermore, the variability of various travel behavior indicators is examined. They are the variability of (1) start time of day; (2) number of trips; (3) total travel time; (4) number of walk trips and (5) total walk time.

Among the 27 samples described in Section 8.3.2, a few samples are excluded in some of the analysis here since the length of GLH data is not sufficient to capture the variance of travel behavior. Respondent 6 is excluded in the analysis of week-to-week variability and month-to-month variability. Respondent 7 and sample 10 are excluded in the analysis of month-to-month variability.

8.6.1.1. Measurement of Variability Index

Previous research employed different measurements for calculating travel behavior/pattern variability, but there is a lack of consensus on a common way to identify similarity/variability of travel behavior/pattern. Some studies considered various travel behavior indicators separately and then measured variability from every single dimension (Li et al., 2018; Pas & Koppelman, 1986; Pas & Sundar, 1995; Raux et al., 2016; Tarigan & Kitamura, 2009). The usual travel behavior indicators are the number of trips per day, daily time use allocation, and dominant mode. However, Schlich & Axhausen (2003) argued that travel behaviors are a complex and multidimensional phenomenon. They applied more complex variability measurements to consider the temporal dimension of trips and the sequence of activity chains.

In this section, intrapersonal variability is computed for each respondent in different temporal scales, including day-to-day variability, week-to-week variability, and month-to-month variability. The travel-related indicators computed at the individual level are (1) start time of day; (2) number of trips; (3) total travel time; (4) number of walk trips and (5) total walk time. We adopt the single dimension approach by measuring the variability of various travel behavior indicators separately. First, the single dimension analysis can help us to investigate the variability of each indicator. The results are usually easier to interpret than the multidimensional measurement. Second, the sample size used in this study is relatively small, so to get meaningful outcomes, it is necessary to reduce the level of complexity in the analysis.

Commonly, there are four statistics to measure variability within one dataset. They are range, interquartile range, standard deviation, and variance. However, they are all absolute variability. To compare variability between individuals with widely different means, we need to calculate the relative variability.

The coefficient of variation (CV), also known as relative standard deviation (RSD) is a dimensionless measure of dispersion. It is widely used to express the repeatability of an experiment. It is calculated as $CV = \frac{\sigma}{\mu}$, where σ represents total variance and μ represents the mean.

8.6.1.2. Intrapersonal Travel Behavior Variability

In Figure 61, we compare the intrapersonal variability (CV) of different travel behavior indicators across different temporal scales. There is no standard rule to interpret high or low variance by using CV. The higher the coefficient of variance, the greater the level of dispersion around the mean. As a rule of thumb, the coefficient of variance larger than 1 indicates a relatively high variation, while the coefficient of variance lower than 1 can be considered low dispersion.

When comparing the intrapersonal variability across different temporal scales, we found that day-to-day travel behaviors are more variable than week-to-week and month-to-month travel behaviors. Especially for monthly indicators, the travel behaviors show a high level of similarity. This suggests that people may have a regularity in their activity participation or have a travel budget that gets articulated over a week or a month. However, it also partly due to the smoothing of values over a larger number of days.

An important question of household travel surveys is how many days of observations are needed to obtain an individual's routine travel behavior, while minimizing costs and respondent burden. Schlich & Axhausen (2003) recommended that travel surveys should span at least two weeks in order to capture an individual's activity pattern. Previous studies in the field of health research suggested that a reliable estimate of physical activity using an accelerometer should monitor over three to seven days (Pedersen et al., 2016). The results from this study indicate that there is low weekly variance because all of the travel measures have CV values less than 1. This suggests that a period of one week maybe be sufficient to capture routine travel behaviors in surveys or other data collection efforts. This is consistent with the previous findings, but it is not conclusive. CV is a descriptive measure and additional statistical methods are needed to explore how many successive days are needed to capture sufficient variation in an individual's travel behavior.

The intrapersonal variability index also differs across the different travel behavior indicators. The start time of the first trip in a day has the least variance, across all of the temporal scales. Since most of the respondents in this study are workers, this suggests more habitual behaviors in the morning, perhaps around commuting behavior

In general, the trip frequencies (number of total trips and number of walk trips) are less variable than the time spent traveling (total travel time and total walk time). The results, however, do support the idea of a travel time budget over a longer time span – weekly or monthly, rather than daily. The idea of travel time budget found in the literature suggests that each person is willing to spend a fixed amount of time to travel across a day or week (Zahavi, 1974). Total walk time has the highest variance among all travel behavior indicators across all of the time scales. While this study does not offer evidence to support why this may be, walking behaviors may be more variable because the decision to walk may be more subject to the weather, physical activity budgets, and the built environment. Thus, one-day travel surveys may not be sufficient to characterize walking activity for travel or physical activity.

It is also interesting to examine the degree of dispersion of the individual variability within each measure of travel behavior. The month-to-month variation has a narrower range of values across all measures of travel behavior. As with the CV index, the start time has the lowest range of variability values across all time scales. Again, this is not surprising given the skewed sample of workers from the same location. A larger sample that includes more workers from various industries and nonworkers may have a wider range of variability. The CV in the weekly number of trips has a large degree of dispersion compared to other time scales and other measures. This is likely due to the differences between weekday and weekend activity patterns.

These results suggest that a traditional one-day survey is not adequate to capture a representative sample of complex travel patterns and that patterns may exist over longer periods of time. It also raises questions about the explanatory factors for the differences in variation across individuals.

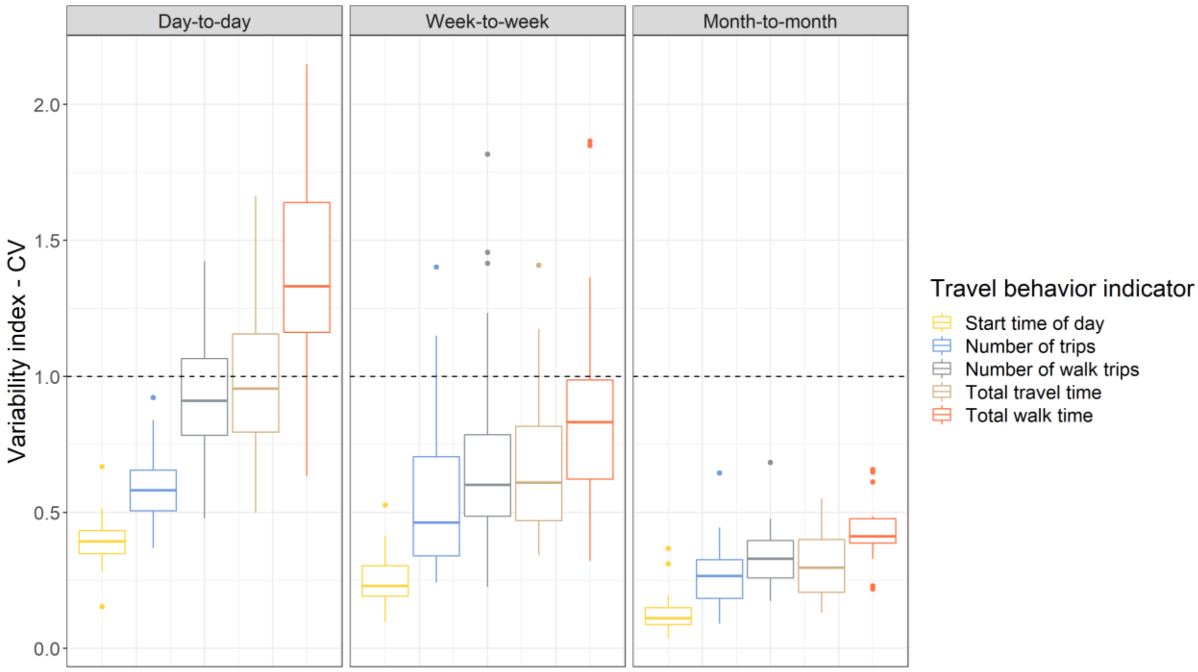


Figure 61 Intrapersonal variability index (CV) of different travel behavior indicators across different temporal scales

8.6.1.3. Impacts of Socio-Demographics on Intrapersonal Travel Behavior Variability

Based on the analysis in the previous section, we found that the degree of intrapersonal variability varies widely across different temporal scales and most travel measures have a wide range of CV values across individuals in the sample. As a result, we will test the contributions that various socio-demographic characteristics might have on the level of intrapersonal travel behavior variability. To test these associations, a set of univariate linear regression models are estimated to investigate the effects of socio-demographic variables on intrapersonal variability. The regression analysis of variability is conducted for the four travel behavior indicators (number of trips, number of walk trips, total travel time, and total walk time) for each of the temporal scales. The intrapersonal variability in the start time of day is not analyzed here because it has little variance observed in the previous section.

The personal and household characteristics shown in Figure 55 are considered as predictors for regression analysis. They are age group, gender, occupation status, income group, household size, number of cars, and number of young children (six years or younger). The number of observations is small in some categories, so several categories are combined. For example, the age group from 35 to 44 is combined with the age group from 45 to 64. Occupation status includes student and full-time employee. Monthly income (euros) has three categories, high-income group (more than 6,000), medium-income group (2,000-6,000), and low-income group (under 2,000).

There are several reasons why we use univariate linear regression models. Due to the small sample size in this study, we reduced the model complexity to better explore the associations with socio-demographic variables. The second reason is the high correlation among socio-demographic variables. As shown in Figure 62, income groups are highly correlated with age groups. Also, household size, income, car access, and the number of young children are correlated. Using simple regression models (a linear regression model with a single explanatory variable) can help us to avoid the multicollinearity issue and directly investigate the impacts of each socio-demographic variable.

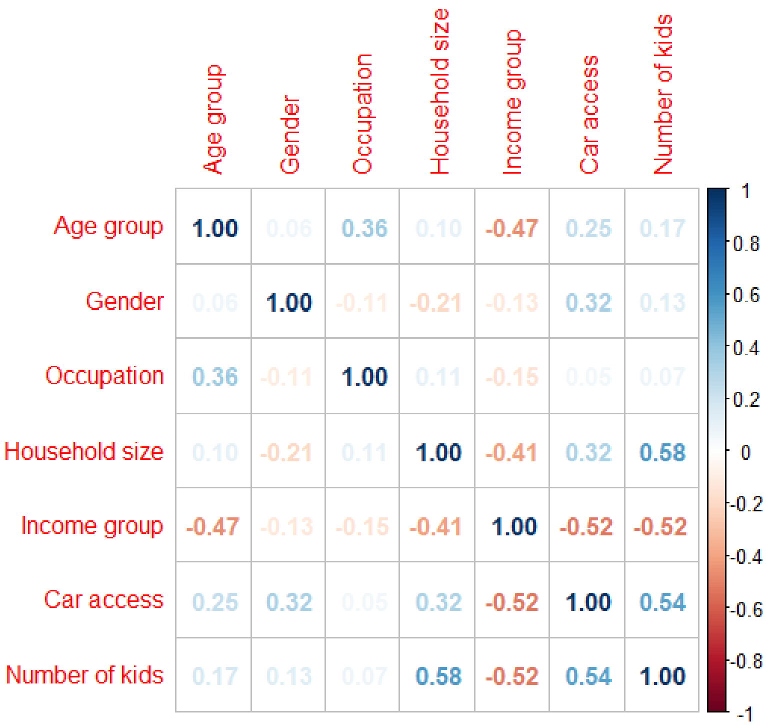


Figure 62 Pearson correlation among socio-demographic variables

Table 19 provides a summary of the coefficients from univariate linear regression models. Significant variables are bolded (Intercepts are not highlighted as they are always significant). Some variables with a significance slightly below 90% were also highlighted when the theory supports their inclusion. The detailed regression model specifications with significance and model fits are presented in Appendix 5. Each table presents the univariate linear regression models for intrapersonal variability in different indicators across three temporal scales.

Table 19 A summary of the coefficients from univariate linear regression models for intrapersonal variability

	Travel time			#Trips			Walk time			#Walk trips		
Variables	Day	Week	Month	Day	Week	Month	Day	Week	Month	Day	Week	Month
Age												
(Intercept)	0.97	0.61	0.31	0.57	0.49	0.24	1.23	0.68	0.45	0.87	0.58	0.29
18 to 24	-0.18	0.02	-0.02	0.00	0.07	0.05	0.18	0.18	0.03	0.03	0.13	0.10
25 to 34	0.08	0.10	0.01	0.03	0.10	0.04	0.17	0.23	-0.02	0.07	0.17	0.05
35 to 64												
Gender												
(Intercept)	1.13	0.82	0.35	0.62	0.68	0.24	1.47	1.06	0.44	0.99	0.88	0.32
Male	**-0.19**	**-0.23**	-0.05	-0.04	**-0.18**	0.05	**-0.17**	**-0.31**	0.00	-0.10	**-0.26**	0.03
Female												
Occupation												
(Intercept)	1.05	0.66	0.28	0.59	0.51	0.24	1.35	0.84	0.41	0.93	0.66	0.30
Student	-0.12	0.06	**0.10**	0.02	**0.18**	**0.09**	0.06	0.10	**0.07**	0.00	**0.20**	**0.11**
Worker												
Household size												
(Intercept)	1.13	0.74	0.30	0.64	0.66	0.26	1.51	0.94	0.41	1.01	0.86	0.35
1	-0.16	0.03	0.09	-0.04	0.01	**0.13**	-0.15	0.03	**0.11**	-0.09	-0.04	**0.09**
2	-0.17	-0.14	-0.01	-0.07	-0.20	-0.03	-0.21	-0.15	0.01	**-0.13**	**-0.27**	**-0.07**
More than 2												
Car access												
(Intercept)	0.90	0.53	0.25	0.53	0.41	0.22	1.42	0.76	0.41	0.95	0.58	0.30
No car access	**0.18**	**0.24**	**0.13**	**0.11**	**0.26**	**0.10**	-0.08	0.18	0.05	-0.04	**0.23**	**0.08**
Has car access												
Income group												
(Intercept)	1.03	0.70	0.32	0.60	0.56	0.27	1.31	0.84	0.41	0.89	0.68	0.33
High	0.04	-0.17	**-0.12**	-0.06	**-0.21**	**-0.07**	0.07	-0.13	-0.04	0.06	-0.14	-0.06
Low	-0.10	0.18	**0.19**	**0.09**	**0.42**	**0.13**	**0.31**	**0.39**	**0.17**	**0.27**	**0.60**	**0.18**
Medium												
Children												
(Intercept)	1.03	0.57	0.21	0.57	0.42	0.23	1.42	0.70	0.37	0.94	0.59	0.32
No children	-0.02	0.13	**0.14**	0.03	0.18	0.06	-0.06	0.20	0.09	-0.02	0.16	0.02
Has children												

Overall, the models show that most socio-demographic variables have no or weak impacts on intrapersonal variability, regardless of travel behavior indicators. The small sample sizes undoubtedly contribute to this. Nonetheless, some characteristics are significant and point to areas for future work.

Models across all travel behavior indicators have some common findings. First, age has no significant impact on intrapersonal variability. Perhaps this can be attributed to the aggregation of ages into broad categories and the small samples sizes. Intrapersonal variability is consistently lower for men, and it is significant in the models of daily and weekly variability. This is aligned with findings in previous study (Raux et al., 2016). Moreover, students have higher intrapersonal variability in travel times at the monthly scale. This may be due to school holidays and semester breaks, leading to larger changes in travel times. Finally, the presence of young children in the household also has almost no significant impact on intrapersonal variability. This may also be an artifact of the small sample of individuals with children.

Car access has the most significant impact on intrapersonal variability of total travel time and the number of trips. Not having a car is systematically associated with high variance in travel time budget and number of trips. This can be explained by people without cars who are more likely to be active travelers. The Department for Transport (Department for Transport, 2019) found that walkers and cyclists were more likely to vary their commute mode than car drivers. Therefore, people having no car tend to have more variance in travel modes, leading to more variance in travel times.

Household size has no important effects on intrapersonal variability of total travel time. However, people living alone have higher variability in month-to-month trip counts, walk trip counts, and walk time. Besides that, two-person households systemically have lower variability in number of walk trips across all temporal scales.

In contrast to our expectations, the model shows that high-income people have lower variability in monthly travel time budget and number of trips, while low-income people have higher variability in monthly travel time budget and number of trips. Low-income people are also correlated to lower variability in walking time and number of walk trips.

8.6.2. Assessment of Transport-Related Physical Activity and Its Comparison with Travel Surveys

Physical inactivity is one of the key risk factors of chronic disease. According to the World Health Organization, over half of the German adults are not sufficiently active (World Health Organization, 2018). One of the easiest ways to accumulate physical activity (PA) is through active transport including walking and cycling. Adults who use active travel are overall more physically active (Sahlqvist et al., 2012). Therefore, researchers in both transport and health science have assessed transport PA to understand its demographic and behavioral determinants.

The results in Section 8.6.1 showed that walking behavior has a highest variance over days. This indicates that a traditional one-day survey may not be adequate to capture the complexity of walking behavior pattern, particularly when calculating physical activity volume for health benefits assessment. However, travel surveys are still the most accessible and widely used data sources for transportation studies. It is important to investigate if travel surveys are suitable for assessing transport-related physical activity. Therefore, this study aims to assess walking and cycling travel behavior and to find out the strengths and limitations of different data sources. A comprehensive comparison will be conducted across four datasets: 1) One-day Household Travel Survey; 2) Mobility Panel Survey; 3) Google Location History data; and 4) Physical activity questionnaire.

8.6.2.1. Introduction of Data Sources

This study used four different data sources that were all conducted in Germany. Two data sources are self-report surveys in which households complete travel diaries designed to capture every occasion in which a household member changes location. One data source is the passive smartphone data (GLH) collected for this dissertation. The other data source is a large-scale physical activity questionnaire. A summary table of each dataset and relevant attributes for this study is given in Table 20.

Table 20 Overview of data sources used for this study

	Mobilität in Deutschland (MiD)	Deutsches Mobilitätspanel (MOP)	German Health Update (GEDA)	Google Location History data (GLH)
Year	2016-2017	2016-2018	2014-2015	2018-2019
Country	Germany	Germany	Germany	Germany
Sample ages	All ages	10 years or older	18 years or older	18 - 64
Sample size	316,361	11,225	24,016	27 (1,668 weeks)
How is walking/cycling recorded?	One-day trip diary	Seven-day trip diary	Questions	GPS tracking

***Mobilität in Deutschland* (MiD)** is Germany's most recent national travel survey (BMVI, 2019b). It is a large cross-sectional survey conducted in 2016-2017 consisting of 156,420 households containing 912,002 trips. Respondents complete 24-hour travel diaries to give on their interactions with the transport network over the course of a day. Trips recorded in MiD with main mode walking and cycling can be used to investigate transport-domain PA. However, distances and travel times are self-reported, limiting the precision of PA estimates. In addition, there is limited information on walking and cycling during trips where another mode is the main mode, such as access and egress to public transport.

The *Deutsches Mobilitätspanel* (MOP) is a nationwide German panel survey conducted every year (BMVI, 2019a). Around 650 households are recruited per year and respondents stay in the panel for 3 years. Each year respondents complete a 7-day travel diary. For this study the MOP is used as a cross-sectional survey in which household IDs are unique each year. Thanks to the 7-day diary period we can capture transport-domain PA more representative of respondents' overall behavior. However, MOP suffers from the same limitations as MiD in that trip details are self-reported and walking and cycling is difficult to capture if not the main mode.

Google location history (GLH) data were introduction in Section 8.1. The 27 valid GLH datasets were used in this study. The semantic location history data that contains the information of travel mode, travel time and travel distance is employed for measuring walking and cycling behavior.

The German Health Update (GEDA) is a cross-sectional health survey (Lange et al., 2017). It is conducted with a large sample size in the German adult population on a regular basis. The latest data which is accessible for research purposes is from the wave of 2014-2015. The European Health Interview Survey (EHIS) PA questionnaire is completely integrated into the GEDA 2014-2015. A total of 24,016 persons aged 18 years or older participated. Transport-related walking and cycling are assessed separately by asking respondents the questions about their time spent on walking/cycling in a typical week:

1) On how many days in a typical week do you walk/cycle at least 10 minutes without stopping to get from place to place?

2) On a typical day, how long do you walk/cycle to get to the place?

8.6.2.1. Assessment and Comparison

Four different datasets were compared against the key indicators presented in Table 21. Travel measures for walking and cycling were compared separately. Due to the nature of GEDA questionnaire, trip-level indicators (e.g., mean walk trips per walker/cyclist, mean walk distance per walker/cyclist, and mean walk time per walker/cyclist) were only compared among MiD, MOP and GLH. Walk and cycle time in GEDA were recorded at week-long level in groups. In GEDA, physical activity is defined as walking/cycling at least 10 minutes without stopping. In other to have a fair comparison, here walkers/cyclists are defined as those with at least one walk/cycle trip over 10 minutes reported in the day or week. In addition, the assessments were only compared among people aged from 18 to 64 in order to compare datasets under the similar context.

Table 21 Overview of indicators used for comparison

		MiD	MOP	GLH	GEDA
Travel measures (walk)	Share of walkers	x	x	x	x
	Mean walk trips per walker	x	x	x	
	Mean walk distance per walker	x	x	x	
	Mean walk time per walker	x	x	x	
	Mean walk time per walker (In group)	x	x	x	x
Travel measures (cycle)	Share of cyclists	x	x	x	x
	Mean cycle trips per cyclist	x	x	x	
	Mean cycle distance per cyclist	x	x	x	
	Mean cycle time per cyclist	x	x	x	
	Mean cycle time per cyclist (In group)	x	x	x	x
Physical activity measure	Transport-related physical activity in MET-hours/week	x	x	x	x

Figure 63 compares walkers' daily and weekly travel behavior across four datasets. The share of walkers (i.e., those with at least one walk trip reported in the day or week), is lower in MiD and MOP than GLH. In other words, MiD and MOP overestimate non-active people.

At the daily level, mean number of walk trips are similar across three datasets, while MiD and MOP recorded longer walk distances and walk times than GLH. This might reveal that walk trips recorded in MiD and MOP are usually long-distance walk trips and short walk trips are poorly captured. Also, due to the fact that MiD and MOP are self-report surveys, respondents tend to overreport travel times and distances of their walk trips.

When it comes to the weekly indicators, MiD shows a much higher mean walk trip and walk distance per walker. This is because MiD only records trips on a single day and the data were multiplied by a factor of 7 to get the weekly walking behavior. People who actively travelled on the survey data are assumed to have the same active level on other days. Without capturing the interpersonal behavior variability, MiD tends to overmeasure walkers' walk trip frequency, distances and times. Thanks to the week-long nature of MOP, it doesn't show overestimation issues, but still MOP captures fewer walk trips at the week level compared to GLH and the mean walk distance is compensated by trips with longer distances.

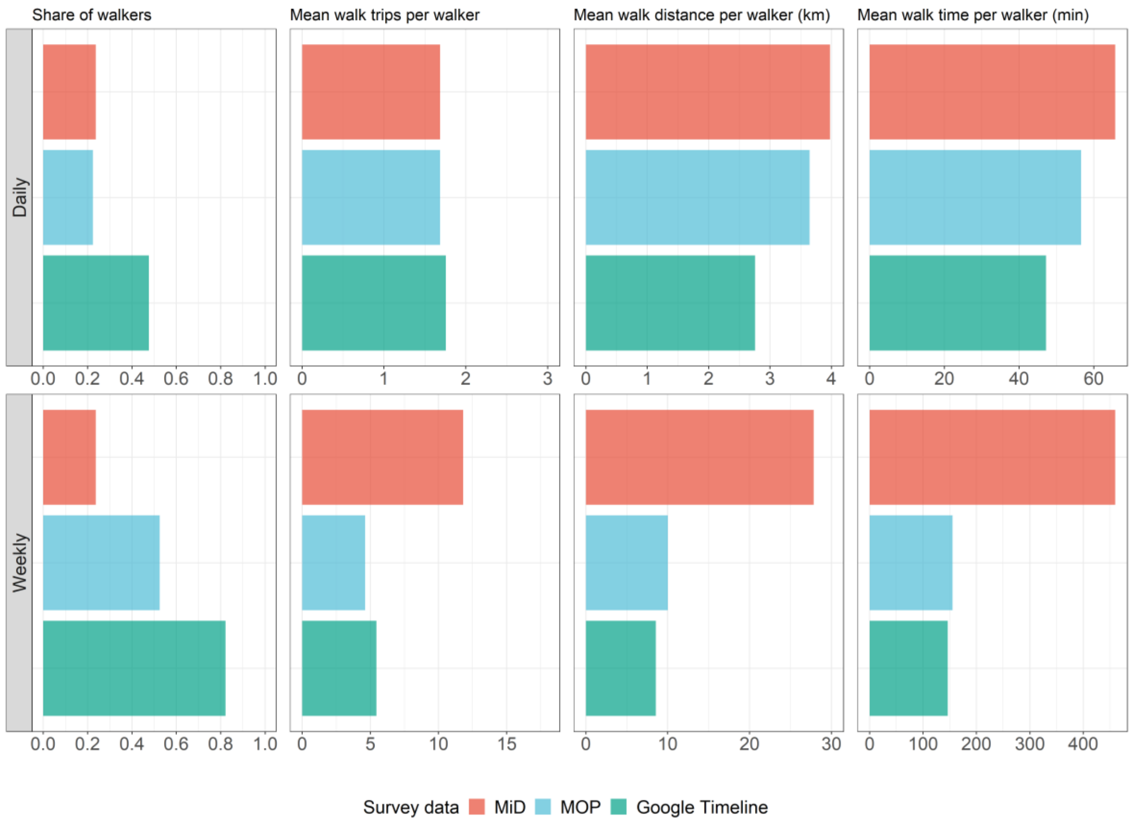


Figure 63 Comparison of walkers' travel behavior at daily and weekly levels across MiD, MOP and Google Timeline datasets

Figure 64 compares the weekly walk time across four datasets. First, without capturing the interpersonal behavior variability, people in MiD have either no walk time or extremely high walk time during a week. Compared to the shares in GEDA, MOP also has a larger share of non-active people. The shares resulted from GLH data are aligned with those in GEDA, except that the share of non-active people is a bit smaller than GEDA. While it is hard to tell what is the ground truth of individuals' weekly walk time, it is obvious that MiD is not suitable for assess people's physical activity level due to the high share of non-active people and less variance across populations.

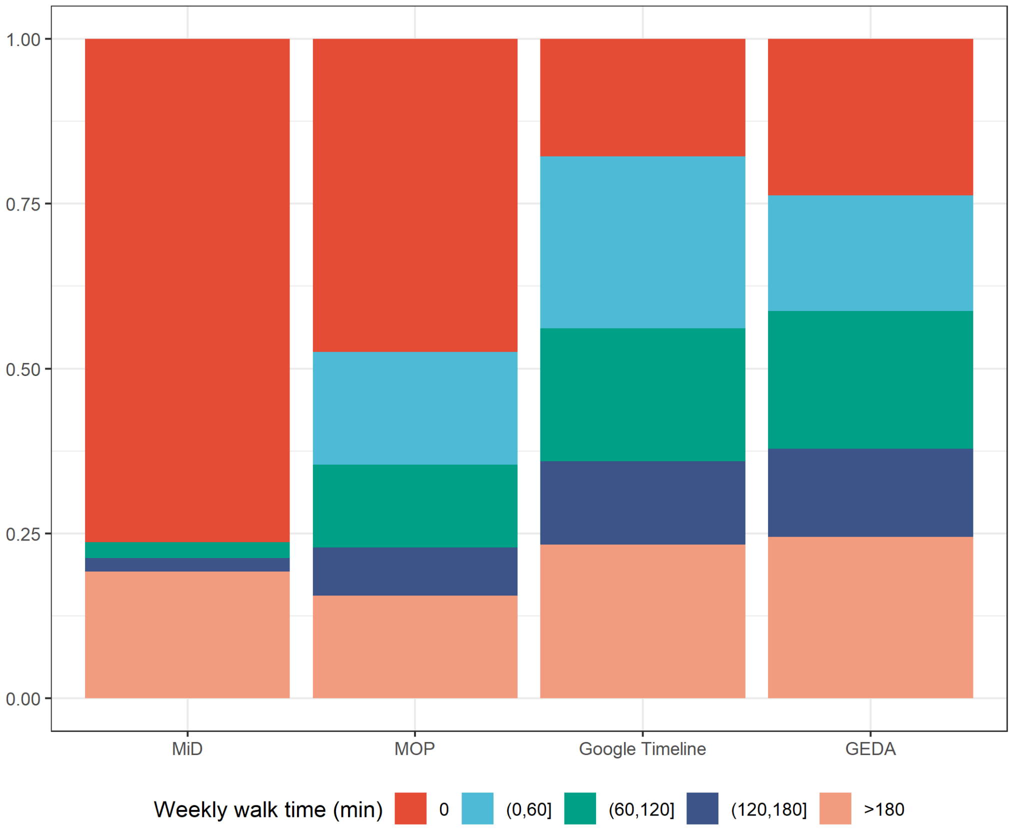


Figure 64 Comparison of weekly walk time across MiD, MOP, Google Timeline and GEDA

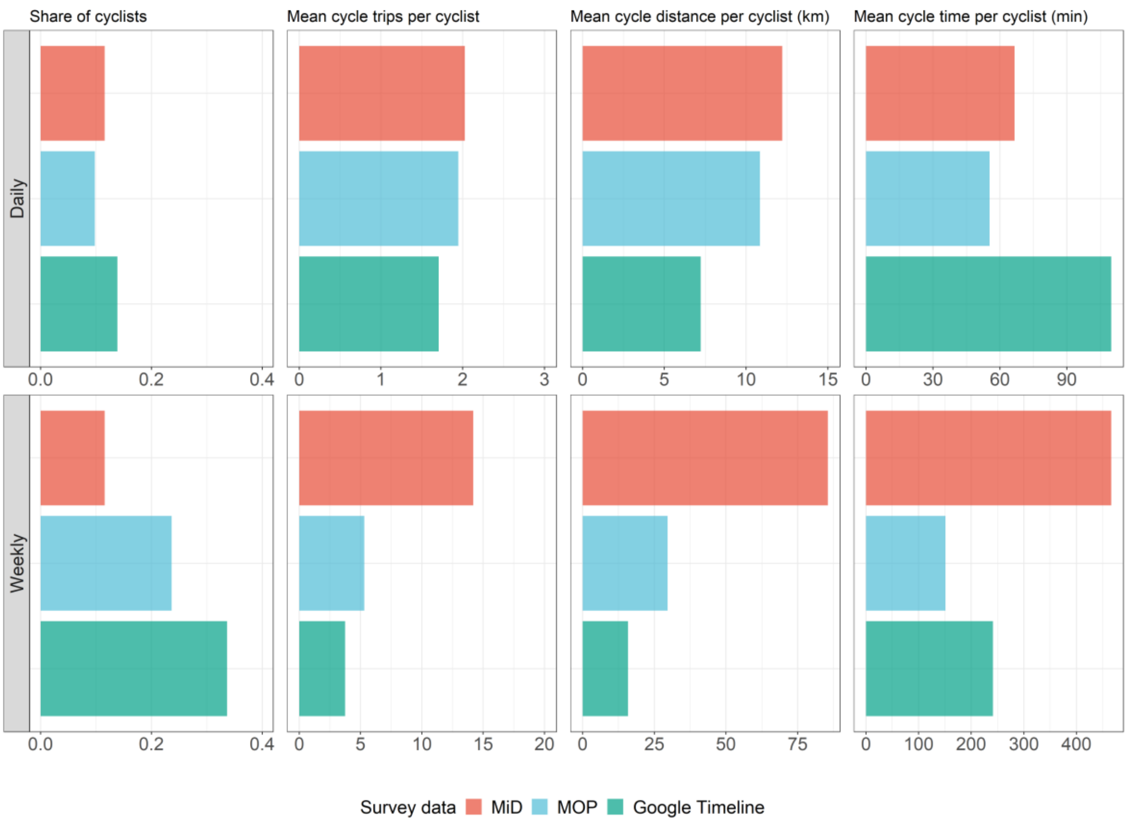


Figure 65 Comparison of cyclists' travel behavior at daily and weekly levels across MiD, MOP and Google Timeline datasets

As mentioned in section 8.5, cyclists are underrepresented in the collected GLH data. Therefore, the comparison results are less intuitive and convincing. Figure 65 shows that at the daily level, GLH data had a higher share of cyclist, less distance cycled per cyclist but much higher cycling time per cyclist. At the weekly level, MiD again shows a much higher mean cycle trip and cycle distance per walker. The differences between MOP and GLH were also large. Figure 66 presents the difference of weekly cycle time distribution across four datasets. MiD still has the same issue as it had for walk travel behavior. The differences among MOP, Google Timeline and GEDA are minor since the majority of the population has no weekly cycle time.

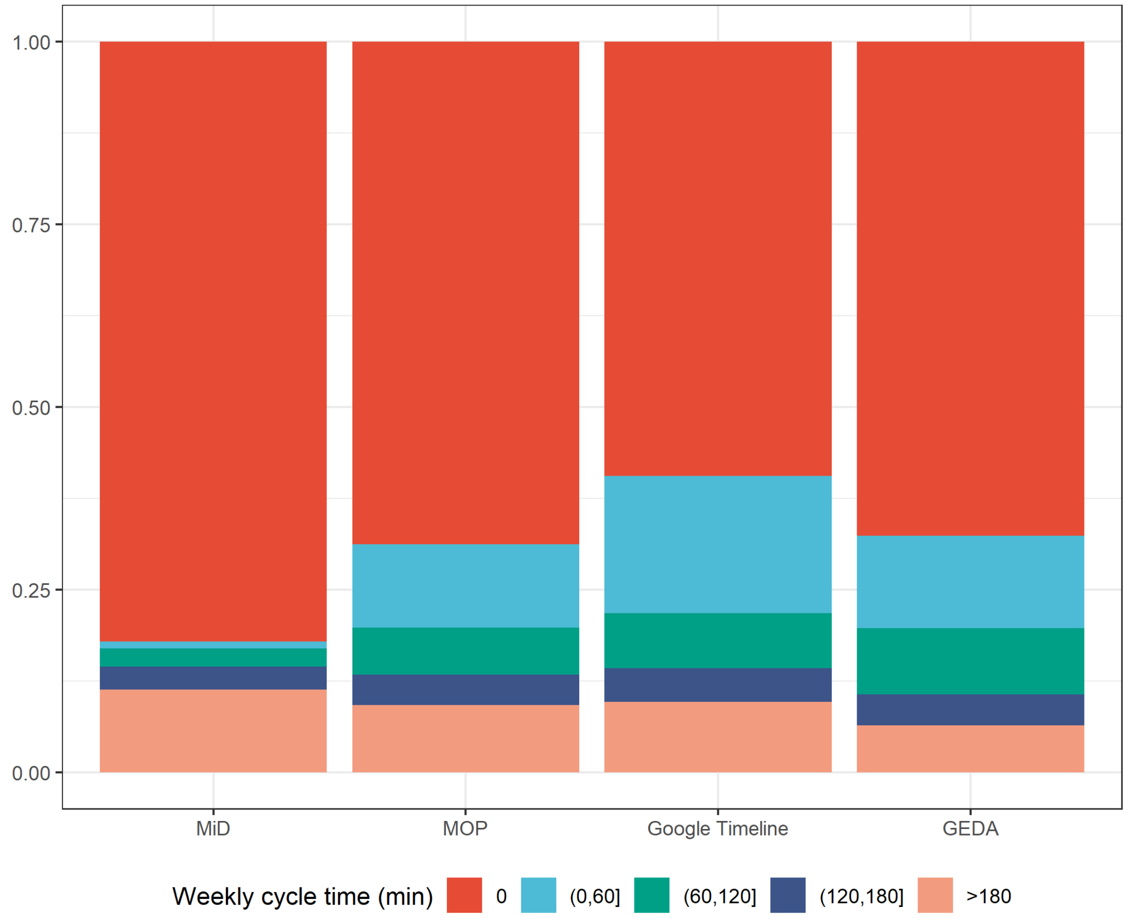


Figure 66 Comparison of weekly cycle time across MiD, MOP, Google Timeline and GEDA

World Health Organization (WHO) recommends that adults accumulate at least 150 minutes of moderate-vigorous physical activity (MVPA) per week. Walking and cycling for transport qualify as being at least moderate intensity activity. Therefore, it is important to assess physical activity level based on four datasets. Figure 67 presents weekly MET/hour distribution across four datasets. These results support the hypothesis that a 1-day diary is not representative of the weekly distribution of transport-domain PA. The proportion of inactive individuals is overestimated, while the number of individuals in the central PA levels is underestimated. The proportion of individuals in the highest PA level is slightly overestimated.

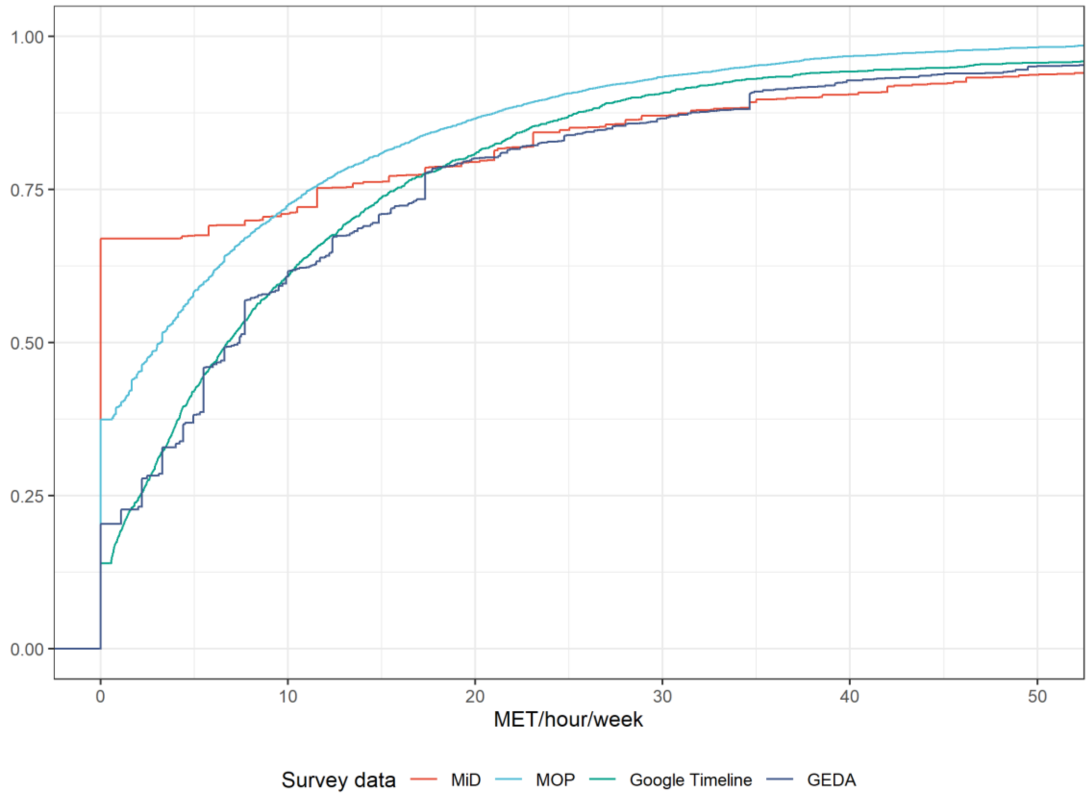


Figure 67 Comparison of weekly MET/hour distribution across MiD, MOP, Google Timeline and GEDA

find out if disagreements and bias exist in some specific person groups

Furthermore, the physical activity level (MET/hour/week) of different datasets were also compared in relation to subgroups in order to find out if disagreements and bias exist in some specific person groups. Figure 68 and Figure 69 compare weekly MET/hour distribution by age group and gender. It indicates that the disagreement is larger in the age group of 18 to 24, may suggesting that people those who from 18 to 24 are poor in report trip diaries in travel surveys.

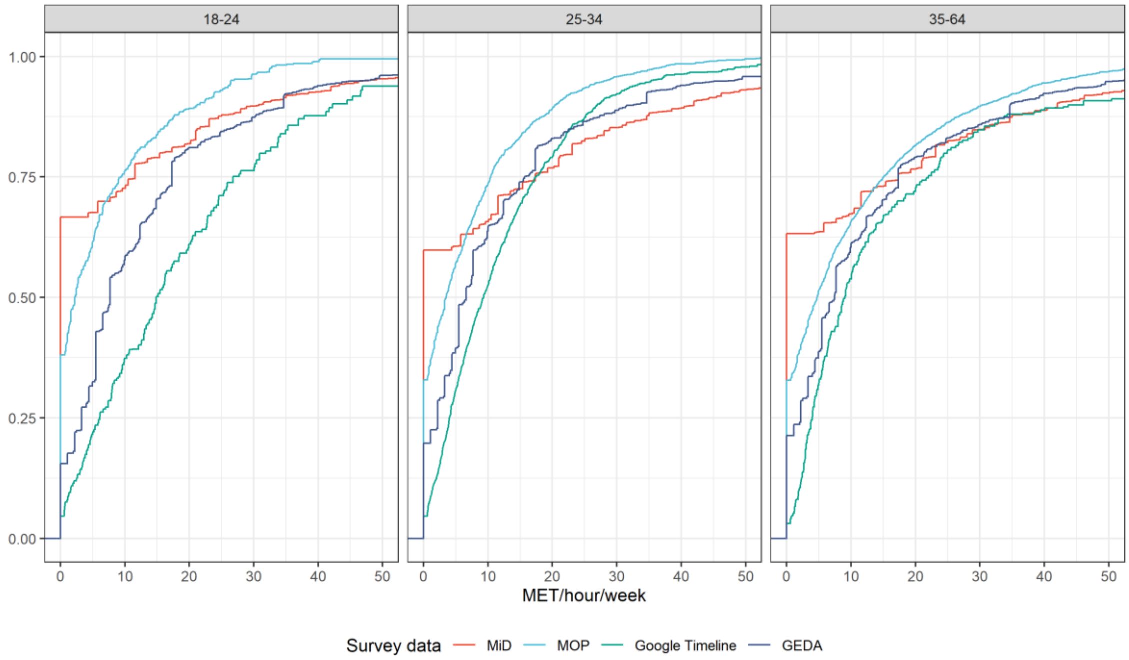


Figure 68 Comparison of weekly MET/hour distribution by age group across four datasets

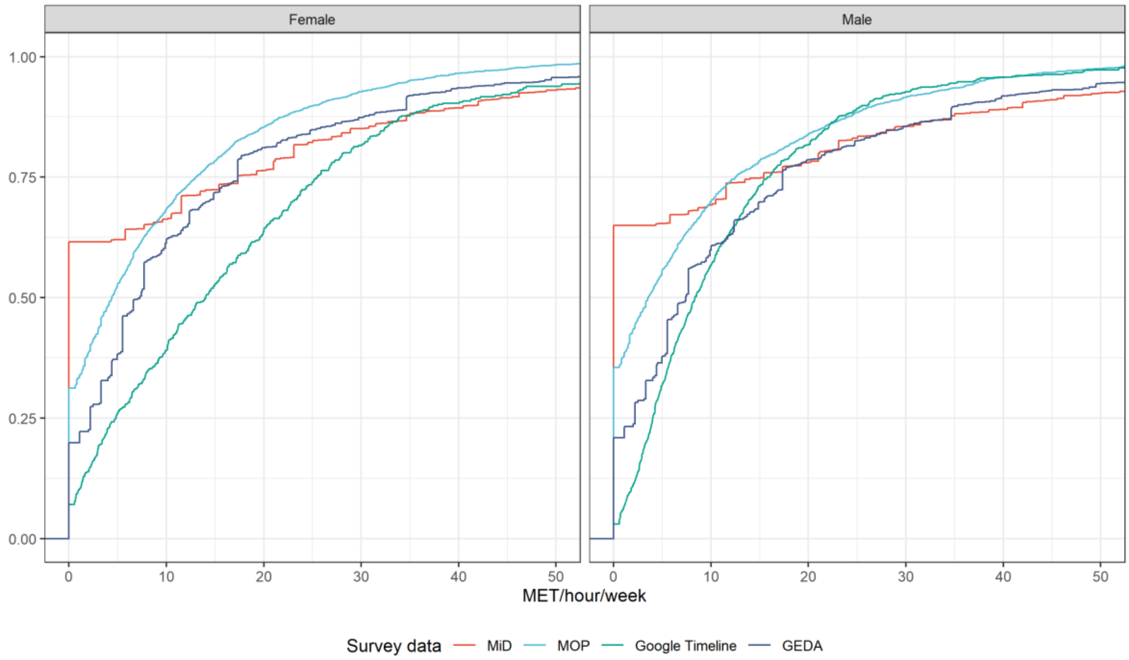


Figure 69 Comparison of weekly MET/hour distribution by gender across four datasets

8.6.3. Determinants of Weekly Travel Behavior

In previous sections, the analysis of travel behavior variability proved that individuals have a great deal of day-to-day variability. Week-to-week travel behaviors have relatively low dispersion. Weekly walk and cycle time are also the common metrics in health research for assessing physical activity volumes. Therefore, it is important to have a closer investigation on weekly travel behavior. This study attempts to find out the potential determinants of weekly walk time, weekly cycle time, as well as weekly physical activity volumes. The following analysis addresses the question of how factors such as personal attributes and weather influence individuals' weekly walking and cycling time.

The 27 valid GLH dataset was used in this analysis. In total, 1667 weeks were recorded. The 27 individuals with varying weeks consist of the panel data for estimation. Linear panel regression models with random effects are employed in this study. The regression equation of panel data with random effects is shown below:

$$y_{it} = \beta_0 + \beta * X_{it} + \alpha_i + U_{it}$$

Where X_{it} is a set of explanatory variables; α_i is the random term capturing unobserved individual factors; U_{it} are the error terms varies by individual and time.

This model estimates panel data where interference variables may be associated across time and across individuals. The difference between intercepts is accommodated by the error terms of each individual. The advantage of using the random effect model is to eliminate heteroscedasticity in individuals and over time.

Three regression models were estimated for dependent variables, including weekly walk time in minutes, weekly cycle time in minutes and weekly physical activity volume in mMET. The weekly physical activity volume was measured by the following equation:

$$PA_i = \sum_{d=1}^{days} \left(\sum_{w=1}^{walk\ trips} \frac{distance_{d,w}}{4.8} * 3.61 + \sum_{c=1}^{cycle\ trips} \frac{distance_{d,c}}{13.9} * 5.44 \right)$$

The explanatory factors included in the models are:

- Socio-demographic characteristics (e.g., age, gender, occupation, and household size)
- Travel-related information (e.g., a set of binary variables about car access, bike access, car share membership and bike share membership)

- Weather-related attributes (e.g., weekly average precipitation in millimeters, weekly average temperature in Celsius, and weekly average snow depth in centimeters)

- Time-related variables (e.g., month, has public holidays in the week or not).

The weather data in this study was obtained from Deutscher Wetterdienst (DWD). The DWD provides open-source data on a wide range of historical and current climate information. The DWD platform was used to download daily climate data from 107 weather stations across Germany for the year 2018 and 2019. Then each trip was mapped to the closet weather station and the matched date to get the precipitation, temperature, and snow depth on that day.

By analyzing the model performance and significance, some of the variables were excluded and the final model specifications are presented in Table 12 Some variables with a significance slightly below 90% were included in the final estimation if theory supported their inclusion.

Individuals aged from 25 to 34 are less active in both walking and cycling. Males and students tend to walk more but cycle less though the influence is minor. Having a car access is a significant to being active. In average, those with car access walked 51 minutes less and cycled 15 minutes less than those without car access. Having bike access increased weekly cycle time by 42 minutes. It is interesting to note that people who use car share tend to walk more. This could be due to the short walking trips to the car. Households with a larger household size tend to walk more but cycle less, which makes sense because walking in a group is easier than cycling together. Walking and cycling time are negatively associated with weekly average precipitation. Heavy rain (defined as a precipitation rate over 7 mm) can reduce walking time by 5.5 minutes and cycling time by 8.2 minutes. The snow depth showed a positive influence on walking and cycling, which is not intuitive. This might be because of the rare snowing days throughout the year, as well as the fact that snow depth in Germany is generally not significant. Having public holidays is a strong predictor for walking time while has almost no impact on cycling time. Walking and cycling time also varied across different months. From August to November, people are generally more likely to walk. This might be because of the vacation season. It is surprising that months such as June and July have a negative correlation with walking time.

In general, the weekly walk time and weekly PA volume models have relatively good R^2, whereas the goodness-of-fit of the weekly cycle time model is poor. This could be due to the small number of cycling trips represented in the collected GLH data.

Table 22 Estimation results of linear panel regression models

	Weekly walk time			Weekly cycle time			Weekly PA volume		
	θ	Pr(>|z|)		θ	Pr(>|z|)		θ	Pr(>|z|)	
Intercept	47.71	0	***	18.76	0	***	5.43	0	***
Age: 25-34	-23.54	0	***	-10.82	0	***	-2.54	0	***
Gender: male	1.06	0	***	-9.73	0	***	-1.02	0	**
Occupation: student	-0.59	0	***	-7.69	0	***	-1.1	0	***
Has car access: yes	-51.52	0	***	-16.47	0	***	-4.68	0	***
Has bike access: yes	-17.36	0	***	43.44	0	***	2.93	0	***
Has car share membership: yes	2.05	0	***	-5.55	0	***	-0.5	0.09	.
Has bike share membership: yes	-15.85	0	***	33.61	0	***	2.43	0	***
Household size = 2	1.44	0	***	-17.19	0	***	-2.1	0	***
Household size > 2	7.35	0	***	-21.13	0	***	-1.9	0	***
Weekly average precipitation	-0.79	0	***	-1.17	0	***	-0.16	0	***
Weekly average snow depth	0.13	0	***	0.01	0	***	0.01	0	***
Weekly average temperature	n.s.			0.41	0	***	0.04	0	***
Has public holiday: yes	14.82	0	***	0.37	0	***	0.3	0	***
Weekly total number of trips	4.39	0	***	-11.1	0	***	-0.17	0.08	.
Month: February	-7.03	0	***	-6.13	0	***	-1.01	0	***
Month: March	-6.82	0	***	-4.67	0	***	-0.87	0	***
Month: April	-3.6	0	***	17.99	0	***	1.55	0	***
Month: May	-3.14	0	***	13.68	0	***	1.18	0	***
Month: June	-11.92	0	***	4.58	0	***	n.s.		
Month: July	-16.02	0	***	5.8	0	***	n.s.		
Month: August	11.78	0	***	-4.3	0	***	0.37	0.1	
Month: September	17.36	0	***	-1.63	0	***	0.92	0	***
Month: October	15.66	0	***	-6.17	0	***	0.35	0.04	*
Month: November	3.79	0	***	-1.82	0	***	n.s.		
Month: December	-9.63	0	***	-5.27	0	***	-1.09	0	***
R-Squared:	0.48			0.21			0.42		

Note: Pr(.|z|) = the level of significance; ***, **, *, '.' = 99.9%, 99%, 95% and 90% level of significance, respectively

8.6.4. Impacts of Life Events on Weekly Travel Behavior

Changes in travel behavior can be influenced by a variety of factors. The previous section explored its association with weather that changes on a daily basis. Some circumstances that change occasionally during life, such as household relocation, car/bike purchases and changes of employment status, may also have influences on the change of travel behavior.

In this section, the investigation tried to look at how major life events influence weekly travel behavior. However, due to the limited number of life events observed among the survey participants, the analysis was conducted through a series of descriptive statistics.

Appendix 6 contains descriptive analyses of changes in travel behavior by different event types. The changes in monthly travel time, monthly walk time, monthly cycle time and monthly PA volume were calculated for each event case. Take monthly walk time as an example, changes in monthly walk time are defined as the percentage increase/decrease in walk time by comparing the average monthly walk time before and after event occurrence. Average monthly walk time is defined as the mean value of walk time over three months before/after the month of occurrence. Since behavioral change caused by life events usually has a transition phrase, the walk time generated in the month of occurrence was not used for capturing travel behavior change. According to this calculation method, the GLH data for analyses needs to have at least three successive months (the month of occurrence, month of occurrence +1 and month of occurrence -1). However, some of the respondents did not have a three-month-long GLH data and were excluded from this analysis. Finally, 32 out of 90 recorded event cases were presented in Appendix 6.

The time-varying factor such as weather could also have joint influences on travel behavior changes. Due to the rare samples in each event type, factors such as weather and months were not controlled in the analysis. It is noted that the change in travel behavior could be affected by multiple life events. Overall, this analysis only serves as a demonstration of how to use the collected GLH data for exploring the relationship between major life events and changes of travel behavior. There is no intuitive conclusion that can be drawn based on such a small sample size.

8.6.5. Defining Travel Routines and Disruptions Using Machine Learning Approaches

The theory-driven approach mentioned in Section 5.4.1 proved that sociodemographic characteristics can to some extent explain the change in travel behavior. However, travel behaviors are complex and are influenced by many factors. Some factors can change daily, including the built environment, weather, and availability of mode, while the other factors may change occasionally during life such as life-changing events (e.g., household relocation, car/bike purchase) and social intervention. Socio-demographics are only a small part of the determinants. Also, the evidence of sociodemographic is less successfully found when using more complex behavioral measures (Raux et al., 2016).

Recently, the data-driven machine learning approach has been popular for understanding unobserved and complex phenomena since it does not rely on explicit knowledge and theory. Making use of the advances in data science, this section employs clustering algorithms and machine learning techniques to examine variations in travel over various periods based upon the Google Location History (GLH) data for a small sample of individuals. This analysis aims to define the concept of routine behavior as well as identify disruptions to this routine. Given that this research observes travel behavior with a large number of location points over years, sufficient data are available to apply these and other machine learning techniques. For example, clustering algorithms, such as those widely used in travel behavior research to explore activity patterns (Cui et al., 2018; El Mahrsi et al., 2017), will be used to group days with similar travel characteristics based on mode usage and time of day. The resulting clusters represent the regularity in travel behavior, while the difference between clusters represents some level of disruption.

8.6.5.1. Methodology of Cluster Algorithm

Cluster analysis is a machine learning technique to find the natural grouping of data points so that the points in the same cluster have similar characteristics. It is widely used for pattern recognition in many fields. In this research, we consider each day as a data point, then use the clustering algorithm to group the days with travel patterns into one cluster considering activity types, departure and arrival times at activities, and activity duration. Here are the steps of the data analysis:

- Data filter and process
- Data transformation: generate a time-activity binary matrix for each day
- Similarity measurement of each day pairs S_{ij}
- Cluster formation
- Cluster interpretation and evaluation

First, GLH data are processed to be ready for the following analysis. For example, we convert time steps from Unix epoch format to standard date format in seconds. Moreover, activity types are defined for all trips as the type with the highest confidence. Afterward, we create a time-activity type map for each day. As seen in Figure 70, activity types are categorized into four groups, and time is recorded at a rather fine-scale (in one-minute bins) to also capture short walk trips.

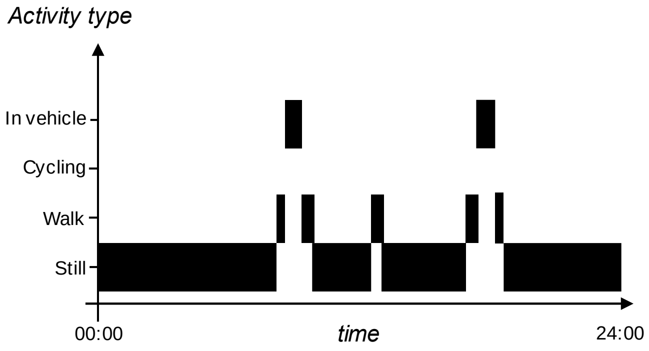


Figure 70 Time-activity type map of an example day.

After that, we convert the maps into binary matrices with 4 [activities] x 1440 [1-minute time step] entries (Equation 1).

$$\begin{bmatrix} x_{11} & \cdots & x_{1j} \\ \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} \end{bmatrix} \tag{1}$$

where x_{ij} is a binary variable. x_{ij} equals 1 when activity type i happened at time bin j; otherwise x_{ij} equals to 0.

The measurement of the distance between two data points is crucial for the clustering algorithm. In this study, the distance of points is defined as the similarity of the binary matrix of two days. The Jaccard similarity coefficient is used to measure similarities for binary matrices (Equation 2).

$$J = \frac{M_{11}}{M_{01}+M_{10}+M_{11}} \tag{2}$$

where M_{11} represents the total number of cells where two days both have a value of 1. M_{01} and M_{10}represent the total number of cells where one matrix has a 0 and the other matrix has a 1. Finally, we generate a similarity matrix for all the point pairs (Equation 3). The matrix is mirrored (i.e., $S_{1,2}$ is the same as $S_{2,1}$).

$$\begin{bmatrix} 0 & S_{1,2} & \cdots & S_{1,j-1} & S_{1,j} \\ S_{2,1} & 0 & \cdots & S_{2,j-1} & S_{2,j} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ S_{i-1,1} & S_{i-1,2} & \cdots & 0 & S_{i-1,j} \\ S_{i,1} & S_{i,2} & \cdots & S_{i,j-1} & 0 \end{bmatrix} \quad (3)$$

There are many common cluster algorithms that can be utilized. Here, we use the graph-theoretic algorithm for clustering data points. The key is to construct a similarity graph from a similarity matrix. Each day denotes a node i in the graph. Every two nodes are connected by an edge when the Jaccard similarity S_{ij} is above some threshold. After generating a similarity graph, a cutting algorithm is used to cluster nodes into several groups.

To interpret the regularity in travel behavior, we will analyze several characteristics of the nodes in the same cluster, such as travel characteristics, weather, time-related attributes (month and public holidays), spatial-related attributes (Walk score, population density, accessibility). The difference between clusters represents some level of disruption. Special attention will be given to sudden changes in the similarities over time. For example, someone might have walked to reach activities most of the time, but at some point, switches to driving as the predominant mode. The questionnaire from the respondents will reveal if life-changing events, such as a move from the city to the suburbs, triggered this behavior change.

8.6.5.1. Analysis Examples of Cluster Algorithm

The methodology mentioned above was applied to all respondents. In this section, we take respondent sample 5 as an example to show the analysis process. First, an activity map is generated for all the observed days (see Figure 71) and then we transfer the activity map to a 1 x 5760 (= 4 x 1440) binary vector to represent the activity chain for each day.

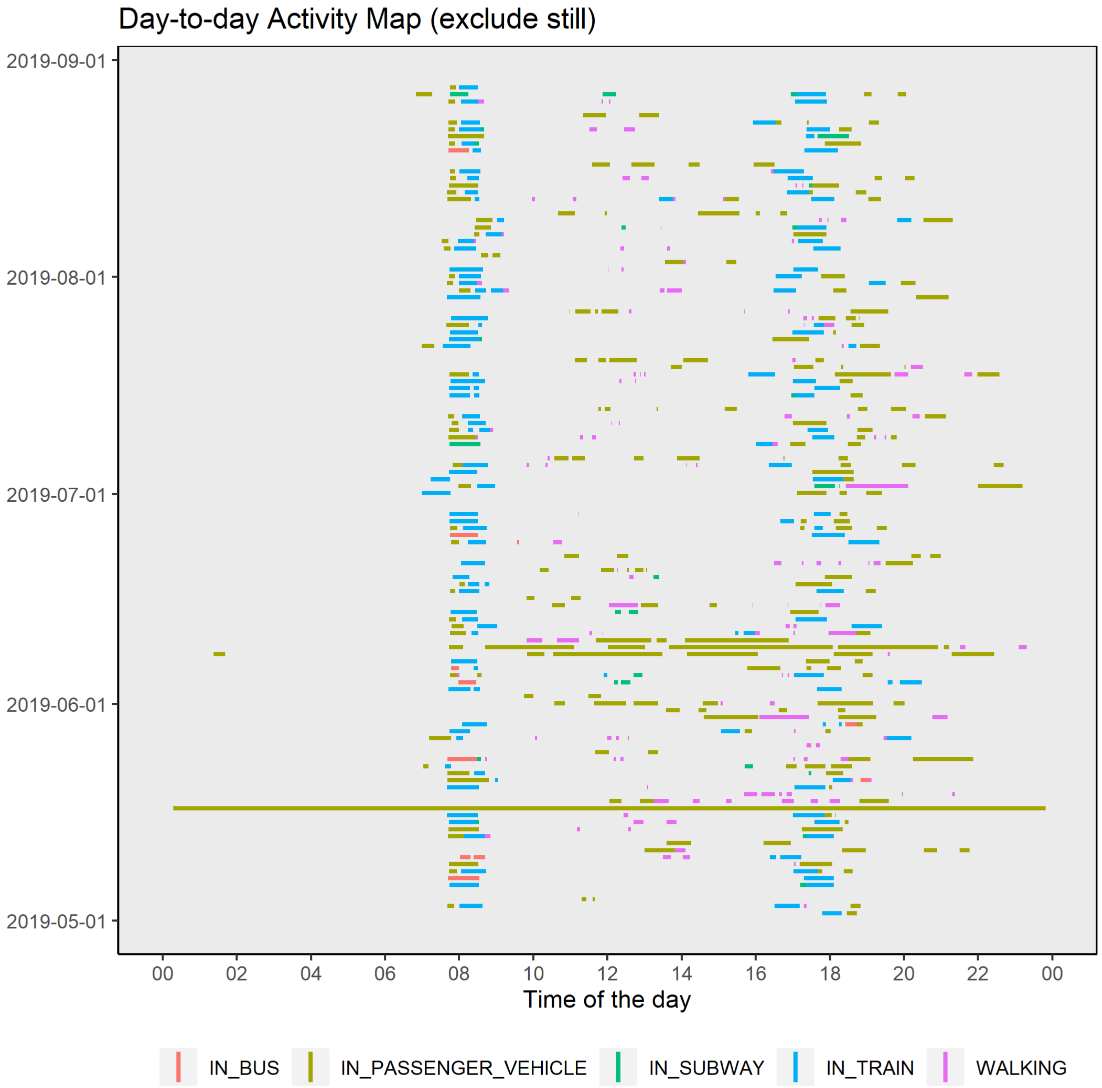


Figure 71 Time-activity type map of sample 5

Then, the Jaccard similarity index is measured for every pair of days (see Figure 72). In this case, every two nodes are connected by an edge when the Jaccard similarity is above the threshold of 0.8. The appropriate threshold is defined by a series of experiments. The rule of thumb is that the number of clusters and the node in each cluster is reasonable.

Day	123	124	126	127	128	129	...	234	235	236	238	239	240
122	0.83	0.92	0.86	0.86	0.82	0.85	...	0.82	0.80	0.85	0.83	0.76	0.88
123		0.85	0.84	0.85	0.86	0.84	...	0.81	0.85	0.78	0.86	0.76	0.90
124			0.86	0.86	0.84	0.85	...	0.82	0.85	0.92	0.85	0.79	0.93
126				0.99	0.92	0.99	...	0.89	0.84	0.79	0.94	0.84	0.92
127					0.91	0.98	...	0.90	0.85	0.79	0.93	0.83	0.92
128						0.92	...	0.89	0.83	0.78	0.94	0.83	0.89
129							...	0.89	0.84	0.79	0.94	0.85	0.91
...								...	...	...	...	...	...
234									0.80	0.76	0.88	0.76	0.87
235										0.78	0.84	0.75	0.90
236											0.79	0.74	0.86
238												0.84	0.90
239													0.81

Figure 72 An example of the Jaccard similarity matrix of sample 5

In the third step, the nodes and edges are fed into Gephi, and the community detection algorithm is conducted to define the modularity class of each node. For this dataset, four clusters are detected (see Figure 73).

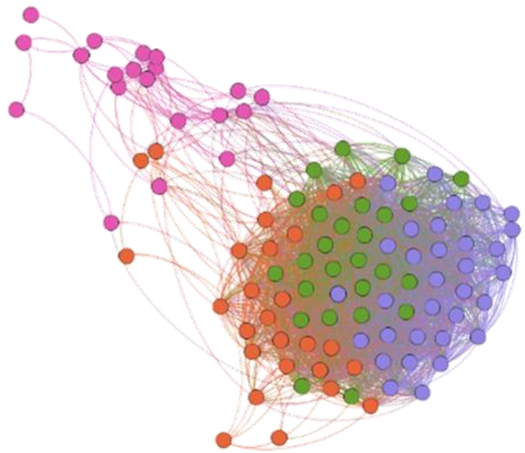


Figure 73 The visualization of clusters generated in Gephi

After cluster formation, several characteristics are summarized to interpret the similarity within clusters and the difference between clusters. From Table 23, we can see that cluster 1 represents the weekend pattern with a late start time of the day and higher travel time. Cluster 2 represents the weekdays with a high share of walking activities. Based on these characteristics, we cannot tell the significant difference between clusters 3 and 4. The days in cluster 3 and 4 are all weekdays with relatively less travel time and low share of walk trips. The start time of the day in cluster 4 is slightly earlier than that in cluster 3. There could be some unobserved factors that can distinguish the days in clusters 3 and 4.

Table 23 Day and travel behavior characteristics of each cluster

Cluster	1	2	3	4
Cluster size	22	32	20	31
Number of weekend days	18	2	0	0
Number of public holidays	1	0	0	0
Mean still time (min)	1283	1323	1329	1333
Mean total travel time (min)	156	117	107	108
Mean total trips	5	6	4	5
Mean walk time (min)	13	15	6	7
Mean walk distance (m)	770	1006	470	472
Mean share of walk trips	0.16	0.28	0.16	0.19
Mean start time of the day (hour)	10.84	7.90	8.23	7.88

8.7. Discussion and Limitations

The collected GLH data is a novel and rich source and helped to better understand the travel behavior variability over days, weeks, and months. However, the work presented in this section has some important shortcomings that limit the ability to draw broad conclusions or examine correlates of behavioral variations. The foremost among these is the small sample size of participants. Given the novelty of these data and the detailed spatial travel patterns, it is no surprise that many potential participants were reticent to engage in our study. In future studies, this could be mitigated by providing potential recruits with more examples of how these data would be analyzed, including the analyses in this dissertation. Offering recruits some compensation for their time, effort, and data might also increase the likelihood of participation.

The analysis of GLH data confirmed that HTS was poor at capturing walking activities and biased in self-reported travel distance. Although limitations exist in the traditional HTS, this work did not use the GLH data to estimate econometric models (such as walk mode choice models and walk destination choice models) for the integrated model presented in Chapter 6. There are several reasons for not applying GLH data for model estimation in pedestrian travel demand models.

The main reason is the small sample size of participants. The limited size of GLH data was a bottleneck to estimate stable econometric models. To illustrate the variation across population, it is important to add essential socio-demographic attributes such as age and gender to the models. However, it was generally not possible to control age and gender in the models because of the limited number of records available.

While the samples were all recruited from the Munich region, they were not representative of either the Munich population or the German population. The cross-data comparison between survey respondents and German census data in Section 8.3.1 revealed that GLH samples are skewed towards participants who are aged 25 to 34. Also, due to ethical considerations, young children or teenagers (under age 18) were not included in the survey recruitment. These sample biases restricted the use of GLH data to the integrated model that simulates the travel demand across the entire synthetic population.

More importantly, HTS data remains the most commonly used data source for travel demand models. We cannot overlook the strength of HTS, such as large sample size, better representation of population distribution, and more accessible to researchers. As a result, it is more important to inform about HTS limitations rather than to replace them in the travel demand models.

The research suggests that HTS can be validated against objective methods. Because of the small sample size in this study, the German HTS could not be validated by using GLH data. Once the GLH data is collected with a larger sample size and better representation of population distribution, cross-data validation can be conducted for different person groups in terms of travel metrics such as trip generation, travel time, and travel distances. Calibration factors can be derived for each person group and then used to modify the self-reported travel information in HTS. Furthermore, GLH data can be used to calculate the variability index of different person groups. Then the variability index can be used to impute an individual's week-long travel behavior based on the one-day HTS.

Overall, the focus of this study is to explore the capabilities of using Google Timeline data for travel behavior studies. It is not suitable for exploring the variations across the population, while it has great potential for investigating the variations within an individual. The findings from this analysis reveal some insights into travel behavior and perhaps more importantly provide some better guidance on the design of future data collection efforts and the utility of GLH for transportation analysis.

9. Conclusion

This dissertation improved pedestrian travel demand modeling in terms of model accuracy, sensitivity, transferability, and efficiency. It targeted to advance pedestrian modeling both theoretically and practically. First, it enhanced the stand-alone pedestrian model MoPeD in different aspects, including run-time upgrade, reconstruction of new pedestrian accessibility measures, and enhancement of the walk destination choice model. Next, the integrated modeling framework that incorporates MoPeD into MITO was developed to provide more accurate travel outcomes for evaluating policies and scenarios. Finally, novel data on travel behavior variability was collected. Detailed data analyses examined travel behavior variability, identified the limitations of HTS, examined potential determinants of weekly travel behavior, and investigated individuals' travel routines and disruptions by using machine learning techniques.

While the key findings and limitations were discussed at the end of each chapter, this section will outline them in a more general view and focus on some broader issues. This chapter begins by highlighting some key findings and lessons learned in each part of the dissertation. Then, Section 9.2 summarizes the contributions and implications of this study, including innovations in data, fundamental research in the travel behavior field, and advances in modeling approaches. The limitations of the work and possible improvements in future research are then discussed in Section 9.3. Finally, Section 9.4 makes some broad recommendations for future pedestrian modeling.

9.1. Key Findings

Fine spatial resolution was employed in this study, with 80-meter grid sizes in the Portland context and 100-meter grid sizes in the Munich context. It is proved that using fine spatial resolution can improve the model accuracy in mode shares and walk trip lengths while the model run time increases exponentially. This work managed to upgrade run time to a few minutes for running large-scale areas with fine spatial resolution. However, this study does not advocate for always using the finest spatial resolution for travel demand models. The selection of the appropriate spatial resolution highly depends on the type of application, the model aims to address and the level of details of the available data differing from locations.

The new pedestrian accessibility measurement proposed in this study follows a simple specification. It is easy to construct and transfer to other study areas. It showed positive and significant influences on walk mode choice. However, the goodness-of-fit of walk mode choice models in MoPeD 2.0 was poor. It is unsurprising because the travel distance is not known at the mode choice stage. This work would underline that the challenge for the walk mode choice model is significantly increased when mode choices are made prior to destination selection. Future research can focus on different modeling sequences or more advanced modeling approaches for joint mode and destination models.

The walk destination choice models in MoPeD 2.0 captured the relationship between built environment variables and destination utility. Network density and the accessibility to shops and retail stores showed strong positive effects on destination selection, whereas the proportion of industrial jobs, cross motorway, and slope were barriers. Different from previous studies that choose walk destinations at the TAZ level, the destination choice models in this study were employed at small scales. However, the model estimation became challenging with the use of small-scale destination zone alternatives. The models for selecting superPAZ destinations (400-meter grid cells) had good performances and showed intuitive associations with the built environment, while PAZ-level (80-meter grid cells) destination choice models had a low goodness-of-fit. This could be due to fewer variations across small-scale destination zones or the lack of important factors in the model. Future studies need to investigate more factors, such as micro-level or street-level built environment variables (e.g., pavement condition and the number of trees).

The MITO/MoPeD integrated model improved the estimation of travel outcomes such as mode shares and walk trip length distributions, but it did not contribute to the better assessment of physical activity volume. One major reason is that the access and egress walk trips are neglected in the integrated model. Active trips navigating to public transport are also a major source of physical activity. Those trips need to be considered in the model to simulate physical activity volumes more precisely. However, the travel information for access and egress trips was limited in HTS. More data needs to be collected to understand these trips. Another reason is that the integrated model does not capture pedestrian travel behavior variability. The physical activity volume is simply measured by multiplying walk time/distances by 7 to represent the weekly physical activity. To better assess physical activity volumes, it is crucial to either find out better methods of imputing weekly walk trips from single-day behavior or move travel demand models from one day to week-long.

The MITO/MoPeD integrated model is an open-source planning tool that can be transferred to other contexts and applied to various policies and scenarios. The main finding of the model integration task was that the technical integration of two models is straightforward though involving a large amount of coding work. However, the spatial transfer of econometric models from the estimation context to the application context is not trivial. Data needs to be collected at the same level of detail and applied using coherent definitions. For example, income categories need to be matched in two contexts because they are usually defined differently in different countries. More importantly, models need to be calibrated at an aggregate level by updating constants. More advanced approaches can also be applied to calibrate the differences in the magnitude of variation in observed and unobserved factors.

The collected GLH data is a novel and rich source and helped to better understand the travel behavior variability over days, weeks, and months. Although we suggest that a one-week period of observation could be sufficient to capture the variability of some travel metrics, there is no solid conclusion on how many days of observations are enough for capturing an individual's routine travel behavior. The analysis of GLH data confirmed that HTS was poor at capturing walking activities and biased in self-reported travel distance. Although limitations exist in the traditional HTS, this work does not underline the need to replace HTS with passively collected data such as GPS and smartphone data. More importantly, the work provided guidance on the design of HTS data collection and suggested that HTS data be validated against objective methods. The analysis of GLH data attempted to find out the potential determinants of travel behavior stability/variability. It is proved that socio-demographics have impacts on travel behavior variability, particularly students have higher intrapersonal variability than workers, and individuals who have no car are more variable than those who have cars. Weather and the days of public holidays can also disrupt an individual's routine. However, due to the limited number of life events observed among the survey participants, the relationship between life events and travel behavior was not clearly established in this study.

9.2. Contributions

This work makes major contributions to the improvement of pedestrian travel demand modeling in the broad areas of data collection, travel behavior studies, and modeling approaches.

First, the longitudinal travel behavior survey itself involved novel elements. One notable aspect was the collection of Google Location History data. Few transportation studies have utilized GLH, in part, because of their proprietary nature. Indeed, the effort to recruit individuals and have them download and share their individual data was not insignificant. However, these data are detailed sources of longitudinal activity and travel diaries that provide rich information for exploring travel behavior variability. Another novelty was the wide time spans of GLH data. Previous studies conducted self-reported surveys or collected GPS data for multiple days or weeks (K. Axhausen et al., 2002; Xianyu et al., 2017). The passively collected location data (GLH) collected in this dissertation recorded respondents' travel diaries over months, which provides great opportunities for exploring longer-term variations in travel behavior.

Additionally, this dissertation provides additional evidence toward growing literature on the field of travel behavior variability. Although the findings were not as novel as the contributions, the new data, along with the advanced approaches, provided numerical evidence in a quantitative way. For instance, the analyses in Chapter 8.6 confirmed the large intrapersonal travel behavior variability, and its degree varies widely across different temporal scales and travel metrics. The study also confirmed the limitations of HTS for capturing pedestrian activities and suggested that HTS assessment be validated against objective methods. Several relatively well-established findings about the determinants of travel behavior were also confirmed, such as association with weather and the built environment.

More importantly, this dissertation contributes to better pedestrian modeling practices. While MoPeD was a well-established pedestrian planning tool, the enhancement made in this dissertation upgrades the performance of MoPeD, particularly for model transferability and efficiency. Furthermore, the integrated model presented in Chapter 6 is one of the first to incorporate pedestrian modeling into the agent-based transport model. It provides more accurate model outcomes, resulting in a clear picture of future transport demand.

Overall, this dissertation makes strides towards a more accurate, sensitive, transferable, and efficient pedestrian planning tool for delivering travel outcomes as well as evaluating policies and scenarios. The work demonstrates that this planning tool is capable of evaluating a variety of policies and scenarios, including transportation infrastructure investments and land use planning. Perhaps the greatest contribution of this work is in the illustration of the use of longitudinal data. The findings from this analysis reveal some insights into travel behavior and perhaps more importantly provide some better guidance on the design of future data collection efforts and the utility of GLH for transportation analysis.

9.3. Limitations and Future Work

While the models developed in this dissertation made significant advances in pedestrian modeling, they are not without limitations. First, A number of additional factors can be added to the models to yield a better representation of pedestrian travel behavior, particularly around walk mode choice and destination choice. For instance, this application work in Chapter 5 reveals that the pedestrian planning tool appears to be only sensitive to the level of street connectivity rather than the quality of street connectivity, which refers to micro-level data such as the design and pavement conditions of the pedestrian street. Due to the large-scale application of the pedestrian planning tool, such small-scale information on the street-level built environment was not able to obtain for the entire study area. Also, having such fine-grained information in the model would increase the data collection burden, leading to lower model transferability to other study areas. Thus, the micro-level built environment factors were eliminated in this study. As more information becomes available, these street-level factors may be incorporated into pedestrian models. Also, this limitation may be addressed in the future when pedestrian models are applied to small-scale planning, such as neighborhood design.

Second, more sophisticated modeling approaches could be employed in pedestrian planning tools. The mode choice models only considered the traditional modes, including car, public transport, bike, and walking, but did not incorporate new modes such as e-bike, bike share, and e-scooter, which may compete with walking. This dissertation employed Oregon household travel surveys collected in the year 2011, which did not include new active mobility and micro-mobility. A number of electric bike trips are captured in the most recent travel surveys (e.g., MiD 2017), but the information on bike share and e-scooter is still missing. Furthermore, only trips with walking as the primary mode was simulated in the model. Public transit access and egress trips were excluded due to the lack of data. This may cause an underestimation of total walking activity, leading to an inaccurate assessment of health benefits.

Third, the small sample size of GLH data prevented us from drawing solid conclusions about travel behavior variability across the population. This is also the main reason why the GLH data was not used for developing pedestrian planning tools. Although the GLH data was not suitable for interpersonal travel behavior study, it is a unique data source for intrapersonal travel behavior research. A few analyses were conducted in this dissertation, but the collected data was not fully utilized. For instance, the fine location information in GLH data can be used for pedestrian route choice modeling. The shortest-path route choice model was implemented in the integrated model for simplicity. Future work can explore the relationship between pedestrian route choice and the built environment by using GLH data. Additionally, the analyses of travel behavior variability were conducted without differentiating activity purposes. Activity purposes may reflect different relationships with travel behavior variability. For example, individuals may have more stable travel behavior for commuting and some favorite shops and restaurants frequently visited, while relatively high variance for leisure activities (Hasan et al., 2013; Tarigan & Kitamura, 2009). Because such information is not inferred in the collected GLH data, future work will require algorithms for deducing activity purposes.

Finally, more research is needed to address the issue of GLH data reliability. Although a small-scale data validation test was conducted in this study and concluded that GLH data was reliable for capturing means of transport, there has recently been a debate among researchers. On the one hand, researchers showed evidence and success in using GLH data. Yu et al. (2019) concluded that GLH data reasonably captured the spatial movements of the subjects. They also confirmed that the estimation of daily exposures to ambient particulate matter using GLH had less than 1.2% error when compared to the results using the GPS data logger. On the other hand, researchers argued that GLH data does not capture the locations and trips adequately. Macarulla Rodriguez et al. (2018) pointed out that GLH data has a low accuracy rate in capturing locations in meters. Cools et al. (2021) validated the GLH data with prescribed itineraries and found that GLH was poor in capturing short trips. Disagreements arise mainly from different methodologies, the purpose of the study, and different ground-truth data. Therefore, a more comprehensive validation study needs to be conducted to check the reliability of GLH data for various applications.

In conclusion, more fundamental research is needed to investigate the travel behavior variations within the population and its association with the built environment and life events. Furthermore, modeling practices need to keep pace with the fundamental research by incorporating key influences and using more advanced modeling approaches.

9.4. Recommendations

It has been illustrated that the pedestrian planning tool presented here can provide better prediction of travel outcomes as well as assessment of transportation and land use policies and scenarios. Its capabilities, however, are not limited to this. Future efforts may be interested in linking these model outcomes to other tools for assessing broader issues such as air quality, public health, and road safety. For example, the number of walk trips along or across a corridor can be used as a measure of exposure when calculating crash rates and assessing traffic safety. Physical activity volumes could be used to assess the relative risk of all-cause mortality. Routes of walk trips on a pedestrian network can improve the precision of these metrics but also can be used for measuring air pollutant exposure while traveling. Such a model has a great opportunity of incorporating health, environmental, and equity research.

The next generation of pedestrian planning tools may attempt to integrate with activity-based models (ABM) and shift modeling from one day to a week. Developing pedestrian activities in ABM may have a wide range of benefits. First, ABM generates tours rather than separate trips, which may result in a better representation of the interconnected mode chain. Individuals who use public transit as their primary mode may be more likely to choose to walk as the trip mode. Then, ABM can capture the effect of the built environment on the total number of tours, which has been investigated in the literature (Zhang et al., 2019). Furthermore, ABM may incorporate a broader scope of walking. Walking is not only a trip mode but also a (physical) activity. ABM has a great opportunity of incorporating non-transport-domain physical activity such as leisure domain (e.g., walking a dog, jogging, and exercise). Finally, Kölbl & Helbing (2003) pointed out a law of constant average energy consumption for the physical activity of daily travel. Once a broader scope of walking activities is implemented in ABM, the nature law of physical activity budget may be introduced to improve the representation of pedestrian travel behavior in travel demand models.

More generally, the work in this dissertation pushes the pedestrian travel demand model toward finer spatial resolution and more measures of the pedestrian environment. Future developments in pedestrian travel demand models need to move toward a longer planning horizon and a broader range of pedestrian activities.

Bibliography

Agyemang-Duah, K., & Hall, F. L. (1997). Spatial transferability of an ordered response model of trip generation. *Transportation Research Part A: Policy and Practice*, *31*(5), 389–402. https://doi.org/10.1016/S0965-8564(96)00035-3

Aschauer, F., Hössinger, R., Schmid, B., & Gerike, R. (2018). Reporting quality of travel and non-travel activities: A comparison of three different survey formats. *Transportation Research Procedia*, *32*, 309–318.

Aultman-Hall, L., Lane, D., & Lambert, R. R. (2009). Assessing Impact of Weather and Season on Pedestrian Traffic Volumes. *Transportation Research Record: Journal of the Transportation Research Board*, *2140*(1), 35–43. https://doi.org/10.3141/2140-04

Axhausen, K. W., Löchl, M., Schlich, R., Buhl, T., & Widmer, P. (2007). Fatigue in long-duration travel diaries. *Transportation*, *34*(2), 143–160. https://doi.org/10.1007/s11116-006-9106-4

Axhausen, K., Zimmermann, A., Schönfelder, S., Rindsfüser, G., & Haupt, T. (2002). Observing the rhythms of daily life: A six-week travel diary. *Transportation*, *29*, 95–124. https://doi.org/10.1023/A:1014247822322

Berjisian, E., & Habibian, M. (2019). Developing a pedestrian destination choice model using the stratified importance sampling method. *Journal of Transport Geography*, *77*, 39–47. https://doi.org/10.1016/j.jtrangeo.2019.04.009

BMVI. (2019a). German Mobility Panel. Retrieved from www.mobilitaetspanel.de

BMVI. (2019b). *Mobility in Germany – short report*. Retrieved from www.mobilitaet-in-deutschland.de

Boarnet, M., & Crane, R. (2001). The influence of land use on travel behavior: specification and estimation strategies. *Transportation Research Part A: Policy and Practice*, *35*(9), 823–845. https://doi.org/10.1016/S0965-8564(00)00019-7

Borrmann, A., Kneidl, A., Köster, G., Ruzika, S., & Thiemann, M. (2012). Bidirectional coupling of macroscopic and microscopic pedestrian evacuation models. *Safety Science*, *50*(8), 1695–1703. https://doi.org/https://doi.org/10.1016/j.ssci.2011.12.021

Borst, H. C., de Vries, S. I., Graham, J. M. A., van Dongen, J. E. F., Bakker, I., & Miedema, H. M. E. (2009). Influence of environmental street characteristics on walking route choice of elderly people. *Journal of Environmental Psychology*, *29*(4), 477–484. https://doi.org/10.1016/j.jenvp.2009.08.002

Broach, J., & Dill, J. (2015). *Pedestrian Route Choice Model Estimated from Revealed Preference GPS Data*.

Buliung, R., Roorda, M., & Remmel, T. (2008). Exploring spatial variety in patterns of activity-travel behaviour: initial results from the Toronto Travel-Activity Panel Survey (TTAPS). *Transportation*, *35*, 697–722. https://doi.org/10.1007/S11116-008-9178-4

Centre for Diet and Activity Research. (2018). Integrated Transport and Health Impact Modelling Tool (ITHIM). Retrieved from http://www.cedar.iph.cam.ac.uk/research/modelling/ithim/

Cervero, R. (2003). The Built Environment and Travel: Evidence from the United States. *European Journal of Transport and Infrastructure Research*, *3*(2), 119–137. https://doi.org/Cited By (since 1996) 5\rExport Date 27 September 2011

Chaix, B., Benmarhnia, T., Kestens, Y., Brondeel, R., Perchoux, C., Gerber, P., & Duncan, D. T. (2019). Combining sensor tracking with a GPS-based mobility survey to better measure physical activity in trips: Public transport generates walking. *International Journal of Behavioral Nutrition and Physical Activity*, *16*(1), 1–13. https://doi.org/10.1186/s12966-019-0841-2

Chikaraishi, M., Fujiwara, A., Zhang, J., & Axhausen, K. (2009). Exploring Variation Properties of Departure Time Choice Behavior by Using Multilevel Analysis Approach. *Transportation Research Record*, *2134*, 10–20. https://doi.org/10.3141/2134-02

City of Munich. (2021). Munich Low Emission Zone. Retrieved from https://stadt.muenchen.de/infos/munich-low-emission-zone.html

City of Portland. (2018). *CENTRAL CITY 2035 GOALS AND POLOCIES (Vol. 1)* (Vol. 1). Portland. Retrieved from https://www.portland.gov/bps/cc2035/adopted-central-city-2035-plan-documents

Clifton, K. J., Orrego-Oñate, J., Singleton, P., & Schneider, R. (2019). *Transferability & Forecasting of the Pedestrian Index of the Environment (PIE) for Modeling Applications*. Portland, OR. https://doi.org/10.15760/trec.218

Clifton, K. J., Singleton, P. A., Muhs, C. D., & Schneider, R. J. (2016a). Development of destination choice models for pedestrian travel. *Transportation Research Part A: Policy and Practice*, *94*, 255–265. https://doi.org/10.1016/j.tra.2016.09.017

Clifton, K. J., Singleton, P. A., Muhs, C. D., & Schneider, R. J. (2016b). Representing pedestrian activity in travel demand models: Framework and application. *Journal of Transport Geography*, *52*, 111–122. https://doi.org/10.1016/j.jtrangeo.2016.03.009

Clifton, K. J., Singleton, P. A., Muhs, C. D., Schneider, R. J., Lagerwey, & Peter. (2013). *Improving The Representation Of The Pedestrian Environment In Travel Demand Models – Phase I*. Portland.

Clifton, K. J., Singleton, P., Muhs, C., & Schneider, R. (2015). *Development of a Pedestrian Demand Estimation Tool*. https://doi.org/10.15760/trec.124

Clifton, K., & Muhs, C. (2012). Capturing and representing multimodal trips in travel surveys. *Transportation Research Record*, (2285), 74–83. https://doi.org/10.3141/2285-09

Cools, D., McCallum, S. C., Rainham, D., Taylor, N., & Patterson, Z. (2021). Understanding google location history as a tool for travel diary data acquisition. *Transportation Research Record*, *2675*(5), 238–251. https://doi.org/10.1177/0361198120986169

Costa, S., Ogilvie, D., Dalton, A., Westgate, K., Brage, S., & Panter, J. (2015). Quantifying the physical activity energy expenditure of commuters using a combination of global positioning system and combined heart rate and movement sensors. *Preventive Medicine*, *81*, 339–344. https://doi.org/10.1016/j.ypmed.2015.09.022

Cotrus, A. V, Prashker, J., & Shiftan, Y. (2005). Spatial and temporal transferability of trip generation demand models in Israel. *Journal of Transportation and Statistics*, *8*, 37–56.

Cui, Y., He, Q., & Khani, A. (2018). Travel Behavior Classification: An Approach with Social Network and Deep Learning. *Transportation Research Record*, (June). https://doi.org/10.1177/0361198118772723

Delaware Valley Regional Planning Commission. (2011). *DVRPC Travel Demand Model Upgrade - Travel Improvement Model (TIM) 1.0*.

Department for Transport. (2019). *Statistical Release Analyses from the National Travel Survey*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attac hment_data/file/775032/2019-nts-commissioned-analyses.pdf

Dharmowijoyo, D. B., Susilo, Y., & Karlström, A. (2014). Day-to-Day Interpersonal and Intrapersonal Variability of Individuals' Activity Spaces in a Developing Country. *Environment and Planning B: Planning and Design*, *41*, 1063–1076. https://doi.org/10.1068/b130067p

Dharmowijoyo, D. B., Susilo, Y., & Karlström, A. (2017). Analysing the complexity of day-to-day individual activity-travel patterns using a multidimensional sequence alignment model: A case study in the Bandung Metropolitan Area, Indonesia. *Journal of Transport Geography*, *64*, 1–12. https://doi.org/10.1016/J.JTRANGEO.2017.08.001

El Mahrsi, M. K., Come, E., Oukhellou, L., & Verleysen, M. (2017). Clustering Smart Card Data for Urban Mobility Analysis. *IEEE Transactions on Intelligent Transportation Systems*, *18*(3), 712–728. https://doi.org/10.1109/TITS.2016.2600515

Erdmann, J., & Krajzewicz, D. (2015). Modelling Pedestrian Dynamics in SUMO. *SUMO User Conference 2015*, 103–118. Berlin. Retrieved from http://elib.dlr.de/100554/

Evenson, K. R., Satinsky, S. B., Rodríguez, D. A., & Aytur, S. A. (2012, March 15). Exploring a Public Health Perspective on Pedestrian Planning. https://doi.org/10.1177/1524839910381699

Everett, J. D. (2009). *An Investigation of the Transferability of Trip Generation Models and the Utilization of aSpatial Context Variable* (University of Tennessee). University of Tennessee. Retrieved from https://trace.tennessee.edu/utk_graddiss/47

Ewing, R., & Cervero, R. (2001). Travel and the Built Environment: A Synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, *1780*(01), 87–114. https://doi.org/10.3141/1780-10

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, *76*(3), 265–294. https://doi.org/10.1080/01944361003766766

Fairnie, G. A., Wilby, D. J. R., & Saunders, L. E. (2016). Active travel in London: The role of travel survey data in describing population physical activity. *Journal of Transport and Health*, *3*(2), 161–172. https://doi.org/10.1016/j.jth.2016.02.003

Gehrke, S., & Clifton, K. (2014). Operationalizing Land Use Diversity at Varying Geographic Scales and Its Connection to Mode Choice. *Transportation Research Record: Journal of the Transportation Research Board*, *2453*(December 2014), 128–136. https://doi.org/10.3141/2453-16

Goulet-Langlois, G., Koutsopoulos, H., Zhao, Z., & Zhao, J. (2018). Measuring Regularity of Individual Travel Patterns. *IEEE Transactions on Intelligent Transportation Systems*, *19*, 1583–1592. https://doi.org/10.1109/TITS.2017.2728704

Greenwald, M. J., & Boarnet, M. G. (2001). Built Environment as Determinant of Walking Behavior: Analyzing Nonwork Pedestrian Travel in Portland, Oregon. *Transportation Research Record: Journal of the Transportation Research Board*, *1780*(1), 33–41. https://doi.org/10.3141/1780-05

Haggerty, B., Hyde, J., & Payne, M. (2011). *Rapid Health Impact Assessment : Vancouver Comprehensive Growth Management Plan 2011*.

Hahm, Y., Yoon, H., & Choi, Y. (2019). The effect of built environments on the walking and shopping behaviors of pedestrians; A study with GPS experiment in Sinchon retail district in Seoul, South Korea. *Cities*, *89*, 1–13. https://doi.org/10.1016/j.cities.2019.01.020

Hanson, S., & Huff, J. O. (1982). ASSESSING DAY-TO-DAY VARIABILITY IN COMPLEX TRAVEL PATTERNS. *Transportation Research Record*. Retrieved from https://www.semanticscholar.org/paper/6de94bad89a09df87351fec6e54d6706197cb9e9

Hanson, S., & Huff, J. O. (1986). Classification issues in the analysis of complex travel behavior. *Transportation*, *13*, 271–293. https://doi.org/10.1007/BF00148620

Hanson, S., & Huff, O. (1988). Systematic variability in repetitious travel. *Transportation*, *15*, 111–135. https://doi.org/10.1007/BF00167983

Hasan, S., Schneider, C. M., Ukkusuri, S. V., & González, M. C. (2013). Spatiotemporal Patterns of Urban Human Mobility. *Journal of Statistical Physics*, *151*(1–2), 304–318. https://doi.org/10.1007/s10955-012-0645-0

Health Effects Institute (HEI). (2010). *Traffic-related air pollution: A critical review of the literature on emissions, exposure, and health effects* (H. E. Institute, Ed.). Boston, MA: Health Effects Institute (HEI). Retrieved from Health Effects Institute (HEI) website: http://pubs.healtheffects.org/view.php?id=334

Heinen, E., & Chatterjee, K. (2015). The same mode again? An exploration of mode choice variability in Great Britain using the National Travel Survey. *Transportation Research Part A: Policy and Practice*, *78*, 266–282. https://doi.org/10.1016/j.tra.2015.05.015

Horni, A., Nagel, K., & Axhausen, K. W. (2016). *The Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press. https://doi.org/http://dx.doi.org/10.5334/baw

Huntsinger, L. F., & Rouphail, N. M. (2013). Temporal stability of generation choice models. *Transportation Research Record*, (2344), 79–87. https://doi.org/10.3141/2344-09

Jones, P., & Clarke, M. (1988). The significance and measurement of variability in travel behaviour. *Transportation*, *15*, 65–87. https://doi.org/10.1007/BF00167981

Kahlmeier, S., Kelly, P., Foster, C., Götschi, T., Cavill, N., Dinsdale, H., … Racioppi, F. (2014). *Health Economic Assessment Tools (HEAT) for walking and for cycling economic assessment of transport infrastructure*. Retrieved from http://www.euro.who.int/en/health-topics/environment-and-health/Transport-and-health/publications/2011/health-economic-assessment-tools-heat-for-walking-and-for-cycling.-methodology-and-user-guide.-economic-assessment-of-transport-infrastructure-and-polic

Karasmaa, N. (2003). *The Transferability of Travel Demand Models: An analysis of transfer methods, data quality and model estimation*. Helsinki University of Technology.

Khan, M., Kockelman, K. M., & Xiong, X. (2014). Models for anticipating non-motorized travel choices, and the role of the built environment. *Transport Policy*, *35*, 117–126.

Kielar, P. M., & Borrmann, A. (2016). Modeling pedestrians' interest in locations: A concept to improve simulations of pedestrian destination choice. *Simulation Modelling Practice and Theory*, *61*, 47–62. https://doi.org/10.1016/j.simpat.2015.11.003

Koh, P. P., & Wong, Y. D. (2013). Influence of infrastructural compatibility factors on walking and cycling route choices. *Journal of Environmental Psychology*, *36*, 202–213. https://doi.org/10.1016/j.jenvp.2013.08.001

Kölbl, R., & Helbing, D. (2003). Energy laws in human travel behaviour. *New Journal of Physics*, *5*(1), 48. https://doi.org/10.1088/1367-2630/5/1/348

Kuehnel, N., Zhang, Q., Staves, C., & Moeckel, R. (2022). The Perfect Match?

Assessment of Excess Commute and Transport Externalities using an Agent-Based Transport Model. *Under Review*.

Kunert, U., Kloas, J., & Kuhfeld, H. (2002). Design Characteristics of National Travel Surveys: International Comparison for 10 Countries. *Transportation Research Record: Journal of the Transportation Research Board*, *1804*(1), 107–116. https://doi.org/10.3141/1804-15

Landeshauptstadt Muenchen. (2022). Bevoelkerung. Retrieved from https://stadt.muenchen.de/infos/statistik-bevoelkerung.html#:~:text=Auch im Jahr 2020 ist,innen mit Hauptwohnsitz in München.

Lange, C., Finger, J. D., Allen, J., Born, S., Hoebel, J., Kuhnert, R., … Ziese, T. (2017). Implementation of the European health interview survey (EHIS) into the German health update (GEDA). *Archives of Public Health*, *75*(1), 40. https://doi.org/10.1186/s13690-017-0208-6

Lassarre, S., Papadimitriou, E., Yannis, G., & Golias, J. (2007). Measuring accident risk exposure for pedestrians in different micro-environments. *Accident Analysis and Prevention*, *39*(6), 1226–1238. https://doi.org/10.1016/j.aap.2007.03.009

Lee, C., & Moudon, A. V. (2006). Correlates of Walking for Transportation or Recreation Purposes. *Journal of Physical Activity and Health*, *3*(s1), S77–S98. https://doi.org/10.1123/jpah.3.s1.s77

Li, W., Houston, D., Boarnet, M., & Park, H. (2018). Intrapersonal day-to-day travel variability and duration of household travel surveys: Moving beyond the one-day convention. *Journal of Transport and Land Use*. https://doi.org/10.5198/JTLU.2018.984

Liu, F., Evans, J. E. (Jay), & Rossi, T. (2012). Recent Practices in Regional Modeling of Nonmotorized Travel. *Transportation Research Record: Journal of the Transportation Research Board*, *2303*(1), 1–8. https://doi.org/10.3141/2303-01

Lue, G., & Miller, E. J. (2019). Estimating a Toronto pedestrian route choice model using smartphone GPS data. *Travel Behaviour and Society*, *14*, 34–42. https://doi.org/10.1016/j.tbs.2018.09.008

Macarulla Rodriguez, A., Tiberius, C., van Bree, R., & Geradts, Z. (2018). Google timeline accuracy assessment and error prediction. *Forensic Sciences Research*, *3*(3), 240–255. https://doi.org/10.1080/20961790.2018.1509187

Maizlish, N., Woodcock, J., Co, S., Ostro, B., Fanai, A., & Fairley, D. (2013). Health cobenefits and transportation-related reductions in greenhouse gas emissions in the San Francisco Bay Area. *American Journal of Public Health*, *103*(4), 703–709. https://doi.org/10.2105/AJPH.2012.300939

Manley, E., Zhong, C., & Batty, M. (2016). Spatiotemporal variation in travel regularity through transit user profiling. *Transportation*, *45*, 703–732. https://doi.org/10.1007/s11116-016-9747-x

Merom, D., van der Ploeg, H. P., Corpuz, G., & Bauman, A. E. (2010). Public Health Perspectives on Household Travel Surveys. *American Journal of Preventive Medicine*, *39*(2), 113–121. https://doi.org/10.1016/j.amepre.2010.04.007

Miranda-Moreno, L. F., & Lahti, A. C. (2013). Temporal trends and the effect of weather on pedestrian volumes: A case study of Montreal, Canada. *Transportation Research Part D: Transport and Environment*, *22*, 54–59. https://doi.org/10.1016/j.trd.2013.02.008

Moeckel, R., & Donnelly, R. (2015). Gradual rasterization: redefining spatial resolution in transport modelling. *Environment and Planning B: Planning and Design*, *42*(5), 888–903. https://doi.org/10.1068/b130199p

Moeckel, R., Kuehnel, N., Llorca, C., Moreno, A. T., & Rayaprolu, H. (2020). Agent-

Based Simulation to Improve Policy Sensitivity of Trip-Based Models. *Journal of Advanced Transportation*, *2020*, 1902162. https://doi.org/10.1155/2020/1902162

Molloy, J., & Moeckel, R. (2017). Automated design of gradual zone systems. *Open Geospatial Data, Software and Standards*, *2*(1), 19. https://doi.org/10.1186/s40965-017-0032-5

Moreno, A., & Moeckel, R. (2018). Population Synthesis Handling Three Geographical Resolutions. *ISPRS International Journal of Geo-Information*, *7*(5), 174. https://doi.org/10.3390/ijgi7050174

Nerella, S., & Bhat, C. R. (2004). Numerical Analysis of Effect of Sampling of Alternatives in Discrete Choice Models. *Transportation Research Record: Journal of the Transportation Research Board*, *1894*(1), 11–19. https://doi.org/10.3141/1894-02

Okrah, M. B. (2016). Evaluation of Methods for Estimation of Intrazonal Travel Impedances. *Transportation Research Record: Journal of the Transportation Research Board*, *2563*(1), 114–121. https://doi.org/10.3141/2563-16

Okrah, M. B., Moeckel, R., & Wulfhorst, G. (2017). Finding the optimal level of spatial resolution for handling non- motorized travel in macroscopic travel demand models. *World Conference on Transport Research*. Shanghai.

OpenStreetMap Contributors. (2017). Planet dump retrieved from https://planet.osm.org. Retrieved from https://www.openstreetmap.org

OpenStreetMap Wiki. (2020). *Key:highway --- OpenStreetMap Wiki{,}*. Retrieved from https://wiki.openstreetmap.org/w/index.php?title=Key:highway&oldid=1985676

Pas, E. I. (1983). A Flexible and Integrated Methodology for Analytical Classification of Daily Travel-Activity Behavior. *Transportation Science*, *17*, 405–429. https://doi.org/10.1287/TRSC.17.4.405

Pas, E. I. (1987). INTRAPERSONAL VARIABILITY AND MODEL GOODNESS-OF-FIT. *Transportation Research Part A: General*, *21*, 431–438. https://doi.org/10.1016/0191-2607(87)90032-X

Pas, E. I., & Koppelman, F. (1986). An examination of the determinants of day-to-day variability in individuals' urban travel behavior. *Transportation*, *13*, 183–200. https://doi.org/10.1007/BF00165547

Pas, E. I., & Sundar, S. (1995). Intrapersonal variability in daily urban travel behavior: Some additional evidence. *Transportation*, *22*, 135–150. https://doi.org/10.1007/BF01099436

Pedersen, E. S. L., Danquah, I. H., Petersen, C. B., & Tolstrup, J. S. (2016). Intra-individual variability in day-to-day and month-to-month measurements of physical activity and sedentary behaviour at work and in leisure-time among Danish adults. *BMC Public Health*, *16*(1), 1222. https://doi.org/10.1186/s12889-016-3890-3

Portland Bureau of Transportation. (2019). Congressman Earl Blumenauer Bicycle and Pedestrian Bridge. Retrieved from https://www.portlandoregon.gov/transportation/71338

Portland METRO. (2020). *2018 Trip-Based Travel Demand Model Methodology Report*. Portland, Oregon.

Prelipcean, A. C., Gidófalvi, G., & Susilo, Y. (2018). *Longest common subsequences: Identifying the stability of individuals' travel patterns*. Retrieved from https://www.semanticscholar.org/paper/1ea9778a6191aafc43d65eda0c0a32acd6558390

Raux, C., Ma, T. Y., & Cornelis, E. (2016). Variability in daily activity-travel patterns: the case of a one-week travel diary. *European Transport Research Review*, *8*(4). https://doi.org/10.1007/s12544-016-0213-9

Reiffer, A., Barthelmes, L., Kagerbauer, M., & Vortisch, P. (2022). Representation of Work-Related Trip Patterns in Household and Commercial Travel Surveys. *Transportation Research Record*, 03611981221091559.

Rhodus, J., Fulk, F., Autrey, B., O'Shea, S., & Roth, A. (2013). A Review of Health Impact Assessments in the U.S.: Current State-of-Science, Best Practices, and Areas for Improvement. In *U.S. Environmental Protection Agency*. Washington, DC.

Rodríguez, D. A., Merlin, L., Prato, C. G., Conway, T. L., Cohen, D., Elder, J. P., … Veblen-Mortenson, S. (2015). Influence of the Built Environment on Pedestrian Route Choices of Adolescent Girls. *Environment and Behavior*, *47*(4), 359–394. https://doi.org/10.1177/0013916513520004

Saelens, B. E., & Handy, S. L. (2008). Built Environment Correlates of Walking. *Medicine & Science in Sports & Exercise*, *40*(7), S550–S566. https://doi.org/10.1249/MSS.0b013e31817c67a4

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine*, *25*(2), 80–91. https://doi.org/10.1207/S15324796ABM2502_03

Sallis, J. F., Bull, F., Burdett, R., Frank, L. D., Griffiths, P., Giles-Corti, B., & Stevenson, M. (2016). Use of science to guide city planning policy and practice: how to achieve healthy and sustainable future cities. *The Lancet*, *388*(10062), 2936–2947. https://doi.org/10.1016/S0140-6736(16)30068-X

San Francisco County Transportation Authority. (2002). *San Francisco Travel Demand Forecasting Model Development: Executive Summary* (Vol. 29).

Saunders, L. E., Green, J. M., Petticrew, M. P., Steinbach, R., & Roberts, H. (2013). What Are the Health Benefits of Active Travel? A Systematic Review of Trials and Cohort Studies. *PLoS ONE*, *8*(8), e69912. https://doi.org/https://doi.org/10.1371/journal.pone.0069912

Schlich, R., & Axhausen, K. (2003). Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation*, *30*(1), 13–36. https://doi.org/https://doi.org/10.1023/A:1021230507071

Schnittger, S., & Zumkeller, D. (2004). *Longitudinal Microsimulation as a tool to merge transport planning and traffic engineering models - the MobiTopp model*. Karlsruhe.

Sikder, S., Pinjari, A. R., Srinivasan, S., & Nowrouzian, R. (2013). Spatial transferability of travel forecasting models: a review and synthesis. *International Journal of Advances in Engineering Sciences and Applied Mathematics*, *5*(2–3), 104–128. https://doi.org/10.1007/s12572-013-0090-6

Singleton, P. A., Schneider, R. J., Muhs, C., & Clifton, K. J. (2014). The pedestrian index of the environment: Representing the walking environment in planning applications. *Transportation Research Board 93rd Annual Meeting*. Washington DC.

Singleton, P. A., Totten, J. C., Orrego-Oñate, J. P., Schneider, R. J., & Clifton, K. J. (2018). Making Strides: State of the Practice of Pedestrian Forecasting in Regional Travel Models. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(35), 58–68. https://doi.org/10.1177/0361198118773555

Singleton, P. A., & Wang, L. (2014). Safety and Security in Discretionary Travel Decision Making. *Transportation Research Record: Journal of the Transportation Research Board*, *2430*(1), 47–58. https://doi.org/10.3141/2430-06

Staves, C. (2020). Physical Activity Assessment and Modelling using Household Travel

Surveys. *Master Thesis, Technical University of Munich*. Retrieved from https://www.bgu.tum.de/fileadmin/w00blj/msm/theses/Staves_Thesis_2020.pdf

Streit, T., Allier, C., Weiss, C., Chlond, B., & Vortisch, P. (2015). Changes in Variability and Flexibility of Individual Travel in Germany. *Transportation Research Record*, *2496*, 10–19. https://doi.org/10.3141/2496-02

Susilo, Y., & Axhausen, K. (2014). Repetitions in individual daily activity–travel–location patterns: a study using the Herfindahl–Hirschman Index. *Transportation*, *41*, 995–1011. https://doi.org/10.1007/S11116-014-9519-4

Targa, F., & Clifton, K. (2004). Built Environment and Trip Generation for Non-Motorized Travel. *National Household Travel Survey Conference: Data for Understanding Our Nation's Trave*, (301), 13.

Tarigan, A., Fujii, S., & Kitamura, R. (2012). Intrapersonal variability in leisure activity-travel patterns: the case of one-worker and two-worker households. *Transportation Letters*, *4*, 1–13. https://doi.org/10.3328/TL.2012.04.01.1-13

Tarigan, A., & Kitamura, R. (2009). Week-to-Week Leisure Trip Frequency and Its Variability. *Transportation Research Record*, *2135*, 43–51. https://doi.org/10.3141/2135-06

The Department for Transport. (2020). National travel survey 2019 technical report. In *The Department for Transport*.

Tian, G., & Ewing, R. (2017). A walk trip generation model for Portland, OR. *Transportation Research Part D: Transport and Environment*, *52*, 340–353. https://doi.org/10.1016/j.trd.2017.03.017

Travel Forecasting Resource. (2019). Destination Choice Set Formation. Retrieved July 1, 2019, from http://tfresource.org/Destination_Choice_Set_Formation

Vogt, M., Researcher, S., Researcher, S., Bosa, P., Patton, T., Pederson, C., … Stein, B. (2015). *2015 Trip-Based Travel Demand Model Methodology Report Modeling Services Manager*.

Wolf, J., Simas Oliveira, M., & Thompson, M. (2003). Impact of Underreporting on VMT and Travel Time Estimates: Preliminary Findings from the California Statewide Household Travel Survey GPS Study. *Transportation Research Record*, *1854*, 189–198. https://doi.org/10.3141/1854-21

Woodcock, J., Franco, O. H., Orsini, N., & Roberts, I. (2011). Non-vigorous physical activity and all-cause mortality: systematic review and meta-analysis of cohort studies. *International Journal of Epidemiology*, *40*(1), 121–138. https://doi.org/10.1093/ije/dyq104

Woodcock, James, Edwards, P., Tonne, C., Armstrong, B. G., Ashiru, O., Banister, D., … Roberts, I. (2009). Public health benefits of strategies to reduce greenhouse-gas emissions: urban land transport. *The Lancet*, *374*(9705), 1930–1943. https://doi.org/10.1016/S0140-6736(09)61714-1

Woodcock, James, Givoni, M., & Morgan, A. S. (2013). Health Impact Modelling of Active Travel Visions for England and Wales Using an Integrated Transport and Health Impact Modelling Tool (ITHIM). *PLoS ONE*, *8*(1), e51462. https://doi.org/10.1371/journal.pone.0051462

Xianyu, J. J., Rasouli, S., & Timmermans, H. (2017). Analysis of variability in multi-day GPS imputed activity-travel diaries using multi-dimensional sequence alignment and panel effects regression models. *Transportation*, *44*, 533–553. https://doi.org/10.1007/S11116-015-9666-2

Yu, X., Stuart, A. L., Liu, Y., Ivey, C. E., Russell, A. G., Kan, H., … Yu, H. (2019). On the accuracy and potential of Google Maps location history data to characterize individual mobility for air pollution health studies. *Environmental Pollution*, *252*, 924–930. https://doi.org/10.1016/j.envpol.2019.05.081

Zahavi, Y. (1974). *Travel Time Budgets and Mobility in Urban Areas*.

Zhang, Q., Clifton, K. J., & Moeckel, R. (2019). Investigate an Appropriate Spatial Resolution for Large-scaled Pedestrian Travel Demand Model. *Transportation Research Procedia*, *41*, 324–327. https://doi.org/10.1016/j.trpro.2019.09.053

Zhang, Q., Clifton, K., Moeckel, R., & Orrego-Oñate, J. (2019). Household Trip Generation and the Built Environment: Does More Density Mean More Trips? *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(5), 596–606. https://doi.org/10.1177/0361198119841854

Zhang, Q., Moeckel, R., & Clifton, K. (2022). Assessing pedestrian impacts of future land use and transportation scenarios. *Journal of Transport and Land Use*, *15*(1), 547–566. https://doi.org/10.5198/jtlu.2022.2117

Zhao, Y., & Kockelman, K. M. (2002). The propagation of uncertainty through travel demand models: An exploratory analysis. *The Annals of Regional Science*, *36*(1), 145–163. https://doi.org/10.1007/s001680200072

Appendix 1: Portland Central City Scenario

Projected household and employment growth distribution plan at PAZ level for 2035

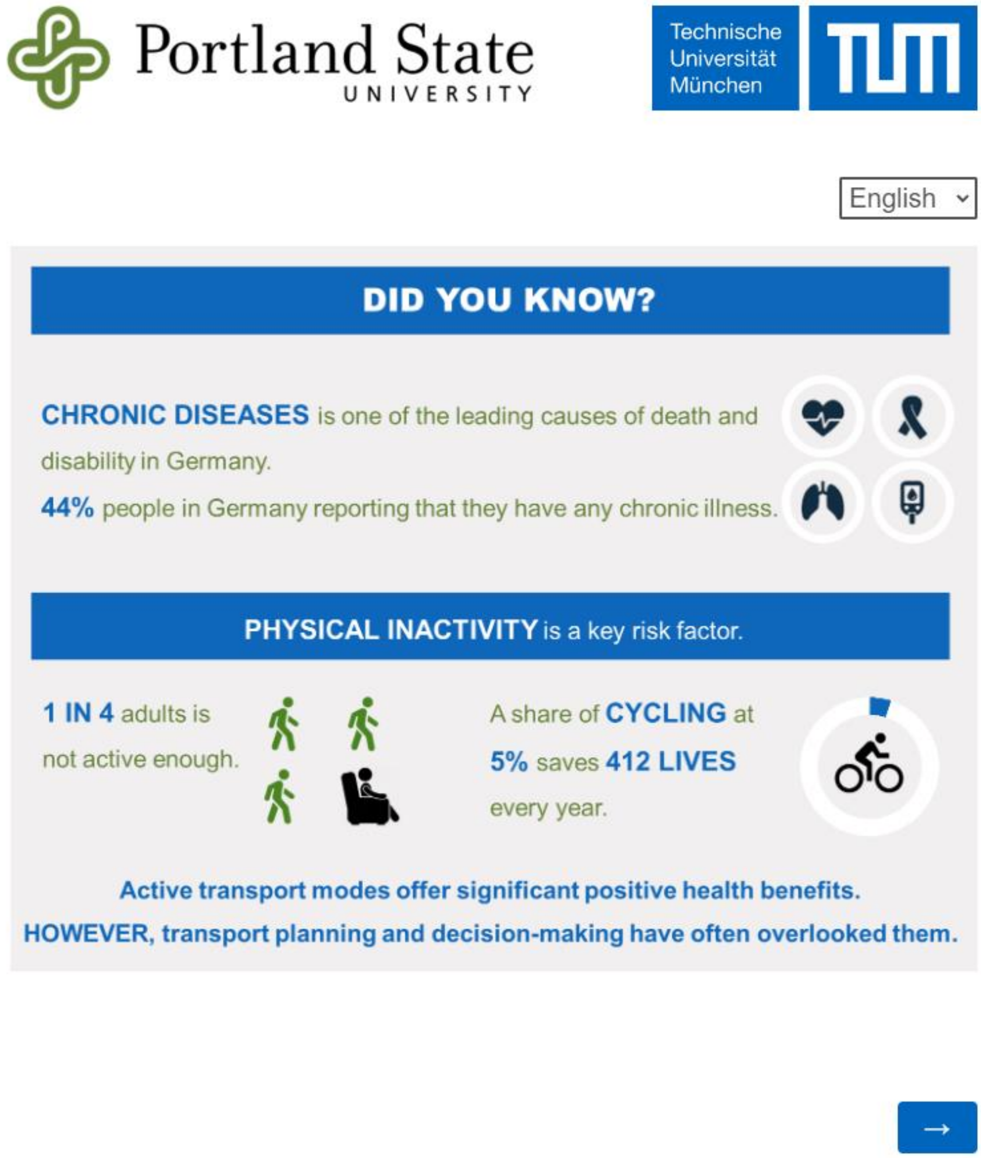
District	Total growth	Land use type	#PAZs allocated	Population and employment growth in each PAZ					
				Household	Service	Retail	Finance	Government	Industrial
Central Eastside	7,000 Households 8,000 Jobs	Housing	15	467	-	-	-	-	-
		Commercial (lower triangle)	41	-	73	24	24	24	-
		Commercial (remaining)	13	-	16	16	16	-	-
		Industrial	290	-	-	-	-	-	2
Downtown	3,000 Households 7,000 Jobs	Housing	10	700	-	-	-	-	-
		Commercial	6	-	333	167	500	-	-
		Government	2	-	-	-	-	500	-
Goose Hollow	1,000 Households 2,000 Jobs	Housing	13	77	-	-	-	-	-
		Commercial	4	-	200	160	-	-	-
		Institutional	1	-	-	-	-	200	-
Lloyd	8,000 Households 9,000 Jobs	Housing	33	242	7	14	-	-	-
		Commercial (office core)	14	-	40	16	161	161	-
		Commercial	9	-	163	175	-	-	-
Lower Albina	200 Households 200 Jobs	Housing	1	200	-	-	-	-	-
		Industrial	3	-	-	-	-	-	67
Old Town/ Chinatown	2,000 Households 3,000 Jobs	Housing	8	25	-	-	-	-	-
		Commercial	9	-	222	111	-	-	-
The Pearl	6,000 Households 4,000 Jobs	Housing	23	261	-	29	-	-	-
		Commercial	9	-	182	29	45	-	-
		Institutional	2	-	-	29	-	250	-
West End	3,000 Households 3,000 Jobs	Housing	8	375	63	63	-	-	-
		Commercial	7	-	143	143	-	-	-
University District	3,000 Households 4,000 Jobs	Housing	10	300	-	43	-	-	-
		Commercial	7	-	143	43	-	-	-
		Institutional	6	-	-	43	-	333	-
South Waterfront	4,000 Households 10,000 Jobs	Housing	22	182	-	18	-	-	-
		Commercial	3	-	333	18	-	-	-
		Institutional	30	-	-	18	-	267	-

Appendix 2: Summary of Oral Feedback Session with Survey Pilots

- Was Google Location History on by default?
  - Android
    - On by default: 2 people
  - Apple
    - Turned on manually: 1 person
    - Turned off by default or turned off manually, not sure: 1 person
- Regarding the welcome page (1st page)
  - Extra information sheet: Only 1 person clicked on the extra information sheet linked in the first page. The rest skipped it.
    - Perhaps this should be consolidated somehow?
  - "To the point"
  - The page does not clarify what the researcher is looking for. It could be acceptable to be entirely forthcoming about this since the data is from the past and shouldn't be able to be influenced by the survey. The survey questions ask about facts, not opinions.
  - "Jargon" – the page should use more layman's terms
  - Needs to establish a relationship with person
  - Show user that you appreciate them
  - "This is incredibly private information" so need a better hook. Give connection to health? Environment? Need more altruistic purposes to inspire the respondent in the letter but need to truthfully be able to link it to the purpose of the research.
  - "Think about explaining it to your grandmother"
- Regarding the turning on tracking and responding in 4 weeks
  - Remove this.
  - The survey should boot people that do not already have location history turned on.
- Regarding the length of the survey
  - Survey itself is not long or difficult to complete
  - If you had to go turn on location history and then come back to the survey (for the 4 weeks tracking purpose), this took a long time
- Regarding the question format/style
  - Drag and drop is "Hard to do on phone," especially when trying to drag/drop the last option in the list
  - Drag and drop is "awkward looking"
  - "Maybe make it a checkbox – does not need to be this fancy"
- Regarding questions asking "Did any of the following event(s) happen in your household since January 2018?
  - User error happened – more than one person was thinking about the current year and not 2018
  - Could use since January 1st
  - "Am I selecting month that I moved or the whole time I was gone?" – one respondent was temporarily located elsewhere while house was getting remodeled
  - Option 3 is confusing (saying two different things)
  - Say "in the past 18 months" instead of January

- Regarding question asking “What was your total household annual income in 2018?
    - Everyone just did their taxes, so it was easier than normal
    - Assumptions were made of partner’s income in some cases
    - “Can be awkward to ask roommates about income”
- Regarding question about getting/buying bicycle
    - Needs to confirm that these questions are looking at net change of bicycles in household
    - How do your account for if you are buying a bicycle for your child?
- Regarding question about employment status
    - The question is not being read through thoroughly. It should be up front (question stem – see reading).
    - Maybe “Did anything change about your work” and list potential options
    - Somebody stepped down from an academic position, but felt weird about saying “I was demoted”
    - Perhaps add “Other” or “lateral shift or change in job title”
    - People who are self-employed may feel better about saying “Other” as they may not necessarily work from home
- Regarding downloading data and privacy
    - Takes a long time to download the data
    - Need to verify that the data file itself is anonymized (no IP address, username, etc.)
    - Does Qualtrics collect IP? Could potentially link this file then to Qualtrics respondent
    - If collecting contact information on Qualtrics for providing incentives, how is this anonymized?
- Regarding uploading data
    - Someone accidentally selected a different file on their phone and was unable to swap files, therefore no data was uploaded
    - File size is good – on the order of 10 of megabytes
- Other comments
    - bike share should be listed among the bike options
    - “I was happy to be able to choose that I work and go to school”
    - “Add pickup or SUV to car” because some people are very proud of their pickup or SUV
    - How to handle children that do not spend all of their time in a household, such as
        - Divorced
        - Children in college
    - “Why am I as a PhD student grouped together with a high school student”
    - “Demographic questions should always be at the end” – J Dill
    - “Sometimes the survey asks about the respondent, other times it asks about the household. Be clear.
    - What is important? how many people have bicycles or how many bicycles in a household.
    - Why ask about children under age 6 specifically? What is the purpose of doing this?
    - Did not ask about gender, race, ethnicity, disability, age, or “other normal demographic data”

Appendix 3: Survey



English

Dear participant,

To understand more about the links between travel patterns and health, we kindly ask for your help with this transportation study. The research goal is to understand how various life events interact with transportation choices, including the use of active modes.

This research consists of two parts. All participants will be invited to take part in both stages. Your responses to both parts will be confidential and you will not be identified in any reports or analysis.

> 1) The first part of the survey will take about 10 to 15 minutes to complete. It asks questions about you and your household. In addition, it makes use of location data history that is automatically collected by your smartphone. In this stage, you will be asked to share this information with us. More information about these data and how we will ensure your confidentiality will be presented on the next page.
>
> 2) The second part is to fill out a short, online questionnaire that asks questions about the occurrence of major life events (marriage, children, changing jobs, etc.) since January 2018. It can be completed within 7-8 minutes.

Your responses to both parts will be confidential and you will not be identified in any reports or analysis.

If you complete the survey, you will have the opportunity to enter a draw to win one of ten Amazon gift cards. If you wish to enter the drawing, you will be asked to provide an email address, which will be deleted after prizes are awarded.

The data that you share will be useful in understanding how people make changes to their transportation patterns after life events. This research can only be successful with the generous help of people like you. We hope you will enjoy answering our questions and look forward to receiving your responses.

More information about this study, the data, and our confidentiality procedures will be provided on the next page.

← →

English

Please review the ***Informed Consent*** document, which provides detailed information about this survey and our research.

This research is being carried out by a doctoral candidate, Qin Zhang, a research associate at the Technical University of Munich (TUM) advised by Dr. Rolf Moeckel of TUM and Dr. Kelly Clifton of Portland State University (PSU). If you have questions or concerns, you may contact us by email at qin.zhang@tum.de or by calling +49-89-289-22698.

← →

English

By proceeding to the next page, you are consenting to participating in this study. Your participation in this research is voluntary and you may choose to withdraw from the study and request deletion of your personal data at any time.

← →

English

Thank you for participant our research project! Please answer the following questions to confirm your eligibility.

Are you 18 years old or older?

Yes

No

← →

English

What is the brand of **your primary mobile phone**?

If you have a private mobile phone and a business mobile phone, please answer these questions for your private one.

- Samsung
- Apple (iPhone)
- LG
- Motorola
- Google
- Blackberry
- NOKIA
- ZTE
- Huawei
- SONY
- Other (Please specify)
- Unsure

→

English

Are you familiar with Google Timeline?

In Google Maps App, location history data are collected passively by your mobile devices. Your Google Timeline tells you the places you've been, the kilometers you have walked, the hours you have spent in the vehicle and so on.

←

→

English

Do you currently have Location History enabled on your primary mobile phone?

Here are the instructions on how to check whether Location History is enable in your mobile phone.

*For **iPhone users**, please click here to get the instructions.*

*For **other mobile phone users**, please click here to get the instructions.*

Yes

No

← →

English

First, we have a few questions *about your transportation-related characteristic.*

If you want to change your answer to a previous question, you may go back by clicking the Back Button on the bottom left of each page. Do not use your browser's back button.

→

Portland State UNIVERSITY

Technische Universität München

English

Do you have a current:

	Yes	No	Unsure/Prefer not to answer
Driver's License	○	○	○
Transit Pass (e.g. weekly/monthly/yearly public transit ticket, semester ticket)	○	○	○
Car-sharing membership (e.g. Car2Go, Drive Now, Sixt Share, etc.)	○	○	○
Bike-sharing membership	○	○	○
Ride-share account (e.g. Uber)	○	○	○

←

→

Portland State UNIVERSITY

Technische Universität München

English

How many **automobiles (cars and motorcycles)** can you access for your personal use? Only count those in working condition that are privately owned or leased by you or people you live with.

- 0
- 1
- 2
- 3
- more than 3
- Unsure/Prefer not to answer

How many **bicycles** can you access for your personal use? Only count those in working condition that are privately owned or leased by you or people you live with.

- 0
- 1
- 2
- 3
- more than 3
- Unsure/Prefer not to answer

← →

English

Next, we have a few questions *about you and your household* so that we may understand the characteristics of our survey respondents. We will keep this information confidential and it is not linked to your name.

←

→

English

What is your age?

18 - 24

25 - 34

35 - 44

45 - 54

55 - 64

65 - 74

75 - 84

85 or older

Unsure/Prefer not to answer

←

→

English

How do you describe yourself?

- Female
- Male
- Transgender or other
- Unsure/Prefer not to answer

What is your current employment status? (Choose all that apply)

- I am predominantly employed outside my home.
- I work predominantly from home.
- I go to school/university.
- Not employed/enrolled at this time.
- Retired.
- Other (Please specify)
- Unsure/Prefer not to answer

←

→

English

INCLUDING YOURSELF, how many people currently live in your household?

- 1 (Just me)
- 2
- 3
- 4
- 5
- more than 5
- Unsure/Prefer not to answer

How many people in your household are ***under 6 years old?***

- 0
- 1
- 2
- 3
- more than 3
- Unsure/Prefer not to answer

← →

Portland State UNIVERSITY

Technische Universität München

English

How many people in your household are ***between 6 and 18 years old?***

0

1

2

3

more than 3

Unsure/Prefer not to answer

How many people in your household are ***employed (full-time or part-time)***?

0

1

2

3

more than 3

Unsure/Prefer not to answer

←

→

English

Including yourself and the people you share finances with, what is your approximate total monthly income (before taxes)?

under 1,000€	5,000€ - 6,000€
1,000€ - 2,000€	6,000€ - 7,000€
2,000€ - 3,000€	7,000€ or more
3,000€ - 4,000€	Unsure/Prefer Not to Answer
4,000€ - 5,000€	

← →

English

Thank you for your response! You are about halfway done!

Now, we would like to ask you to share your Google Timeline data which is passively collected by your smart phone. It will take approximately five minutes to complete.

← →

English

Please follow the instructions to *export your Google Timeline data*, then upload the zip file here.

- Please click here to get the instructions.
- If you uploaded a wrong file by mistake, don't worry. Please click any area in the grey box below, then you can re-select the correct file.

Drop files or click here to upload

← →

English

Thank you for your response! You have finished the first part of the survey!

The Part II survey will take approximately 7 to 8 minutes to complete. Afterwards, you will receive an entry into the draw to win one of ten Amazon gift cards.

← →

English

In part II of the survey, you will be asked about changes you have experienced *since January 1st 2018*, including changes to your:

- Household size or structure
- Transportation resources
- Employment status or location
- Student status

←

→

Portland State UNIVERSITY

Technische Universität München

English

Did any of the following event(s) happen to you *since January 1st 2018*? (Check all that apply)

A family member/partner/roommate moved out of my residence.

A family member/partner/roommate (who lived with me) passed away.

A family member/partner/roommate moved into my residence.

A new child/children arrived in my household (through birth, adoption, step child etc.).

None of the above events happened to my household.

Other event(s) that changed the number of adults or number of children in my household (Please specify):

←

→

English

Did any of the following event(s) happen to you *since January 1st 2018*? (Check all that apply)

Got access (i.e., bought yourself, received as a gift, rented long-term) to a car/motorcycle.	Registered for a car-sharing membership.
Lost access to a car/motorcycle.	Got rid of car-sharing membership.
Gained a driver's license.	None of these events.
Got rid of a driver's license.	Other event(s) that relate to automobiles ownership (Please specify):

← →

English

In which month(s) did these events happen? If it happened more than once, please select all the months that apply.

(For example, if one child was born in January and another was adopted in December, please check both 01/2018 and 12/2018.)

	01/2018	02/2018	03/2018	04/2018	05/2018
A family member/partner/roommate moved out of my residence.	☐	☐	☐	☐	☐

←

→

English

Did any of the following event(s) happen to you *since January 1st 2018*? (Check all that apply)

Got access (i.e., bought yourself, received as a gift, rented long-term) to a car/motorcycle.	Registered for a car-sharing membership.
Lost access to a car/motorcycle.	Got rid of car-sharing membership.
Gained a driver's license.	None of these events.
Got rid of a driver's license.	Other event(s) that relate to automobiles ownership (Please specify):

← →

English

In which month(s) did these events happen? If it happened more than once, please select all the months when it happened.

(For example, if you bought a car in January and another one in March separately, please check both 01/2018 and 03/2018.)

	01/2018	02/2018	03/2018	04/2018	05/2018
Got rid of car-sharing membership.	☐	☐	☐	☐	☐
Other event(s) that relate to automobiles ownership (Please specify):	☐	☐	☐	☐	☐

← →

English

Did any of the following event(s) happen to you *since January 1st 2018*? (Check all that apply)

Got a transit pass.

Got rid of a transit pass.

None of these events.

←

→

English

In which month(s) did these events happen? If it happened more than once, please select all the months when it happened.

	01/2018	02/2018	03/2018	04/2018	05/2018
Got a transit pass.	☐	☐	☐	☐	☐

←

→

Portland State UNIVERSITY

Technische Universität München TUM

English

Did any of the following event(s) happen to you *since January 1st 2018*? (Check all that apply)

- Got access to a bicycle.
- Lost access to a bicycle.
- Registered for a bike-sharing membership.
- Got rid of bike-sharing membership.
- None of these events.
- Other event(s) that relate to bicycles ownership (Please specify):

←　→

English

In which month(s) did these events happen? If it happened more than once, please select all the months when it happened.

(For example, if you bought a bike in January and another one in March separately, please check both 01/2018 and 03/2018.)

	01/2018	02/2018	03/2018	04/2018	05/2018
Got access to a bicycle.	☐	☐	☐	☐	☐
Registered for a bike-sharing membership.	☐	☐	☐	☐	☐

←

→

English

In which month(s) did these events happen? If it happened more than once, please select all the months when it happened.

(For example, if you bought a bike in January and another one in March separately, please check both 01/2018 and 03/2018.)

2019	09/2019	10/2019	11/2019	12/2019	01/2019	02/2020	Unsure/Prefer not to answer
☐	☐	☐	☐	☐	☐	☐	☐
☐	☐	☐	☐	☐	☐	☐	☐

←

→

Portland State UNIVERSITY

Technische Universität München TUM

English

Did you change your employment status *since January 1st 2018*?
- Internship is also considered.
- Changes in employment status indicates to start or lose a job, changes in job location or working hours, changes in job title and so on.

Yes

No

Unsure/Prefer not to answer

←

→

English

Which of the following event(s) happen to you *since January 1st 2018*?

I started a new job.	I increased my working hours.
I lost my job.	I decreased my working hours.
I changed my job location.	None of the above events.
I was promoted to a position with more responsibilities.	Other event(s) that relate to my employment status (Please specify):

← →

English

In which month(s) did these events happen? If it happened more than once, please select all months that apply.

	01/2018	02/2018	03/2018	04/2018	05/2018	06/2018	07/2
I started a new job.	☐	☐	☐	☐	☐	☐	☐
I changed my job location.	☐	☐	☐	☐	☐	☐	☐
I was promoted to a position with more responsibilities.	☐	☐	☐	☐	☐	☐	☐
I increased my working hours.	☐	☐	☐	☐	☐	☐	☐

←

→

English

Did you change your student status *since January 1st 2018*?
- Study status indicates changes in enrollment status, school location and so on.

Yes

No

Unsure/Prefer not to answer

←

→

Portland State UNIVERSITY

Technische Universität München

English

Did the following event(s) happen to you *since January 1st 2018*?

I started attending a school/university.

I graduated from school/university.

I dropped out of school/university.

I changed my school/university location.

None of the above events.

Other event(s) that relate to my study status (Please specify):

←

→

English

In which month(s) did these events happen? If it happened more than once, please select all months that apply.

	01/2018	02/2018	03/2018	04/2018	05/2018	06/2018	0…
I started attending a school/university.	☐	☐	☐	☐	☐	☐	
I graduated from school/university.	☐	☐	☐	☐	☐	☐	
I dropped out of school/university.	☐	☐	☐	☐	☐	☐	
I changed my school/university location.	☐	☐	☐	☐	☐	☐	

← →

Portland State UNIVERSITY

Technische Universität München

English

Congratulations, you made it to the end! As a thank you, we are offering you the chance to win an Amazon gift card. Are you interested in entering this draw for a gift card?

Yes

No

←   →

Portland State UNIVERSITY

Technische Universität München

English

We will need your email address to inform you if you have won the draw.

In order to keep survey responses anonymous, after you submit this survey, you will be *redirected to an independent form that will collect only your email address.* It will *not be tied to your survey responses.*

←   →

Portland State UNIVERSITY

Technische Universität München TUM

English

Thank you for your help! If you have any other thoughts/comments/feedback on this survey, please either include them below or contact us by email qin.zhang@tum.de.

Submit

Portland State UNIVERSITY

Technische Universität München TUM

English

Thank you for your support! Please provide an email address to enter the draw.

The information you provide is private and will be *treated confidentially*. It is being used *only for the purposes of this drawing*, and it will be *deleted immediately after prizes are awarded*.

Submit

Appendix 4: Information sheet of survey recruitment

Dear Sir or Madam,

We would like to invite you to participate in the research project "Exploring the correlation between individual's travel routine and life events using longitudinal travel diary data". The objective of this research is to use longitudinal travel data to observe changes individual's transportation patterns over time, with a specific interest in walking and health.

All participants will be invited to participation in the following stages:

1) The first part asks information about your socio-demographics, transportation resources and requests your location history data from your smartphone (since January 2018). Here, participants are asked to retrieve their location history from Google Timeline (see below) and upload it to a password-protected server. This research will only focus on the data of the years 2018 and 2019. You can fill in the survey on your PC or your mobile device. We will assign each participant a subject ID and all records will be coded with the subject ID rather than personal identifiers.
2) The second part is an online questionnaire that will take 7-8 minutes to complete. We ask information about your life events since January 2018.

Survey records and location history data collected in steps 1 and 2 will be filtered and aggregated by using automated scripts to record the mode, travel distance, and travel duration. All information will be stored on protected severs at Portland State University and the Technical University of Munich. All personal identifying information will be removed, and participants' identities will remain confidential. In addition, all participant information stored on our protected server will be identified by a randomly assigned subject ID only. Only the researchers collaborating in this research project and requiring access to this information will have access to these data and will be subject to a strict duty of confidentiality. At the conclusion of this project, all data records will be deleted.

We do not see any major personal risks in participating in this study. However, we do recognize that you will be asked to provide personal information about your transportation patterns and location history, which are personal and perhaps sensitive. Thus, the biggest risk is the disclosure of your personal location history. We are taking action to minimize this concern by storing the information on a secure server, limiting the access to this information to select members of the research team, anonymizing participant identity, filtering out only the data necessary for our study, and deleting information at the end of our study.

No personal information (e.g., name, address, phone number) will be collected in the whole procedure. Each participant will hold a subject ID and all survey records and location history data will be associated with this subject ID.

As a participant, you have the right to withdraw from the project at any time and request that your information be deleted, without any negative consequences for you.

Results of this research may be reported in scientific journals, meetings and conferences. None of these reports will show your personal information or data that can point to any person who took part in the study.

What is Google Timeline data?

Google timeline data makes use of the global positioning system (GPS) capabilities of your smartphone and track information about your travel patterns, including location, travel mode, travel duration, and activity. More information can be found on the Google Maps Help page: https://support.google.com/maps/answer/6258979?co=GENIE.Platform%3DDesktop&hl=en#

Further Questions?

You are welcome to ask any questions for clarification or additional information. For any information about this research, you can contact Qin Zhang (the Principal Investigator of this project) by email qin.zhang@tum.de or by calling 0049-89-289-22698.

Appendix 5: Univariate linear regression models for intrapersonal variability

Table A: Univariate linear regression models for intrapersonal variability in *total travel time* across different temporal scales (daily, weekly, and monthly)

*Significant variables are bolded (Intercepts are not highlighted as they are always significant)

	Day			Week			Month		
Variables	Coef.	P-value	R^2	Coef.	P-value	R^2	Coef.	P-value	R^2
Model - age			0.07			0.02			0.01
(Intercept)	0.97	0.00		0.61	0.00		0.31	0.00	
18 to 24	-0.18	0.48		0.02	0.93		-0.02	0.88	
25 to 34	0.08	0.66		0.10	0.54		0.01	0.85	
35 to 64	Base								
Model - gender			0.09			0.18			0.03
(Intercept)	1.13	0.00		0.82	0.00		0.35	0.00	
Male	**-0.19**	**0.13**		**-0.23**	**0.03**		-0.05	0.42	
Female	Base								
Model - occupation			0.03			0.01			0.14
(Intercept)	1.05	0.00		0.66	0.00		0.28	0.00	
Student	-0.12	0.37		0.06	0.58		**0.10**	**0.08**	
Worker	Base								
Model -household size			0.05			0.08			0.11
(Intercept)	1.13	0.00		0.74	0.00		0.30	0.00	
1	-0.16	0.37		0.03	0.86		0.09	0.23	
2	-0.17	0.29		-0.14	0.30		-0.01	0.87	
More than 2	Base								
Model - car access			0.08			0.18			0.24
(Intercept)	0.90	0.00		0.53	0.00		0.25	0.00	
No car access	**0.18**	**0.16**		**0.24**	**0.03**		**0.13**	**0.02**	
Has car access	Base								
Model - income group			0.01			0.13			0.47
(Intercept)	1.03	0.00		0.70	0.00		0.32	0.00	
High	0.04	0.83		-0.17	0.23		**-0.12**	**0.05**	
Low	-0.10	0.66		0.18	0.31		**0.19**	**0.01**	
Medium	Base								
Model - children			0.00			0.03			0.17
(Intercept)	1.03	0.00		0.57	0.00		0.21	0.00	
No children	-0.02	0.89		0.13	0.39		**0.14**	**0.05**	
Has children	Base								

Table B: Univariate linear regression models for intrapersonal variability in the *number of trips* across different temporal scales (daily, weekly, and monthly)

*Significant variables are bolded (Intercepts are not highlighted as they are always significant)

	Day			Week			Month		
Variables	Coef.	P-value	R^2	Coef.	P-value	R^2	Coef.	P-value	R^2
Model - age			0.01			0.01			0.02
(Intercept)	0.57	0.00		0.49	0.01		0.24	0.00	
18 to 24	0.00	0.99		0.07	0.77		0.05	0.61	
25 to 34	0.03	0.63		0.10	0.56		0.04	0.56	
35 to 64									
Model - gender			0.03			0.08			0.04
(Intercept)	0.62	0.00		0.68	0.00		0.24	0.00	
Male	-0.04	0.42		**-0.18**	**0.14**		0.05	0.36	
Female									
Model - occupation			0.01			0.08			0.15
(Intercept)	0.59	0.00		0.51	0.00		0.24	0.00	
Student	0.02	0.70		**0.18**	**0.15**		**0.09**	**0.07**	
Worker									
Model -household size			0.05			0.11			0.35
(Intercept)	0.64	0.00		0.66	0.00		0.26	0.00	
1	-0.04	0.60		0.01	0.95		**0.13**	**0.03**	
2	-0.07	0.26		-0.20	0.17		-0.03	0.51	
More than 2									
Model - car access			0.18			0.18			0.18
(Intercept)	0.53	0.00		0.41	0.00		0.22	0.00	
No car access	**0.11**	**0.03**		**0.26**	**0.03**		**0.10**	**0.04**	
Has car access									
Model - income group			0.11			0.32			0.28
(Intercept)	0.60	0.00		0.56	0.00		0.27	0.00	
High	-0.06	0.35		**-0.21**	**0.14**		**-0.07**	**0.24**	
Low	**0.09**	**0.29**		**0.42**	**0.02**		**0.13**	**0.06**	
Medium									
Model - children			0.01			0.04			0.03
(Intercept)	0.57	0.00		0.42	0.01		0.23	0.00	
No children	0.03	0.66		0.18	0.30		0.06	0.40	
Has children									

Table C: Univariate linear regression models for intrapersonal variability in *total walk time* across different temporal scales (daily, weekly, and monthly)

*Significant variables are bolded (Intercepts are not highlighted as they are always significant)

Variables	Day			Week			Month		
	Coef.	P-value	R^2	Coef.	P-value	R^2	Coef.	P-value	R^2
Model - age			0.03			0.05			0.02
(Intercept)	1.23	0.00		0.68	0.00		0.45	0.00	
18 to 24	0.18	0.52		0.18	0.55		0.03	0.75	
25 to 34	0.17	0.39		0.23	0.29		-0.02	0.75	
35 to 64									
Model - gender			0.06			0.17			0.00
(Intercept)	1.47	0.00		1.06	0.00		0.44	0.00	
Male	**-0.17**	**0.21**		**-0.31**	**0.03**		0.00	0.96	
Female									
Model - occupation			0.01			0.02			0.10
(Intercept)	1.35	0.00		0.84	0.00		0.41	0.00	
Student	0.06	0.68		0.10	0.53		**0.07**	**0.14**	
Worker									
Model -household size			0.07			0.05			0.15
(Intercept)	1.51	0.00		0.94	0.00		0.41	0.00	
1	-0.15	0.41		0.03	0.90		**0.11**	**0.10**	
2	-0.21	0.21		-0.15	0.41		0.01	0.81	
More than 2									
Model - car access			0.01			0.05			0.04
(Intercept)	1.42	0.00		0.76	0.00		0.41	0.00	
No car access	-0.08	0.58		0.18	0.24		0.05	0.34	
Has car access									
Model - income group			0.09			0.14			0.35
(Intercept)	1.31	0.00		0.84	0.00		0.41	0.00	
High	0.07	0.70		-0.13	0.50		-0.04	0.45	
Low	**0.31**	**0.16**		**0.39**	**0.12**		**0.17**	**0.01**	
Medium									
Model - children			0.00			0.03			0.08
(Intercept)	1.42	0.00		0.70	0.00		0.37	0.00	
No children	-0.06	0.77		0.20	0.35		0.09	0.19	
Has children									

Table D: Univariate linear regression models for intrapersonal variability in the *number of walk trips* across different temporal scales (daily, weekly, and monthly)

*Significant variables are bolded (Intercepts are not highlighted as they are always significant)

	Day			Week			Month		
Variables	Coef.	P-value	R^2	Coef.	P-value	R^2	Coef.	P-value	R^2
Model - age			0.02			0.03			0.06
(Intercept)	0.87	0.00		0.58	0.01		0.29	0.00	
18 to 24	0.03	0.87		0.13	0.66		0.10	0.29	
25 to 34	0.07	0.55		0.17	0.43		0.05	0.48	
35 to 64									
Model - gender			0.05			0.12			0.02
(Intercept)	0.99	0.00		0.88	0.00		0.32	0.00	
Male	-0.10	0.26		**-0.26**	**0.08**		0.03	0.54	
Female									
Model - occupation			0.00			0.06			0.24
(Intercept)	0.93	0.00		0.66	0.00		0.30	0.00	
Student	0.00	0.98		**0.20**	**0.21**		**0.11**	**0.02**	
Worker									
Model -household size			0.06			0.11			0.33
(Intercept)	1.01	0.00		0.86	0.00		0.35	0.00	
1	-0.09	0.46		-0.04	0.85		**0.09**	**0.12**	
2	**-0.13**	**0.22**		**-0.27**	**0.14**		**-0.07**	**0.16**	
More than 2									
Model - car access			0.01			0.09			0.12
(Intercept)	0.95	0.00		0.58	0.00		0.30	0.00	
No car access	-0.04	0.63		**0.23**	**0.12**		**0.08**	**0.11**	
Has car access									
Model - income group			0.16			0.32			0.43
(Intercept)	0.89	0.00		0.68	0.00		0.33	0.00	
High	0.06	0.61		-0.14	0.41		-0.06	0.22	
Low	**0.27**	**0.06**		**0.60**	**0.01**		**0.18**	**0.01**	
Medium									
Model - children			0.00			0.02			0.00
(Intercept)	0.94	0.00		0.59	0.00		0.32	0.00	
No children	-0.02	0.90		0.16	0.45		0.02	0.76	
Has children									

Appendix 6: Descriptive Analysis of Travel Behavior Changes by Event Types

Category	Event type	Cases	Sample id	Occurrence		Changes in …			
				Month	Year	Monthly travel time	Monthly walk time	Monthly cycle time	Monthly PA volume
Household events	Moved home	1	12	9	2018	-27%	24%	168%	30%
	Person moved in	1	24	6	2018	-8%	-47%	812%	-32%
	Person moved out	2	11	8	2019	68%	1396%	-	1415%
			24	12	2018	-58%	-75%	-10%	-61%
Job events	Changed job location	3	19	10	2018	-54%	-15%	-5%	-15%
			20	9	2019	48%	-1%	-	12%
			24	8	2018	1%	81%	11346%	131%
	Lost job	1	12	9	2018	-27%	24%	168%	30%
	Promotion	3	12	7	2018	-21%	-62%	6162%	-59%
			13	2	2019	-18%	47%	669%	172%
			29	9	2019	105%	143%	473%	169%
	Started new job	3	20	9	2019	48%	-1%	-	12%
			24	8	2018	1%	81%	11346%	131%
			28	8	2019	41%	-8%	172%	33%
	Decreased work hour	1	20	4	2019	-11%	-7%	-100%	-7%
	Increased work hour	3	13	8	2018	-42%	-21%	188%	9%
			24	5	2019	83%	3%	62%	34%
			29	9	2019	105%	143%	473%	169%
Mobility events	Registered bike sharing	1	28	7	2019	19%	-22%	229%	38%
	Got bike access	5	12	5	2019	270%	512%	-11%	387%
			16	4	2019	45%	93%	-33%	21%

			19	3	2019	8%	-22%	6423%	136%
			20	9	2019	48%	-1%	-	12%
			28	2	2019	56%	19%	340%	39%
	Lost bike access	1	20	5	2018	-9%	21%	-18%	18%
	Lost car access	2	11	8	2019	68%	1396%	-	1415%
			12	5	2018	-30%	-3%	142%	0%
School events	Graduated from school/university	3	18	9	2019	-23%	-14%	-	30%
			20	4	2019	-11%	-7%	-100%	-7%
			24	3	2019	231%	1376%	1686%	1595%
	Started school/university	2	8	4	2018	-43%	-88%	-	-88%
			12	10	2018	-36%	304%	373%	314%