MSc Urban, Port and Transport Economics
Erasmus School of Economics
Erasmus University Rotterdam

What makes people enjoy walking?

Physical determinants of perceived walkability in Rotterdam

Master thesis by:
Benjamin Majcen (481582)

Supervisor:
Anna Bornioli, PhD

Second assessor:
Susan Vermeulen, MSc

Date final version:
30th of April 2022

Erasmus
University
Rotterdam

Abstract

How does the physical environment explain where we enjoy walking? In 2019, the municipality of Rotterdam launched a crowdsourced “heart map”, giving pedestrians the chance to mark their preferred and less preferred walking spots across the city. This paper juxtaposes the “hearts” and “broken hearts” of this map with a wide range of spatial data on the city’s physical environment, obtained from BGT and CBS. After transforming the spatial data by means of GIS-based landscape metrics, two hedonic utility regressions (basic & index-based regression) have been employed. The density of green areas and the proximity of water courses were found to have a significant positive effect on the pedestrians’ perceived walkability, while the proximity of non-pedestrian traffic infrastructure like arterial roads and public transit rails was found to be significantly negative.

Table of Contents

A. Introduction

Despite serious efforts to promote more active modes of transportation, many cities continue to accommodate heavy rates of motorization, putting the livability and accessibility of urban space at stake at risk (Nieuwenhuijsen & Khreis, 2016). High motorization rates are associated with a wide range of challenges, including higher rates of air pollutant concentration (Health Effects Institute, 2021), higher local noise levels (Foraster et al., 2011), physical inactivity (Leslie et al., 2007) and a higher frequency of fatal traffic accidents (WHO, 2020). In the wake of rising geopolitical tensions, it should also be noted that the predominance of car mobility comes with an increased reliance on foreign oil reserves (Santos et al., 2010). Yet, with 30% of all urban car journeys in Europe covering distances below three kilometers, many could potentially be replaced by more active mobility modes (WHO, 2022). One such active mobility mode is walking, an "essential enabler of sustainable development, contributing to both the people and the environment" (Liao et al., 2020).

To harness the multiplicity of benefits associated with walking, academia has devoted increasing attention to making cities more *walkable* in the past decade and a half (Grasser, Van Dyck, Titze & Stronegger, 2016). Although *walkability* has taken on various definitions in past literature, most studies refer to walkability as a measure to denote the conduciveness of the built environment to walking as a mode of transport: Leslie et al. (2007) define walkability as "the extent to which characteristics of the built environment and land use may or may not be conducive to residents in the area walking for either leisure, exercise or recreation". Frank et al. (2009) consider walkability as a measure to isolate the effect of the urban form on walking behaviour from the effect of other determinants such as socio-demographic characteristics.

While the most prominent research on the determinants of walkability originates from the United States, a rising number of studies are arising in other parts of the globe too (Grasser, Van Dyck, Titze & Stronegger, 2016). Findings of such studies have identified various determinants – including household density, job density, intersection density, population density, retail floor area ratio, land-use mix, proximity of supermarkets and the proximity of green areas. Yet, most of these studies take on either an objective approach, where they seek to draw a correlation between physical characteristics and measured walking rates, or a subjective approach, where communities are asked to report their perception on the walkability of their neighbourhoods (Liao, Van Den Berg, Van Wesemael, & Arentze 2020). This research strives to contribute to existing literature in the field, by reaping the benefits of both objective and subjective approaches: Drawing on data originating from a crowdsourced heart map composed during the 2019 "Do you want to walk with me?" campaign in Rotterdam (where respondents placed hearts or broken hearts on a map according to whether they found the place walkable or not), this study strives to explain the subjective, perceived walkability across Rotterdam with objective, GIS-based physical characteristics – by means of a basic multiple linear regression analysis and an index-based hedonic utility model.

The following central research question has been formulated:

- **RQ: "How does the physical environment explain perceived walkability?"**

The central question is further divided into five sub-questions, aimed at identifying the kinds of physical environments that have the most important role in explaining walkability. The first sub-question looks into the association between the traffic infrastructure and perceived walkability. The second sub-question focuses on the role of natural features like green areas and water bodies. The third sub-question investigates the role of built structures. The fourth sub-question examines the role of the density of essential amenities like daily good stores, schools and libraries. Finally, the fifth sub-question shifts focus on the diversity of land use, i.e. how the land-use mix explains perceived walkability.

Research question: "How does the physical environment explain perceived walkability?"

- SQ1: "Does the distribution of traffic infrastructure explain perceived walkability?"
- SQ2: "Does the distribution of natural features explain perceived walkability?"
- SQ3: "Does the distribution of built structures explain perceived walkability?"
- SQ4: "Does the distribution of essential amenities explain perceived walkability?"
- SQ5: "Does the diversity of land use explain perceived walkability?"

The remainder of this paper is organized as follows: First, existing literature will be reviewed, centered on definitions, measurement approaches, benefits and policy implementations of walkability. In the subsequent chapter, the spatial data on Rotterdam's physical environment will be presented and visualized. Thereafter, the employed landscape metrics and research methodology will be explained. This is followed by a presentation of the results, possible interpretations, and the associated limitations.

B. Literature Review: Past attempts to capture walkability

a. What makes a city walkable? Defining walkability

In the past fifteen years, a growing number of studies have examined the association between the built environment and physical activity. Research in this field predominantly points at the mechanisms explaining how community design can positively influence local rates of active modes of transport, and how policy can implement the results of this research to create environments that are conducive to physical activity (Sugiyama et al., 2010). Since walking is the most prevalent mode of active transport worldwide, studies that examine the physical determinants of walking rates have already received substantial attention from academia, policy and beyond (Grasser, Van Dyck, Titze & Stronegger, 2016). However, with a growing range of stakeholders actively striving to harness the benefits of walking, studies have taken on different perspectives on what makes a place conducive to walking in first place. As a result, the definitions and measurements of walkability vary according to the area of study and the context of the research.

In the reviewed literature, there is a consensus that walkability is a measure that defines walking rates from physical characteristics of an area. Liao, Van Den Berg, Van Wesemael & Arentze (2020) give a rather general definition of walkability, presenting it as "a measure of how friendly an area is to walking". Leslie et al. (2007) define walkability as "the extent to which built environment characteristics and land use may or may not be conducive to area residents walking for leisure, exercise, access service, travel or work". Meanwhile, some papers define walking rates rather as the result of the urban design: Brownson (2009) view walkability as the "proximity and directness of routes from home to destinations". Speck (2012) argues that "where most people own a car, you need to offer a walk that is just as good or better than a drive". Speck's *General Theory of Walkability* summarizes walking rates in a city as the result of four dimensions: Firstly, the residents must have a "good reason to walk", in that conveniences are attainable. Secondly, the city must be – and especially feel – safe: As such, distraction of automobiles is an aspect that can truly hinder the walkability of an area. Thirdly, Speck argues that the walk must feel comfortable, which can be achieved through sufficient shade by trees in warm place. Finally, the walk must be interesting, which is why the urban design should accommodate a sense of humanity and stimulate pedestrian interactions. Frank et al. (2006) emphasize that walkability measures specifically aim to capture the effects of other determinants than socio-demographic factors such as age, income, or ethnicity.

Cervero & Kockelman (1997) categorize the analysis of transportation mode choices into three constituents by means of a '3D' model: The first 'D' stands for 'density' (which is further divided into household and job density), 'diversity' (which is determined by land-use mix) and 'design' (which is further divided into intersection and street density). Subsequent walkability analysis has frequently drawn upon this conceptual model, with research expanding the scheme by including constituents such as 'destination' (which is determined by the proxy of job accessibility) or 'retail floor area ratio' (Ewing, 2009). Grasser, Van Dyck, Titze & Stronegger (2013) conceptualize walkability as a measurement constituted by two aspects: The first aspect is 'proximity', which Frumkin et al. (2004) define as the result of density (i.e. the "quantity of people, households or jobs distributed over a unit of area") and land-use mix components (i.e. the "measure of how many types – offices, housing, retail, entertainment, services, and so on – are located in a given area"). The idea is that mixed neighborhoods with low distances from amenity to amenity are more convenient for pedestrians. The second aspect is 'connectivity', which Leslie et al. (2007) denote as the "directness of the pathway between households, shops and places of employment and is based on the design of the street network". The density and connectivity of places is thereby often a function of how old the built environment and urban planning is: Older European cities have, for example, often a higher density of amenities and traffic infrastructure than cities with more recent roots (Grasser et al., 2016).

Depending on the context of the research and the methodology used, studies have come up with further, more specific, determinants of high walking rates: As such, Glazier et al. (2012) and identify retail density as one such determinant. In a Netherlands-based empirical research, Liao, Van Den Berg, Van Wesemael & Arentze (2020) add the number of cafeterias within 1 km and total inland water to their list. Wagtendonk & Lakenveld (2019) also add the density of green areas to their index. Bauman et al. (1999) finds a correlation between walking rates and the proximity to the coast. Humpel et al. (2002) formulate a list of further determinants, including street lighting, moderate differences in altitude (hills) as well as specific types of trees.

Figure 1 conceptualizes the walkability components cited throughout the definitions according to their thematic fields. The traffic environment has been identified as a first thematic field, containing components such as route proximity, traffic safety, intersection density and street lighting density. The second thematic field is the natural environment, referring to components associated with green areas or water bodies. The third thematic field addresses the components related to the built environment, such as the density of building kinds and associated measurements such as floor-area ratios and architectural periods (alluding to higher density as well). A series of practical amenities are thereafter aligned under the "amenities" categorical field. Finally, the land-use mix is identified as another cited walkability component, broadly denoting the evenness of the abovementioned categories in the landscape. Finally, the figure emphasizes the purpose of walkability measures to isolate the effect of the physical environment from possible socio-economic determinants.

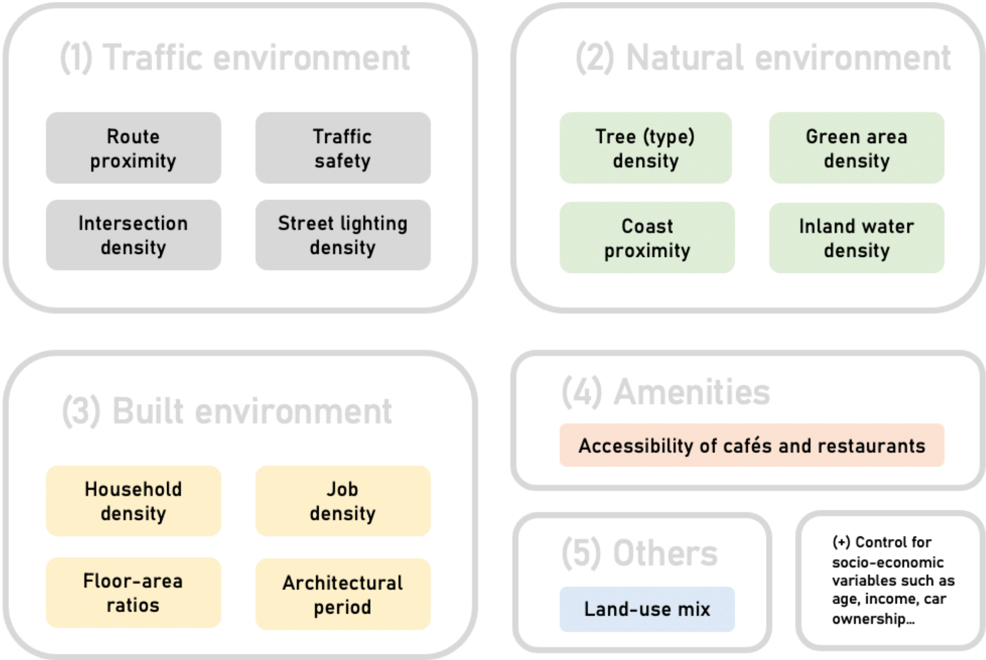

Figure 1: Conceptualization of cited walkability determinants

b. What makes a city walkable? Measuring walkability

To measure walkability, a growing number of research has been published aimed at identifying specific determinants in the physical environment that are conducive to higher walking rates (of which many eventually constitute a so-called *walkability indexes*). These are receiving increasing attention from fields including urban design, transport policy, public health, social anthropology, and urban economics. However, despite the progress that has been made since the eruption of the first attempts, existing walkability indices are still subject to various limitations (Liao, Van Den Berg, Van Wesemael & Arentze, 2020): As such, most indices select variables and assign weights to them merely based on theoretical considerations, with lacking evidence that these methods accurately represent the impact of physical characteristics of the environment on walking rates. Moreover, with many indices resulting from cross-sectional regression analyses, the studies often fail to account for possible self-selection effects: Cao et al. (2009) argue that self-selection is likely to occur because people who prefer walking will be inclined to live in neighborhoods with high walkability. Furthermore, with a vast variety of physical determinants that can be considered in a walkability study, the majority of studies acknowledge the possibility of omitted variables, e.g. crime density, traffic volumes, air quality or street furniture (Wagtendonk & Lakerveld, 2020). On the other hand, other measurement attempts, including the most notorious walkability indices (e.g. Frank et al. (2010)) include measures that are strongly correlated, which can lead to a misestimation of separate effects of the built environment characteristics on walking rates (Liao et al., 2020). Liao et al. identify two broad walkability measurement approaches – (1) *objective* and (2) *subjective* methods.

(1) Objective methods

Objective methods (e.g. walkability indices) measure walkability by describing the built environment as a total of a number of variables. Arguing that subjective measurement approaches predominantly rely on self-reported perceptions of environmental attributes that are juxtaposed to self-reported physical activity, Leslie et al. (2007) stresses the higher acuity enabled by objective measurement methods aimed at understanding the relationship between physical characteristics of an area and its walking rates.

Liao et al. (2020) further categorize objective methods into theory-driven versus data-driven approaches: While theory-driven approaches describe the built environment as a set of variables based on conceptual definitions of walkability, data-driven approaches strive to empirically derive a walkability measure from objective, geographical or GIS-based data. The most widely known example of a theory-driven approach is the walkability index by Frank et al. (2005): Drawing on theory including Ewing & Cervero (2001)'s findings on urban sprawl and the factor analysis on transportation choice mode by Cervero & Kockelman (1997), the research merges the most widely cited physical determinants of high walking rates into a walkability index. Departing from the same index, Grasser et al. (2013; 2016) systematically review previous walkability studies and adjust the weights assigned to the variables to a more European context. Other examples include Duncan et al. (2011), who depart from the notion of density and land-use mix and constitute an index capturing the walking distance to nine different types of amenities (restaurants, grocery stores, etc). Liao et al. (2020) note that theory-based walkability indices often suffer from the correlated hypothesized explanatory variables, e.g. the land use measures 'density' and 'connectivity' in the index by Frank et al. (2005), leading to either an over- or underestimation of separate or pure effects of the respective variables. Moreover, partly basing themselves on fully conceptual definitions of walkability, theory-driven approaches frequently do not provide sufficient evidence on the selection and weighting of the picked factors (Liao, Van den Berg, Van Wesemael & Arentze, 2020).

Meanwhile, data-driven approaches seek to empirically derive to an index capturing an area's friendliness to walking. Using geographical data in Iran, Habibian & Hosseinzadeh (2018) for example employ a linear regression analysis to estimate a fitting model, specifying it according to the highest goodness of fit. As acknowledged by the authors, the disadvantage of data-driven approaches, is their lack of external validity of the concluded results: Studying the explanatory physical characteristics in the city of Rasht (Iran), the model cannot account for peculiarities or cultural affinities in another region the model could possibly be applied to (Liao, Van den Berg, Van Wesemael & Arentze, 2020).

(2) Subjective methods

In the reviewed literature, there is a reoccurring pattern of explanatory factors stressing the importance of perception, e.g. perceived fear of crime (Blečić et al., 2016) or perceived comfort (Lizárraga, Martín-Blanco, Castillo-Pérez & Chica-Olmo, 2022). Examining self-reported perceptions of environmental attributes can therefore have its advantages, as such methods capture more qualitative values in measuring an area's walkability, such as perceived street safety or aesthetics and incorporates variations in cultural affinities across the areas of study, improving the findings' external validity (Park, Deakin & Lee, 2014).

Subjective methods use individuals' perception of the degree of walkability of their environment as a starting point, usually measured through a questionnaire (Liao, Van den Berg, Van Wesemael & Arentze, 2020). As such, Saelens et al. (2003) developed the Neighborhood Environment Walkability Scale (NEWS), which is a questionnaire featuring 68 criteria to evaluate walkability. Building on this scale, Cerin et al. (2006) developed the simplified NEWS-A, limiting the questionnaire to eight environmental characteristics, including widely cited determinants such as residential density, proximity to non-residential land uses, street connectivity, density of walking facilities - but also more dimensions that are more easily measured through a questionnaire, such as the respondents' perceived ease of access to non-residential land uses, aesthetics, pedestrian traffic safety and crime safety.

(3) People-based methods

Most of the reviewed definitions and measurement approaches of walkability define the concept as a measure of a neighborhood's community, thus predominantly comparing walking rates across neighborhood communities and, upon controlling for socio-demographic variables to solely focus on the physical determinants, drawing conlusions on the determinants of a neighborhood's walkability. Referring to the abovementioned definition, again, Brownson (2009) has defined walkability as the "proximity and directness of routes from *home* to destinations". Brownson therefore relates the walkability of a community to the adequacy of the community's residences to walk from or to. Similarly, Wagtendonk & Lakerveld (2019) capture solely the residences of the respondents in their index, not the location of the measured walkability itself.

(4) Place-based methods

Liao, Van den Berg, Van Wesemael & Arentze (2020) consider walkability "a characteristic of the neighborhood". Therefore, the unit of analysis throughout their study is the neighborhood. As such, all variables are measured on a neighborhood basis – or in the Dutch context, on a postcode area basis. Examples of variables include the average number of daily walking trips per person in each neighborhood (dependent variable), the average distance from a list of daily facilities in each neighborhood (candidate physical environment independent variable) or the percentage of households with incomes below €9249 in each neighborhoood (socio-economic control variable).

In conclusion, a categorization of the reviewed analyses of walkability is provided in Figure 2, dividing the framework into (a) objective and subjective approaches, (b) theory- and data-driven approaches and (c) people- and place-based approaches.

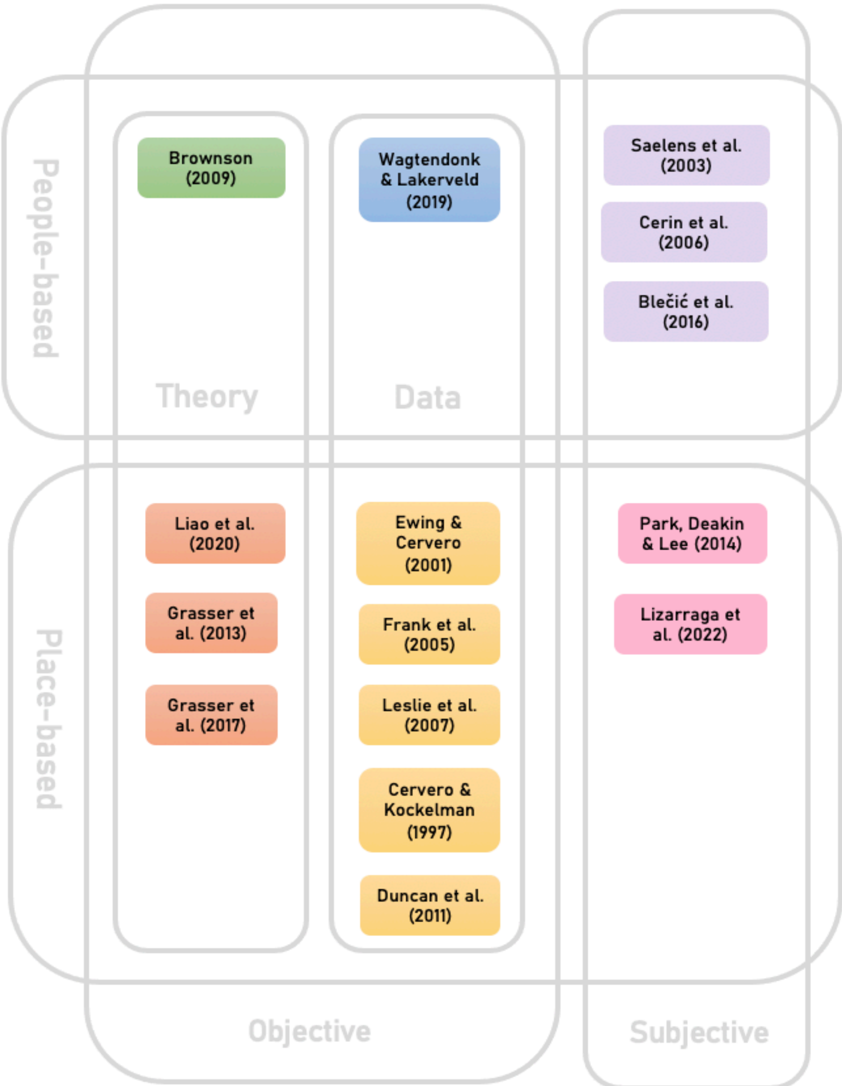

Figure 2: Categorization of the reviewed analyses of walkability

c. Benefits of walkability

(1) Economic benefits

With a wide range of papers addressing walkability from a public health perspective, the health benefits of walkability and associated gains in public spending have received considerable attention: In a study researching the effect of walkability on obesity across the city of San Diego (United States), the walkscores assigned to the different neighborhoods were found to be highly correlated with overweight likelihood (Jones, 2010). Although pedestrians are more exposed to automobile exhaust than automobile drivers themselves in car-intensive environments, the positive effects of daily physical activity such as walking have been found to outweigh the negative impact of poor air quality (College van Rijksadviseurs, 2020). Unwalkable, car-dependent cities have higher asthma rates caused by disproportionate automobile exhaust too (Speck, 2012). Meanwhile, active travel modes were found to induce better cognitive functioning (Yaffe et al., 2001). As a result, public health also pays its financial toll in areas with low walking rates: Decisio (2019) estimates, assuming that active mobility contributes by 20% to a 2% reduction in the number of diabetes patients, that higher walking rates can lead to €13 million of annual savings in public healthcare spending in the Netherlands.

In the field of transportation, walkability has been found to induce public savings by reducing the per capita use of financial resources in the field of transportation (WHO, 2018). Studies also suggest opportunities for consumer savings for members of walkable communities, including savings on vehicle expenses and associated ownership and operating costs (Litman, 2003).

In terms of urban development, walkability is also associated with saving opportunities. Litman (2003) finds cluster land use mixes to engender a substantially more economical and efficient use of land: With less viaducts to support arterial roads taking up urban space, the gained area can host diverse amenities or retail. Walkability improvements can also support regional economic development by shifting consumer expenditures from vehicle purchases and fuel expenses to the consumption other consumer goods that yield higher regional employment and business activity (Litman, 2003). Even in fuel-exporting Texas, consumer expenditures that are shifted from automobiles to a typical bundle of consumer goods provide 71% more regional income and twice the employment activity.

Tinessa, Pagliara, Biggiero & Delli Veneri (2020) find, in a case study conducted in Naples (Italy), that retail owners are willing to pay 3.80%-5.42% higher rents for each minute of walking closer to the pedestrian zone, whereby first activity retailers or bars and clubs (youth- and tourist-attracting places) are willing to pay even considerably more. Conversely, surveys of hypermarket and electronics store owners showed a decrease in bidding rents, fearing a lack of accessibility to customers arriving in cars. Mingardo (2019) also demonstrates that, though motorists spend more per shopping trip, pedestrians visit a shopping area much more often, which makes them economically more relevant. Hedonic price modelling has also shown that walkable neighborhoods can gauge higher residential as well as office real estate prices (Eppli & Tu, 2003), hinting a significant appreciation of walkable space.

(2) Environmental benefits

Active travel modes can also been presented as a strategy to mitigate climate change (Banai, 1998). As such, walkability can also substantially reduce greenhouse gas emissions (WHO, 2018). Especially in walkable cities, walking replaces short rides where energy consumption rates are the highest, per mile, particularly under urban peak conditions. Walkability is also an essential component of transit-oriented development (TOD) as pedestrian friendliness in a half-mile radius from transportation hubs substantially boosts the use of public transit (Canepa, 2017). With less traffic noise and arterial roads, divisions of natural habitats (flora and fauna) are also alleviated (EEA, 2016).

(3) Social benefits

Basing on a wide range of previous scholars, Adams (2006) brings forward that sustainability is not merely a construct combining economic, environmental, and social considerations. Instead, these three segments heavily depend upon each other. Rogers, Gardner & Carlson (2013) emphasize that no economy would survive, would there be no society, just like no society would survive, would there be no be environment. Still, with sustainability historically focusing on the mitigation of environmental risks and predominantly the interplay between environmental and economic considerations of sustainability, the social aspect of sustainability is often disregarded in associated studies, including in the field of walkability. Rogers, Gardner & Carlson (2013) find a strong correlation between walkability and social capital in their study on the social-environmental interface of communities, with walkability meaningfully contributing to social interaction and community building. In a study measuring the impact of urban sprawl on social capital, Nguyen (2010) finds widely cited indicators of social capital, such as volunteering rates or church attendance, to be significantly lower in sprawled areas. Members of walkable communities are also more likely to participate in community currency projects (Kwon, 2017). Litman (2003) emphasizes the inclusivity benefits associated with walkability: Walkable communities boost the accessibility to goods, services, and amenities to socially excluded population segments, such as license-less youth and pensioners, people who cannot afford motorized vehicles and – if the infrastructure is adequate – to wheelchair users, for example. Thereby, walking can also be regarded as a free option to do sports, as opposed to activities requiring a paid gym or sports club membership. Lo & Houston (2017) also find a positive association between compact, walkable urban development and gender equality: Observing that in heterosexual married couples, women conduct their activities further from home on average, their results indicate that compact development of neighborhoods gift married couples with a higher flexibility in terms of how they assign out-of-home household activities to each other, since destinations are more accessible and subject to a lower spatial fixity.

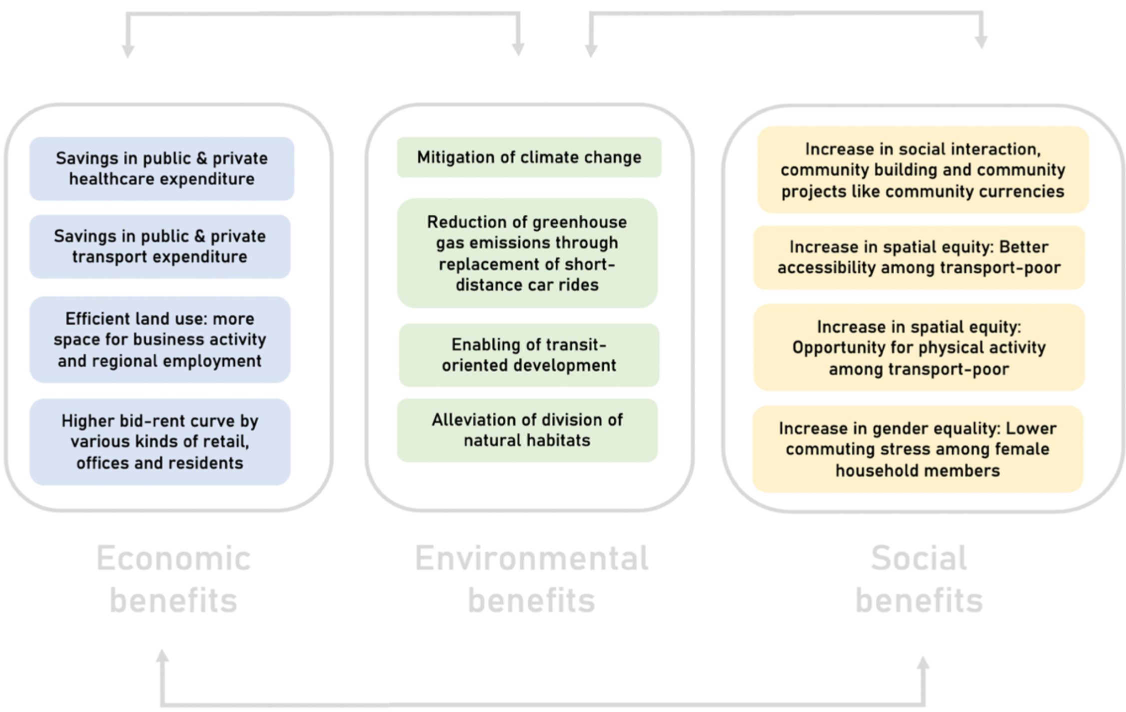

Figure 2: Contributions of walkability to the three pillars of sustainable urban development

Note: The arrows emphasize the contingent relationship between the three pillars

d. Insights from Rotterdam

Misled by the often instinctive and subconscious nature of walking in mobility, conventional transport planning often underestimates the role of walking as a mode of transport. Rietveld (2000) shows that the actual number of non-motorized trips is six times higher than indicated in the then-published Dutch travel survey. Litman (2003) argues that this is due to how transport is measured: Travel surveys often allow respondents to classify their trip only as one of the listed modes of transport, e.g. “automobile”, “public transit” or “walking”. Because respondents are inclined to select the mode which accounts for the largest share of the travel distance, the walking components before, after or in-between motorized segments of the trip are often not reflected in the survey results. Moreover, travel surveys also undercount walking because non-work-related travel, travel among children and recreational travel - all more likely to involve walking - are often disregarded in the conduct of these surveys. Litman (2003) therefore recommends policy to ask instead: “What portion of trips involve some walking?”.

Rotterdam’s Vice Mayor for Urban Mobility acknowledges the instinctive and subconscious involvement of walking in mobility: “When you head outside to hop on your bicycle, catch a tram or metro, or step into your car, you always start by walking. Sometimes it’s just a few steps, and sometimes you have to walk farther. Walking happens so instinctively that we hardly give it any thought” (Bokhove, 2020). With a significant increase in research on and investment into walkability across the municipality, the city uses on a systematic approach to boost walking among its citizens. In its first ever action plan Rotterdam Walks, the municipality takes on a broad view of who is a pedestrian: A pedestrian is defined as anyone who uses their feet to move from one place to another, as well as all other users of pavements and footpaths throughout the city, including people who use wheelchairs, mobility scooters, walking frames and white canes (Gemeente Rotterdam, 2020). With 41% of households in Rotterdam owning no car (CBS, 2019) and a total of 52 million kilometres walked from home to public transport stops and back (Decisio, 2019), the municipality is actively striving to “put the pedestrian on the pedestal” (Gemeente Rotterdam, 2020).

The Rotterdam Action Plan is divided into two “pillars” and three “tracks” denoted in Figure 3. While the bottom two “tracks” are predominantly related with boosting walking rates by mobilizing the different stakeholders (citizens, public organizations, companies, educational institutions, activists and more), the pillars “Healthy & Active” and “Accessible & Nearby” as well as the top track “Link & Place” also pay attention on the role of the landscape in determining walking rates, i.e. within the scope of this paper (marked in dark blue in Figure 3). These three elements are described below.

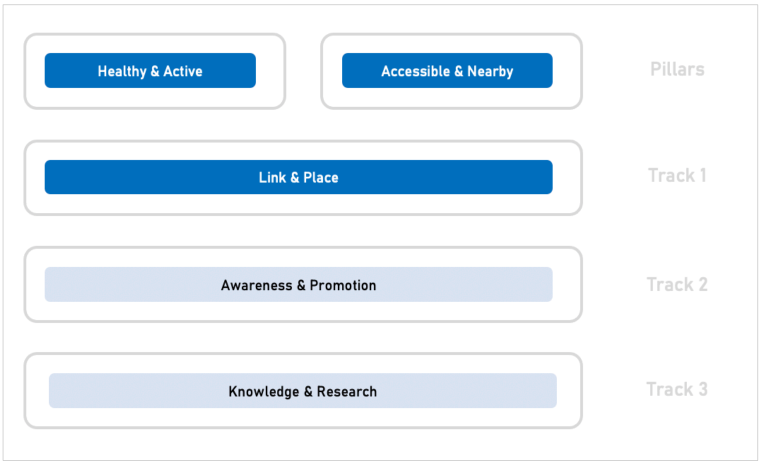

Figure 3: Pillars and tracks of the “Rotterdam Walks!” Action Plan (Gemeente Rotterdam, 2020)

(1) Pillar “Healthy & Active”

As part of the Rotterdam Walks Action Plan, the pillar “Healthy & Active” seeks to address the spatial disparities in community health across the city (Gemeente Rotterdam, 2020): The municipality draws this back to behavioural factors such as differences in lifestyle health (e.g. access to and conduct of physical activity as well as smoking), but also to factors related to the physical environment. Aware of the influence of the physical environment on transport behaviour, the municipality considers a wide range of associated aspects such as accessibility, convenience, noise exposure and traffic exposure. As such walkability has been a guiding principle in various urban development projects such as MerweVierhavens (M4H), Rotterdamsche Droogdok Maatschappij (RDM) or the “Heart of the South”.

(2) Pillar “Accessible & Nearby”

In the frame of the second of two pillars, “Accessible & Nearby”, the municipality centers on rendering walking facilities accessible to the widest possible public. As such, projects include safer crossings and designated walking routes for groups that are the most vulnerable in traffic such as children as well individuals with limited mobility or visual impairment (Rotterdam Veilig Vooruit, 2019). Besides traffic safety, the pillar also implies improvements in terms of social safety, approached among others through physical interventions such as a higher density of lighting to reduce exposure to abuse (Gemeente Rotterdam, 2020). Moreover, as a means to work on the accessibility, the municipality aims to improve the land-use mix, by ensuring the access to basic goods and services such as supermarkets, pharmacies, libraries, public transport hubs, schooling institutions, green and recreation areas, and playgrounds within a 15-minute walking radius.

(3) Track “Link & Place”

The track “Link & Place” concentrates on the density of essential and popular destinations and their linkage (Gemeente Rotterdam, 2020). As such, the municipality planned out a designated “green-blue” route network that links neighborhoods to green space and water bodies and avoids (perceived) barriers such as arterial roads and unattuned traffic lights. So-called “plus routes” thereby put a special emphasis on enabling a functional journey to people with limited mobility, by removing obstacles, placing benches, configuring traffic lights, adding audible signals, placing public toilets and accessibilizing public buildings on their way to essential destinations.

e. Hypotheses

To address the central research question ("How does the physical environment explain where we enjoy walking?") the paper has been divided into five sub-questions, serving as a guiding basis for the formulating of the hypotheses, which are banking on the literature review. The five sub-questions reflect the different kinds of physical characteristics considered in assessing if, and if so how, the physical environment can explain the perceived walkability in a PC5 area. The first sub-question investigates the role of the distribution of motorized traffic infrastructure on perceived walkability. Previous research has linked the presence of arterial roads and other landscape-disrupting infrastructure like railways. An increased proximity and density of such infrastructure might therefore be conducive to a lower perceived walkability. The first hypothesis is therefore:

H1: A higher proximity to motorized traffic infrastructure has a negative effect on perceived walkability.

In contrast, rich natural landscapes have continuously been cited as positively associated with perceived walkability. Previous papers have investigated various kinds of natural landscapes already including green areas like forests, hills, agrarian lands, and grasslands. Specific kinds of water bodies have also been associated with perceived walkability in other area studies. These include seawater, water courses and inland water areas. As a result, municipalities like Rotterdam have been stimulated to build on pedestrian traffic networks in areas close to, and with dense, green areas and water bodies. A higher presence (i.e. higher proximity to and/or density) of these types of natural landscapes are therefore hypothesized to explain higher perceived walkabilities. In other words, the second hypothesis is formulated as follows:

H2: A higher presence of natural features has a positive effect on perceived walkability.

The third sub-question investigates the role of built structures like buildings, but also various street furniture, on perceived walkability. Looking at past papers, research has regularly investigated the association between the density of (different kinds of) the built environment and walking preferences. While floor area ratios are, for example, used as a way to estimate the density of built structure, average building heights can also indicate how much space buildings offer as a proportion of how much land they use. Meanwhile, older cities have also been found to offer a denser land use. Perceiving average building heights and average building ages as a proxy for the density of the built environment, the third hypothesis is therefore:

H3: A higher density of built structures has a positive effect on perceived walkability.

The fourth sub-question focuses on the distribution of practical amenities in a given urban area. As part of its “Link & Place” track, the Rotterdam Walks Action Plan puts a particular emphasis on the accessibility of essential goods and services. As such, its plus routes, for example, strive to ensure a sufficient number of functional places and amenities like supermarkets, cafés, restaurants, cinemas and public transport hubs in the pedestrians’ walking radius. Empirical research also confirms the importance of the proximity and density of a series of amenities such as cafeterias in establishing favourable walking environments. As such, the fourth hypothesis is the following:

H4: A higher presence of practical amenities has a positive effect on perceived walkability.

Finally, the fifth sub-question shifts the focus of the paper on the land-use mix. A higher diversity in terms of the purpose land is used for has repeatedly been associated with higher walkability in past research, alluding to a better accessibility of destinations in such lands. Thus, although thematically related to the previous two sub-questions, this hypothesis looks at the role of the distribution of built space through the lens of a land-use mix entropy index. The fifth and final hypothesis is therefore:

H5: A higher land-use mix has a positive effect on perceived walkability.

D. Data collection and transformation

To measure how the physical environment can explain a pedestrian's perceived walkability, this study gathered a cross-sectional dataset composed of private and open-source spatial data. Data on perceived walkability has been retrieved from Rotterdam's Digital City Panel's crowdsourced heart map. Spatial data containing various categories of physical characteristics has been obtained by the Publieke Dienstverlening Op de Kaart (PDOK), NDW (Nationaal Dataportaal Wegverkeer) and Statistics Netherlands (CBS). The latter institution also provides a wide range of open-source land use and socio-demographic data, also used in this dataset. All entries in the spatial datasets located beyond 1.4km away from the municipality's boundary have been omitted from the analysis. Assuming a walking speed of 5.5km/h, 1.4km namely roughly corresponds to the distance made within the 15-minute radius, a widely used threshold for walking willingness (e.g. Hosford et al., 2022).

In the following section, the reporting on the data collection and transformation will be divided into six categories, namely (1) data on perceived walkability from the Digital City Panel's heart map, (2), data on land use by CBS, (3) data on the traffic environment by PDOK and NDW, (4) data on the natural environment by PDOK, (5) data on the built environment by PDOK, (6) data on amenities by CBS, and (7) data on socio-economic control variables by CBS.

(1) Data on perceived walkability - Digital City Panel

Firstly, the data to compute the dependent variable (perceived walkability) has been obtained from the "Wil je met me lopen?" ("Do you want to walk with me?") campaign by the Rotterdam Digital City Panel. During the campaign, which has been conducted in 2019, survey respondents have been asked to denote what stimulates them to walk and what stops them from walking. More precisely, the respondents placed three 'hearts' at spots they perceived as walkable, and three 'broken hearts' at a spot they perceived as unwalkable. The distribution of hearts and broken hearts is mapped in Fig. 4.

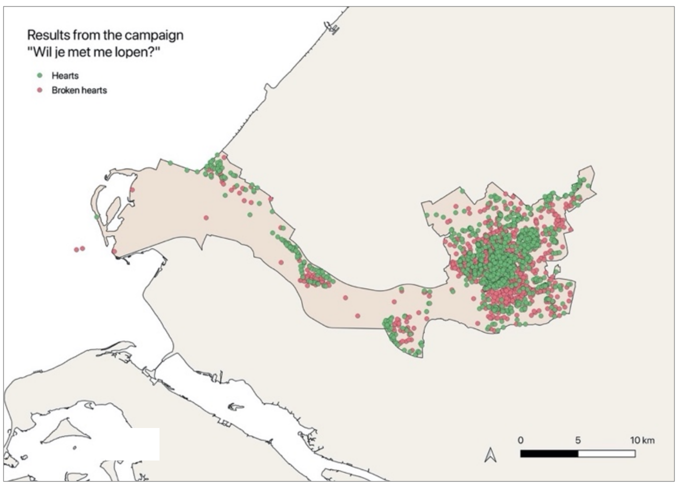

Figure 4: Distribution of hearts (walkable spots) and broken hearts (unwalkable spots) in Rotterdam (NL). Source: Digital City Panel (2019)

The respondent sample consisted of 1893 members of Rotterdam's Digital City Panel, i.e. citizens of Rotterdam who had participated in surveys by the municipality prior to this campaign and who indicated they would like to take part in upcoming surveys as well. When comparing the socio-demographic profile of the respondent sample with the overall population of Rotterdam, several discrepancies come to light: 55% of the respondents identify as male, 44% as female and 1% with another gender (Digital City Panel, 2019). However, only 49% of Rotterdam's actual population were male in 2019, i.e. at the time of the crowdsourcing campaign (CBS, 2019). The distribution of gender among the respondents is therefore somewhat male-skewed (see appendix, Figure A1). Furthermore, the average age of the respondents is above average in the context of Rotterdam: The age segments 45-54, 55-64, 65-74 and 75+ have been overrepresented, if we consider the age distribution across Rotterdam's overall population (see appendix, Figure A2). Furthermore, the campaign acknowledges the over-representation of highly educated individuals, as well as the over-representation of car, bike and public transport subscription owners, in the sample (Lopen in Rotterdam, 2020). The skewedness of the abovementioned socio-demographic characteristics points at possible biases in the upcoming parts of this study: CBS (2019) has found that the over-represented segments in this sample (people aged between 65 and 74, highly educated people and/or public transport subscription owners) all walk longer distances and times compared to the average population of the Netherlands. A final limitation of the perceived walkability dataset

To compute the dependent variable of the empirical analysis (elaborated in Section E), several transformations have been performed in QGIS: By using the "count-in-polygon" function, the number of hearts and broken hearts per PC5 area have been counted. Thereafter, a *perceived walkability score (PWS)* has been computed for every PC5 area (Equation 1). PC5 areas counting less than 5 answers (i.e. hearts + broken hearts) have been omitted from the dataset to reduce the influence of answers in PC5 areas with a low amount of responses (e.g. a PC5 area with 0 hearts and 1 broken heart would result in a PWS of 0, whereas a PC5 area with 1 heart and 40 broken hearts would result in a better PWS in spite of a wider difference in absolute terms between hearts and broken hearts). This would result in a bias in estimating the variations explained by physical characteristics.

$$PWS_i = \frac{hearts_i}{hearts_i + brokenhearts_i} * 100 \qquad (1)$$

…where PWS_i denotes the perceived walkability score in each PC5 area i, $hearts_i$ the number of "hearts" in each PC5 area i, and $brokenhearts_i$ the number of "broken hearts" in each PC5 area i.

(2) Data on land use (Bestand Bodemgebruik) – CBS

To examine the role of the diversity of land use types in explaining perceived walkability, data has been obtained from Statistics Netherlands (2017): The dataset "Bestand Bodemgebruik" contains spatial data at a scale of 1:10000 on the dominant land use at ground level across the Netherlands recorded in 2015 and corrected in 2017. The 34 land use types have been categorized into 9 broader land use categories (Figure 5).

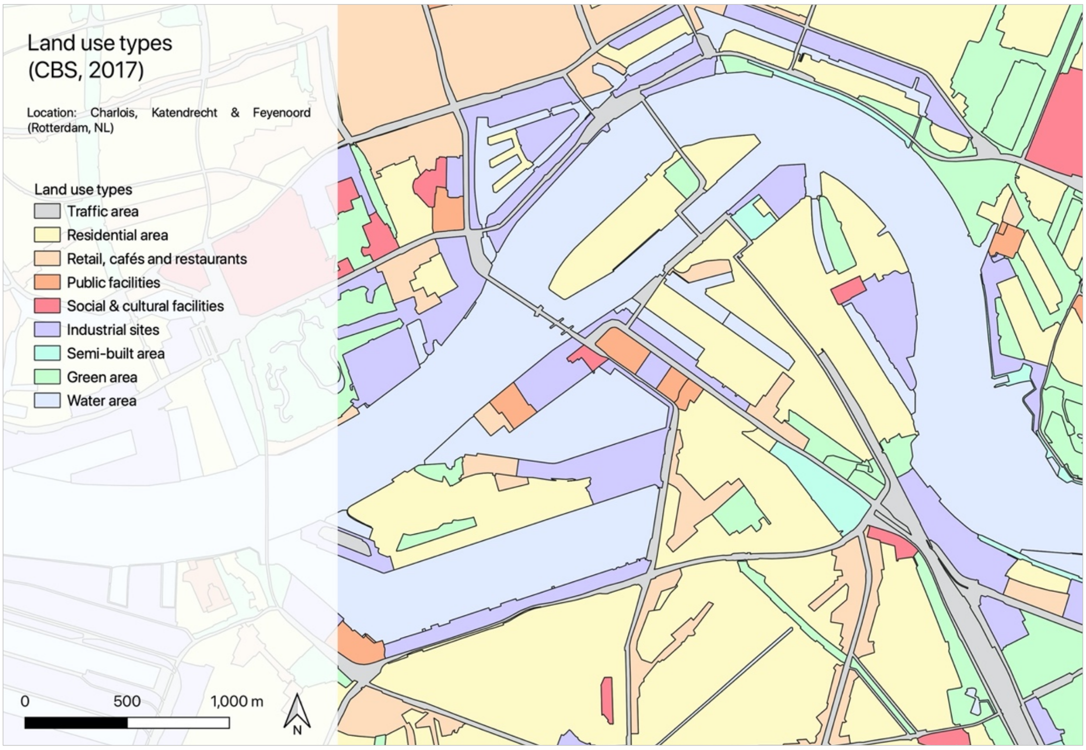

Figure 5: Map and legend with considered land use kinds (Source: CBS, 2017)

As a means to measure the diversity of land use per PC5 area, an entropy index measures the land-use mix. Song, Merlin & Rodriguez (2013) define the entropy index as formulated in Equation 2. The mathematical terms aims to quantify the extent to which the proportion of land in a defined area devoted to a specific type of land use is diversified: If a specific land use type takes on the value P_l=1 for a PC5 area, the numerator (and thus the entropy index) will take on the value 0, indicating the absence of any diversity in land use in the PC5 area. Meanwhile, if all land use types account for the same proportion of the PC5 area's surface, the entropy index will yield the value 1, indicating a maximized land-use mix.

$$entropy_{i,l,m} = -\frac{(\sum_{l=1}^{9} P_l \ln P_l)_i}{\ln m} \qquad (2)$$

…where $entropy_{i,l,m}$ denotes the entropy index of each PC5 area i with a given distribution of land use types l ∈ {1;9}, P_l denotes the percentage of each land use type l in each PC5 area and m denotes the total number of considered land use types (m=9).

(3) **Data on traffic - PDOK**

Based on shapefiles provided by PDOK, the traffic areas are further divided into different kinds of uses. This includes types of streets (local roads, regional roads, rapid roads, highways), types of rails (tram, train, metro) and sidewalks. Additional infrastructure such as sound isolation and tunnels are also taken into consideration. Furthermore, road traffic intensity measurements have been obtained by Nationaal Dataportaal Wegverkeer (NDW). The measurement used in this dataset is the average flow of vehicles between 17:00 and 18:00 on a weekday in 2019, excluding national holidays and holiday periods – a time period with presumably high walking traffic rates.

The geographical data is categorized into (1) **train rails**, (2) **other rails** (tram and metro rails) and (3) **arterial roads** (highways, rapid roads and other roads with traffic measurements with average speeds above the threshold of 40 km/h and/or more than 2200 vehicles per hour according to NDW data). Thereafter, each PC5 area's average air distance to the three respective entities is measured. This is done by first transforming the PC5 polygon into points with an equal distance of 20m to each other, upon which the average distance of all created points to the nearest respective traffic entities is calculated. By choosing average distances to the traffic entities over distances from the PC5 centroids, the calculated distances are more representative for locations in the outskirts of the PC5 polygons. Air distances (calculated using the NNJoin plugin in QGIS) have been chosen over walking distances (which could have been computed with the QNEAT3 plugin in QGIS by choosing sidewalks as a network layer) to account for presumably negative externalities, such as noise, which are rather dependent on the air distance to the noisy street than on the walking distance from the noise source.

(4) **Data on natural environment - PDOK**

Cross-sectional data on the natural environment in Rotterdam has also been obtained by PDOK. These include the geographical distribution of various water body types, green area types and trees. The provided categories of water body types have been categorized into three types: sea, inland water area (lakes, ponds, etc.) and water courses (rivers, canals, etc.). Using a similar technique to the one employed to compute the average distance to traffic entities, for each PC5 area, the average distance to the nearest water body has been computed. However, for water bodies, the QNEAT3 plugin has been chosen to compute the walking distance instead, i.e. by choosing the sidewalk layer as the network layer. As such, the distances are more representative the target population of this paper (pedestrians).

Although the PDOK data also distinguishes between different types of green areas, the categorization of green areas was rendered difficult by the high proportion of areas originally labelled as *"groenvoorziening"* (green facility) which can contain dissimilar types of greenery such as both grasslands and forests. Therefore, as opposed to the water body shapefile, the shapefile containing green areas has not been further categorized. As such, the PC5 area's average walking distances to the nearest green area has been measured, again by using the QNEAT3 plugin. Moreover, the total green area as a proportion of each PC5 area's total surface has been computed to measure the density of green areas per PC5 (Equation 3). Furthermore, to measure the density of trees per PC5 area the count per polygon function has been used.

$greendensity_i = \frac{totalgreenarea_i}{pcarea_i}$ (3)

...where $greendensity_i$ denotes the density of green areas in each PC5 area i, $greenarea_i$ the sum of all green areas in a PC5 area I, and $pcarea_i$ the area size of each PC5 area i.

(5) Data on built environment – PDOK & CBS

As for the constellation of buildings across the city, widely considered variables such as ground floor height of floor area ratio could not be found across the entire area. The paper therefore fails to take into consideration these factors. However, the building height and building area of all covered buildings across the PC5 areas have been provided. As such, after accounting for the building's areas, the average building height in each PC5 area was computed, serving as a proxy for floor area ratio. Due to higher buildings yielding a more compact urban design, the assumption is made that a higher average building height in the area is associated with a higher floor area ratio.
Next, buildings are also categorized according to their year of construction: The first in four defined periods is made up of dwellings built prior to the completion of the New Waterway *(Nieuwe Waterweg)* in 1872. The second period covers dwellings built between 1873 and the bombing of Rotterdam in 1940. The third period includes dwellings built between 1941 and the completion of Piet Blom's Cube Houses (*Kubuswoningen*) in 1984. The fourth period stretches from 1985 to 2018. Buildings built in 2019 or later were eliminated from the spatial dataset, since buildings built after the heart map campaign could not have possibly explained the distribution of the "hearts" and "broken hearts". To provide data for the empirical analysis, the percentage of buildings built in the respective time periods as a proportion of all buildings built in the PC5 areas have been computed for each PC5 area in the cross-sectional dataset. As such, four variables were created: The percentage of buildings built before 1872, the percentage of buildings built between 1873 and 1940, the percentage of buildings built between 1941 and 1984 and the percentage of buildings built between 1985 and 2018.

Moreover, a data interpolation had to be conducted to replace missing values: 4% of buildings in the dataset had no indicated height and 19% of buildings in the dataset had no indicated year of construction. Assuming that "everything is usually related to all else but those which are near to each other are more related when compared to those that are further away" (Tobler, 1970) in terms of building heights and ages, the missing values have been replaced by the average PC5 values.

Furthermore, distinguished buildings taking on highly negative values for their building heights have been dropped to alleviate the effect of any spurious outliers.

Shapefiles of various kinds of street furniture were also provided in the PDOK dataset. However, several street furniture items that have been extracted exhibited an irrealistically low frequency (e.g. irrealistically low number of benches), resulting in an omission from the dataset. Only ad columns have been kept in the dataset eventually, not to measure the association between their density and perceived walkability, but, in fact, to measure the path directness ratio. The path directness ratio (defined as the factor by which the walking distance to a chosen destination exceeds the air distance) has been obtained by measuring the PC5 areas' average walking distance to the nearest ad column (calculated with QNEAT3) and dividing it by the measured PC5 areas' average air distance (calculated with NNJoin).

(6) **Data on amenities – CBS**

Data obtained from CBS' *Wijk- en Buurtkaart* 2019 reveals average distances (i.e. also not centroid distances) to a wide range of basic goods, services, and amenities on a PC5 area basis. As represented in Figure 12, the considered destinations can be grouped into retail (marked in yellow), bars & restaurants (marked in orange), entertainment (marked in red), daycare & schooling (marked in pink), healthcare facilities (marked in purple) and public transport stops or stations (marked in blue). Thereafter, two amenity ranks have been formed capturing the proximity and density of amenities in the PC5 areas, respectively: a distance-based rank and a radius-based rank. To compute the distance-based rank, the PC5 areas have all been ranked according to their proximity to all listed amenities, after which the average of all these ranks has been computed. The same approach has been taken with radii.

Whilst computing the distance-based and radius-based amenity rank per PC5 area, values were missing for some amenity types, while for other amenity types (see entries with no cross in Figure 6), radius counts were missing altogether. As such, the radius-based amenity rank is limited by a lower range of considered amenities. For distance measurements and some radius counts, however, data was available in older editions of the *Wijk- en Buurtkaart* (2018 or 2017). Therefore, for amenities where only parts of the distance measurements or radius counts were missing, the data has been interpolated by replacing it with 2018 data, and if still not available by 2017 data. If the 2017 data was not available either, the data has been, again, manually replaced with the geographically nearest available value.

Location types	Availability at PC5 scale	
	Average distance (m)	Av. count: 1km radius (#)
Supermarket	x	x
Daily good store	x	x
Department store	x	
Café	x	x
Cafeteria	x	x
Restaurant	x	x
Library	x	
Cinema	x	
Museum	x	
Theater	x	
Childcare facility	x	x
School	x	
GP	x	x
Hospital	x	
Pharmacy	x	
Train station	x	
Transport hub	x	

Figure 6: List of location types and provided distance and density metrics (Source: CBS, 2017-2019)

(7) Data on socio-economic characteristics

To isolate the effects of the physical characteristics on perceived walkability from the effect of socio-economic variables, the study takes into account a whole series of potential effects that could stem from societal or economic characteristics. It is true that the heart map campaign measures walking preferences as opposed to walking frequency – thus, socio-economic factors that could explain individuals' walking frequencies would not necessarily have to be controlled for. However, as elaborated in Section E, 48.9% of responses (i.e. hearts and broken hearts) have been placed in the proximity of the respondents' residences. As such, the placement of hearts and broken hearts correlates with the respondents' residences, alluding to a likely concentration of respondents with certain socio-economic characteristics in certain PC5 areas. The considered socio-economic variables are listed in Figure 7.

Variables	Definition or calculation
Gender balance (%)	Men per PC area / Inhabitants per PC area
Age group: 65+ (%)	Individuals aged 65+ per PC area / inhabitants per PC area
Migration background*: None (%)	Individuals with "no migration background" per PC area / inhabitants per PC area
Households: At least one child (%)	Households with children per PC area / households per PC area
Households: Cars per household (#/hh)	Cars per PC area / households per PC area

Figure 7: List of socio-economic characteristics (Source: CBS, 2017-2019)

E. Methodology

To investigate to what extent the physical environment of a place determines the local perceived walkability in Rotterdam, the analysis of this paper explains the perceived walkability score (PWS) of a PC5 zone from the responses' physical environment by means of a hedonic utility model. The aim of the empirical analysis is to demonstrate whether the hypothesized characteristics are significantly associated with a change in the PC5 areas' perceived walkability scores. The paper is thereby guided by five sub-questions, resulting in five hedonic regressions.

a. Hedonic utility models

Hedonic utility models are a revealed preference method used to estimate the demand for a good and therewith the utility it brings to its users (Lancaster, 1966). Lancaster argues that, rather than the good itself, the individual characteristics to which the good can be broken down to, are what essentially generates utility. Under a hedonic utility model, an item's utility is therefore the aggregated utility of the respective utilities of each characteristic. Hedonic utility models are, in fact, predominantly used in the shape of hedonic pricing models, often in the context of real estate valuation (Young, 1996).

However, in the context of this perceived walkability study, the dependent variable would not be the monetary pricing of a good, but the perceived walkability score (PWS) (Equation 1 in Chapter D). Meanwhile, the independent variables are implicit attributes of the physical environment. Thereby, the physical environment is organized according to the thematic categorization in the literature review (Figure 1). The resulting consideration of physical characteristics is gathered in Figure 8.

Applying a hedonic model to a study that investigates how perceived walkability can be explained by a list of physical characteristics, comes with several advantages as well as disadvantages due to the assumptions hedonic regressions rely on. An advantage worth noting is the fact that hedonic regressions allow for the inclusion of observational data (Bateman, 1994). To refer to Figure 2, this implies that an objective (although theory-driven as variables have been selected based on citations in literature) measurement methodology can be applied. This contrasts hedonic utility models from stated preference methods, where respondents are asked about the determinants of their perceived walkability themselves. Bateman (1994) notes that stated preference methods are likely to be subject to biased questions, leading to – in the context of hedonic pricing models in real estate valuation – to a biased willingness to pay (WTP). Similarly, in the context of this perceived walkability study, stated preference methods with self-reported physical attributes can result in biases such as the anchoring of respondents to revealed walkability points.

Thematic fields	Physical characteristics	Units
Traffic environment	Distance to nearest train rails	m
	Distance to nearest other public transit rails	m
	Distance to arterial road	m
	Square metres per car	#/m^2
Natural environment	Distance to nearest seawater	m
	Distance to nearest freshwater course	m
	Distance to nearest inland water area	m
	Density of green areas	m^2/m^2
	Density of trees	#/m^2
Built environment	Average building height	m
	Percentage of buildings built until 1872	%
	Percentage of buildings built between 1873 and 1940	%
	Percentage of buildings built between 1941 and 1984	%
	Percentage of buildings built between 1985 and 2018	%
	Path directness ratio	0 to 1
	Address density	#/km^2
Amenity environment	Distance-based amenity score	0 to N=192
	Radius-based amenity score	0 to N=192
Land use	Land-use mix entropy index	0 to 1
Socio-economic control variables	Gender ratio	%
	Age group: 65+	%
	Migration background: none	%
	Households with at least one child	%
	Cars per household	#

Figure 8: List of considered physical characteristics (listed in the column "physical characteristics") with their units of measurement and thematic field according to the categorization in Figure 1.

As mentioned above, this study also has a series of disadvantages to consider resulting from the use of a hedonic utility model in studying perceived walkability. The main observed disadvantage is the exposure to omitted variable bias (OVB) (Klaiber, 2011). Although limited by the extensiveness of the dataset, OVB can persist on a neighborhood scale. Furthermore, pedestrians are presumably not as well-informed about the physical environment's wide-range bundle of characteristics. This, therefore, distinguishes them from real estate buyers who are usually better-informed about the housing's implicit characteristic attributes, allowing for the assumption that observed prices are in equilibrium (Young, 1996). In addition, occurrences of measurement errors in observed variables and multicollinearity can impede the interpretation of the regression results.

Furthermore, the regression analysis of cross-sectional data cannot fully account for a possible (residential) self-selection effect (Cao et al., 2009). Cao et al. note that self-selection occurs among others because people who prefer walking choose to live in neighborhoods that have high walkability as well. Although respondents of the "Do you want to walk with me campaign" did not per se reside in the PC5 area where they placed a heart and/or broken heart, the dataset indicates that 48.9% of responses (i.e. hearts and broken hearts) have been placed in the proximity of the respondents' residence. As such, the placement of hearts and broken hearts correlates with the respondents' residences, alluding to a likely concentration of respondents with certain socio-economic characteristics in certain PC5 areas. Therefore, both models hereunder control for a list of socio-economic variables, which are listed in Figure 8.

Drawing on the review of hedonic utility modelling, the base model investigating the association between the PC5 area's PWC and its underlying physical characteristics is represented in Equation 5.

$$PWS_i = \alpha_o + \beta_1 T_i + \beta_2 N_i + \beta_3 B_i + \beta_4 a_i + \beta_5 entropy_i + \beta_6 S_i + \varepsilon_i \quad (5)$$

...where PWS_i denotes the perceived walkability score in each PC5 area i, $\beta_1 T_i$ denotes the vector of coefficients of traffic environment characteristics, $\beta_2 G_i$ denotes the vector of coefficients of natural environment characteristics, $\beta_3 B_i$ denotes the vector of coefficients of built environment characteristics, $\beta_4 a_i$ denotes either of the amenity scores (both are tested for their fit), $\beta_5 entropy_i$ denotes the coefficient of the land-use mix and $\beta_6 S_i$ denotes the vector of coefficients of socio-economic control variables. α_o is the constant and ε_i stands for the standard errors.

Besides the base model, a thematic index model is constructed. This model does not investigate the association between the separate features of the physical environment and the perceived walkability. Instead, it merges the separate features of the physical environment into thematic indexes, indexed on the values for the PC5 area with the most available values (excl. interpolated values) for the candidate predictors. The indexes are therefore indexed on the values of the PC5 area 3076J in Rotterdam-Lombardijen, which has been found to exhibit the least amount of missing values. Characteristics of the mentioned PC5 area have been studied in detail, including its land-use mix, traffic environment, natural environment, built environment and accessibility of amenities, to infer any possible biases in the interpretation of the index-based model. For a map representing the 3076J's physical environment, see Figure A3 in Appendix.

The PC5 area perceived walkability index is essentially sought to be predicted by the traffic environment (embodied in the traffic environment score "tindex"), the natural environment (embodied in the water environment score "windex" and green environment score "gindex") and the built environment (embodied in the compactness score "cindex"). The thematic scores are thereby computed as the average of the indexed values of all components in the respective thematic field. The components equal the variables listed in Figure 8. Further predictors include one of the amenity scores indexed on 3076J = 100 and the land-use mix entropy index, again indexed on 3076J = 100. As for the amenity score, both the distance-based and radius-based scores have been tested for the resulting model's goodness-of-fit.

Once constructed, the indexed values are then juxtaposed to the perceived walkability score. In Equation 6, the PC5 area perceived walkability scores are therefore aimed to be explained by the four abovementioned thematic indexes, the indexed value for the amenity score ("amindex") and the indexed value for the land-use mix entropy index ("lindex").

$$PWS_i = \alpha_o + \beta_1 tindex_i + \beta_2 windex_i + \beta_3 gindex_i + \beta_4 cindex_i + \beta_5\, amindex + \beta_5\, lindex + \beta_6 S_i + \varepsilon_i \qquad (6)$$

...where PWS_i denotes the perceived walkability score in each PC5 area i, $tindex_i$ denotes the coefficient of the traffic environment score, $windex_i$ denotes the coefficient of the water environment score, $gindex_i$ denotes the coefficient of green environment score, $amindex_i$ denotes the coefficient of indexed amenity score, $lindex_i$ denotes the coefficient of the indexed land-use mix and $\beta_6 S_i$ denotes the vector of coefficients of the considered socio-economic control variables. α_o is the constant and ε_i stands for the standard errors.

b. Robustness checks

Finally, the models were examined for any exposure to violations of the assumptions of a multiple linear regression. Several robustness tests were conducted: Variance inflation factors (VIF) tested both models for any presence of multicollinearity (see results of final models in Appendix, Figures A4-A7). Model constellations which were tested prior to the formulation of the final models exhibited a range of correlation levels above the threshold of 0.8. Highly correlated with address density, *tree density, path directness ratio* and *m^2 per car* have been eliminated from the cross-sectional dataset, as a way to drive down the mean VIF and stabilize the parameter estimates.

Skewness and kurtosis statistics investigated to which extent the models exhibit a normal distribution of residuals. At a 5% significance level, the hypothesis of normally distributed error terms could not be rejected. Testing the assumption of constant variances of residuals, a Breusch-Pagan/Cook-Weisberg test indicates no significant presence of heteroskedasticity in either of the models (see results in Appendix, Figures A8-A9). Lastly, to observe patterns of any spatial error autocorrelation, the error term of both models have been visually inspected by means of a map representation, with no apparent spatial pattern visible (see map in Appendix, Figure A10-A11).

F. Descriptive statistics

First, looking at the statistical distribution of the perceived walkability scores (PWS), skews towards the left and the right extreme of the histogram can be observed (see appendix, Figure A12). This can be drawn back to the high density of PWSs taking on the values 0 and 100 in the dataset. Such a distribution could be interpreted as the result of a polarized view of the landscape in terms of its walkability, with many PC5 areas being widely viewed as very unwalkable (left bar) and other PC5 areas as very walkable (right bar). When overlaying the PC5 perceived walkability scores (PWSs) with selected physical characteristics, clearer patterns can be observed.

As such, in various locations across the municipality, visualizations suggest a negative relationship between the proximity of arterial roads as well as various types of rails. As seen in the example of Hillegersberg-Zuid/Liskwartier (Figure 9), a majority of PC5 areas adjacent to arterial roads and/or rails exhibit a red colour (denoting a PWS below 20 out of 100). The negative perceived walkability engendered by arterial roads (especially highways) and train rails seems to capture a wider adjacent area than the negative perceived walkability associated with other rail traffic such as trams or overground metro segments. Moreover, on a wider scale, the map demonstrates that many arterial roads and rails run in parallel to each other in tight traffic corridors. Such corridors visually seem to have a particular unfavorable reverberation in its adjacent PC5 areas when looking at the PWSs. On the other hand, the widespread adjacency of rail traffic and arterial roads across the municipality will make the interpretation, regarding which of these two traffic infrastructure bodies influences perceived walkability, somewhat harder. Finally, it must be noted that at various locations, the observations do not apply, with streets such as the Westersingel exhibiting an above-average PWS in spite of its contiguity to tram rails.

Comparing the distribution of the PC5 PWSs with natural environment characteristics also suggests reoccurring patterns across space. A rather apparent observation is the widespread presence of blue and green lungs across the city, often exhibiting high perceived walkability scores in adjacent PC5 areas, or at least a better perceived walkability score than other PC5 areas in the wider neighborhood (Figures 10 & 11). Looking at how PWSs vary according to the type of water body they border, no particular patterns could be visually identified. Finally, similarly to the observation made regarding motorized and rail traffic lanes, the blue and green lungs often run side by side to each other. Examples include PC5 areas found in designated park areas such as Kralingse Bos or Het Park, which have a high density of green areas as well as low average distances to water bodies like inland water areas (lakes, ponds, etc). The same spatial correlation is also suggested when looking at canals such as Heemraadsingel, where the water course has a wide layer of green bordering both shores. Therefore, when interpreting the results of the statistical output, one must again bear in mind the potentially spatially correlated character of natural environment characteristics.

As for the spatial relationship between the built environment and the PC5 areas' perceived walkability, the picture is not as clear. While relationships between the average building height and perceived walkability varies vastly across and within PC5 areas, a clearer relationship can be hinted at when comparing PWSs with the average years of construction. At various locations, similar patterns could be observed as the one in Noordereiland, mapped in Figure 12: PC5 areas containing or contiguous to buildings constructed prior to 1940 exhibit higher perceived walkability scores than those where buildings built after 1940 dominate the landscape.

Finally, the representation of the geographical data did not hint at any apparent spatial relationship between perceived walkability and the two indices used in the base model (amenity index and land-use mix entropy index). However, looking at the distribution of these indices separately, a spatial relationship manifests itself on maps, whereby the indices correlate with the PC5's density of addresses, which can be seen as an indicator for the areas' degree of urbanity. In fact, both indices have their highest values in PC5 areas in the city centre, with the highest amenity score (1016.71) and highest land use mix entropy index (0.22) both located within 200m from the city hall (Stadhuis) Therefore, the possibility of a multicollinear relationship between the indices and other physical characteristics has to be taken into consideration while interpreting the results of the regression. For other statistical summary results, see Figures A13 (base model) and A14 (index model) in Appendix.

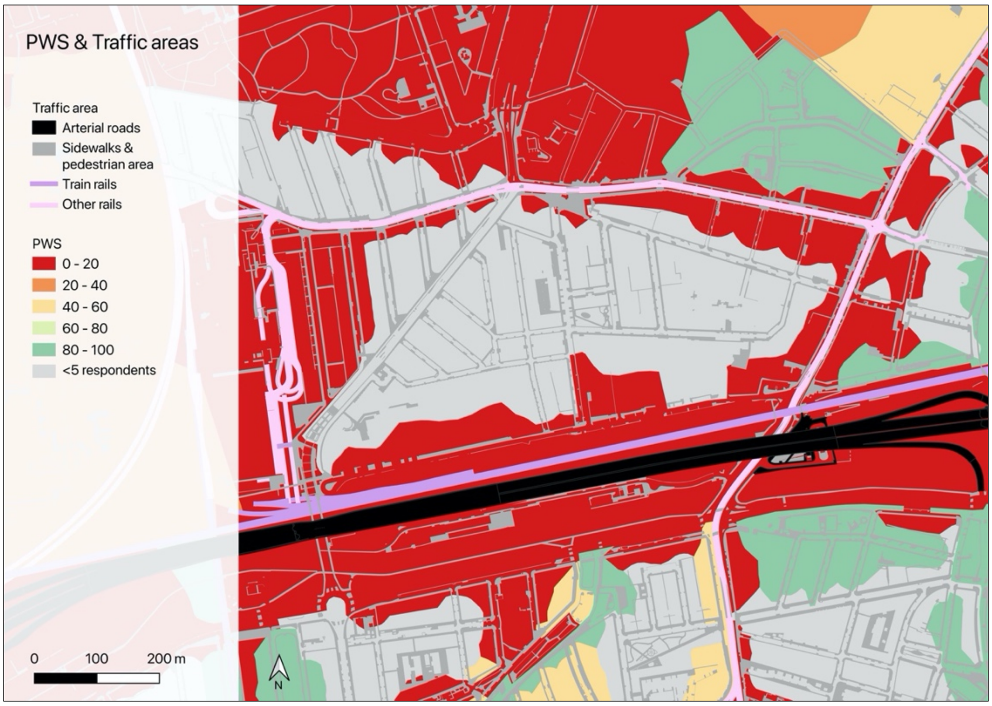

Figure 9: Perceived walkability scores (PWS) across the municipality of Rotterdam (PC5 level), overlayed with traffic environment characteristics

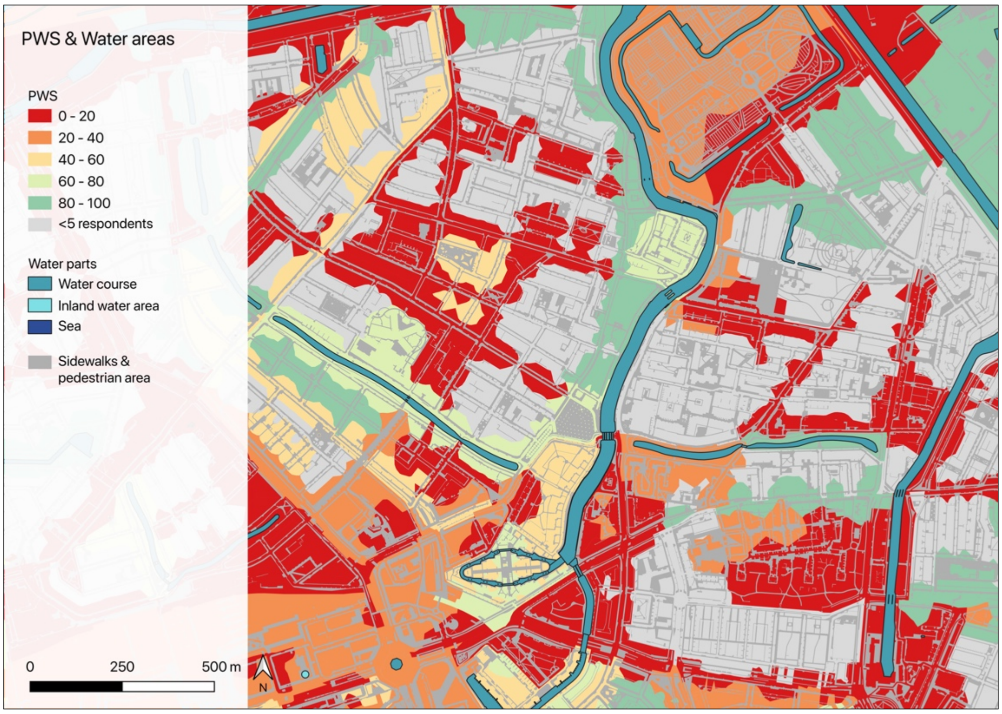

Figure 10: Perceived walkability scores (PWS) across the municipality of Rotterdam (PC5 level), overlayed with water elements

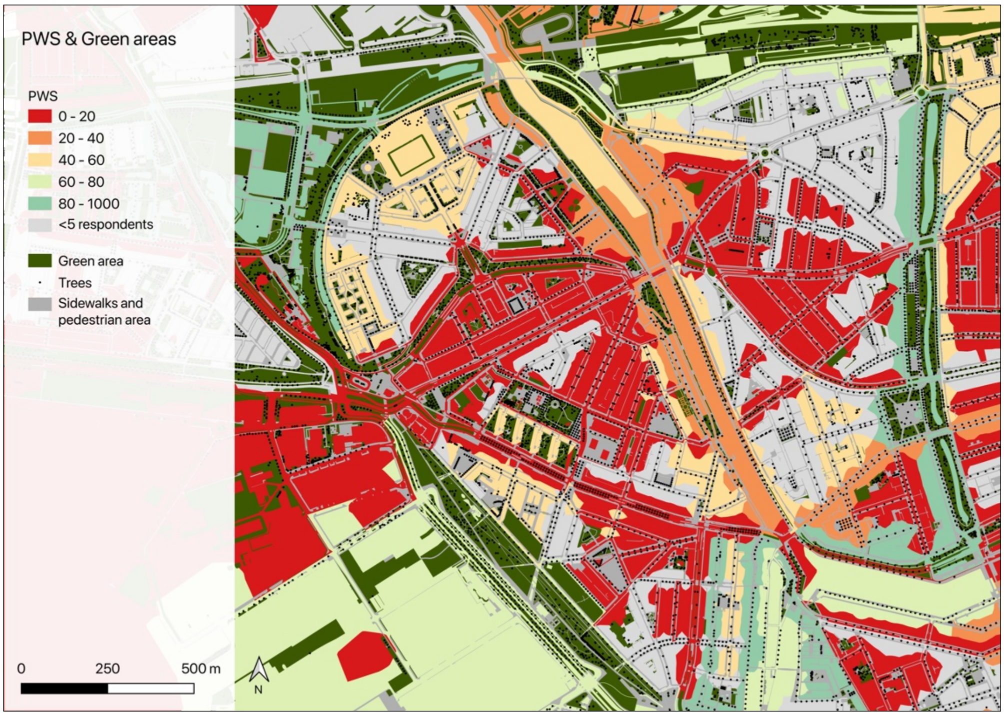

Figure 11: Perceived walkability scores (PWS) across the municipality of Rotterdam (PC5 level), overlayed with green elements

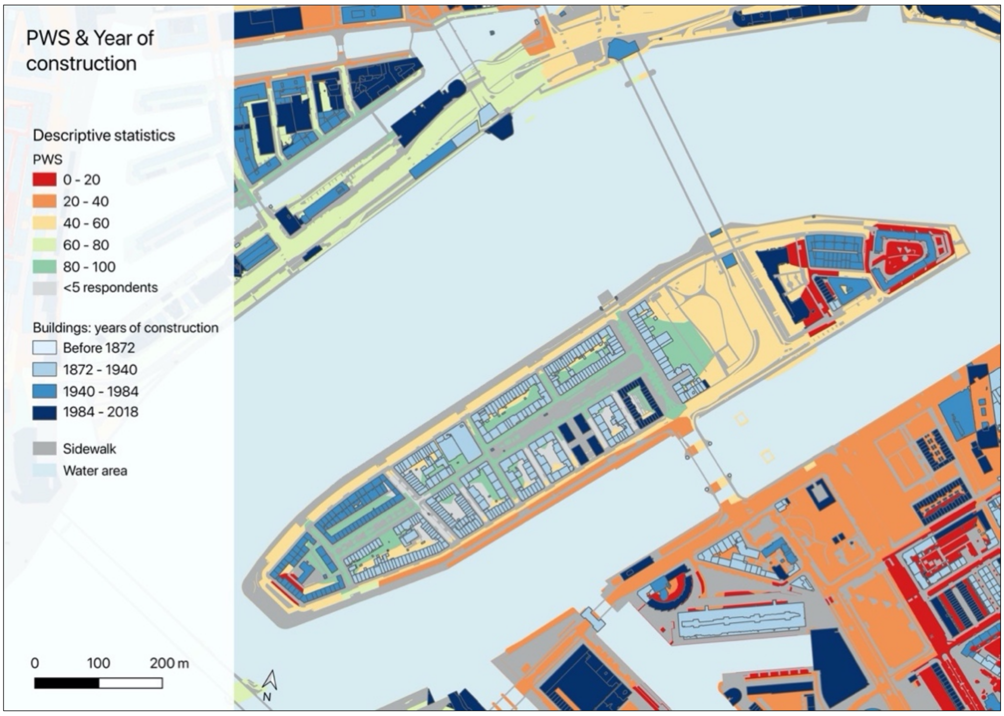

Figure 12: Perceived walkability scores (PWS) across the municipality of Rotterdam (PC5 level), overlayed with built (categorized according to considered periods of construction)

G. Empirical results

In the following chapter, the results of the empirical analysis will be presented. The methodology of the two employed models can be found back in Chapter E (Methodology).

(1) Association with traffic environment

At a 5% significance level, the base model exhibits a significantly positive relationship (P>|t| = 0.009) between the PC5 areas' perceived walkability scores (PWSs) and the PC5 area's average distance to the nearest arterial road (Figure 13). This means that the further the PC5 area is located (on average) from any arterial road, the higher its predicted PWS. The coefficient indicates that, if the distance to the nearest arterial road increases by 100m, the perceived walkability score (measured on a scale from 0 to 100 points) scores 1.3 points higher. Thereby, as elaborated in Chapter D, entities defined as arterial roads can be different kinds of roads: Arterial roads comprise all highways, rapid roads and other roads that have an average traffic speed of at least 40 km/h and/or more than 2200 vehicles per hour. As such, the interpretation does not give an opportunity to infer the association with the separate road kinds. Still, the result would support one of the priorities of the Gemeente Rotterdam (2020), which strives to boost walking rates by implementing route networks which seek to avoid arterial roads.

In contrast, the distance to rail traffic infrastructure seems to play a smaller, statistically insignificant, role in determining the pedestrians' perceived walkability. Having used the forward selection while formulating the base model, different constellations of the traffic entities have been tried out, whereby tram rails have also been considered as a separate layer to examine the association between their proximity and their adjacent PC5's perceived walkability scores: While the signs of the coefficients were consistently negative (i.e. proximity to rails is positively correlated with perceived walkability), coefficients for either of the examined rail types remained insignificant.

Moving on to the index-based model, the interpretation of the association between the PC5 areas' perceived walkability and the distribution of the traffic environment is done by interpreting the coefficient of the traffic environment score (tindex). Tindex computes the average of the indexed values (3076J = 100) of the score components, which are the average PC5 distance to the nearest arterial roads, the nearest train rail and the nearest other kind of rail. Significant at an alpha of 5%, the composure of the "tindex" could be implemented in a final walkability index. However, the separate effects of the respective components remain uncaptured in the index-based model and with the inclusion of variables with various units of measurement, the magnitude of the coefficient cannot be assessed appropriately.

(2) Association with natural environment

Looking at the association between the implicit attributes of the natural environment and the PC5 areas' perceived walkability scores, a clearer picture can be drawn for green attributes than blue attributes. At a 5% significance level, the base model suggests a significantly positive association between the PC5 areas' density of green space and the local perceived walkability scores (Figure 13): Taking on a coefficient of 25.212, the regression results imply that for every (0.1*m2)/m2 of additional green space, the perceived walkability score (0 to 100) increases by 2.52. In sports terms, the result could also be interpreted as follows. Meanwhile, the significance of the PC5 areas' average distance to water bodies varies according to the kind of water body: While distances to the sea and distances to nearest lake/pond are found insignificant, the distance to the nearest water course (i.e. river, canal, etc.) is found to have a significantly negative relationship with perceived walkability. In other words, the proximity of a water course has a significantly positive association with the PWS: For every 100m approached to a water course, the perceived walkability (measured on a scale from 0 to 100) scores 7.3 higher. The results are thus only partly in line with past research: While the positive relationship with green space corresponds to Wagtendonk & Lakerveld (2019)'s inclusion of the green area density into their Netherlands-based walkability index, findings by Liao et al. (2020) (who find a positive relationship between inland water and perceived walkability) and Bauman et al. (1999) (who find a positive relationship between coast proximity and perceived walkability) could not be confirmed by this analysis. It has to be noted, however, that this research is focusing on the relationship between the natural environment and walking frequencies, rather than walking preferences, as is the case in this paper.

Looking at the index-based model, where the role of the natural environment is sought to be captured by means of two indices (windex for water environment characteristics and gindex for green environment characteristics), both indexes are associated with a significant variance in the PC5 area PWSs: While an increase in the green environment index is associated with a significantly positive variance in perceived walkability, increases in the water environment index exhibit a negative relationship. Since gindex is composed of (indexed) density measures, an increase in the index can be interpreted as an increase in green feature density. Meanwhile, windex features (indexed) distance measures. Thus, an decrease in windex is what denotes an increase in the proximity of water entities. Thus, the coefficient of the indices suggest that a richer green and water environment are associated with an increase in perceived walkability. This finding therefore, for example, supports directives of Rotterdam's green-blue network, which seeks to stimulate walking by promoting routes along such natural features.

(3) Association with built environment

As for the built environment, the base model finds none of the considered attributes significant in explaining the variation in perceived walkability across Rotterdam. With P>|t| values far above 0.05, neither of the results suggest a positive association between a compact city structure and perceived walkability. As a result, findings by Frumkin et al. (2004), who define walkability as the "density of people, households or jobs distributed over a unit of area", could not be confirmed by analysing the spatial relationship between address density and perceived walkability. However, again, it has to be noted that this research looks into walking preferences (i.e. what makes people *enjoy* walking) rather than walking rates (i.e. what makes people walk). As for the association with designated periods of construction, neither of the periods differ significantly from buildings constructed after 1984, with buildings built between 1872 and 1940 even exhibiting a negative coefficient sign, and thus unaligned with previous literature suggesting higher walkabilities in older cities.

(4) Association with amenity environment

As elaborated in Chapter D (Data), two amenity scores have been conceptualized to measure the accessibility of a wide range of amenities. Using the distance-based amenity average rank resulted in a lower amount of missing variables. Therefore, it has been selected as an independent variable both in the base model and the index model (indexed on the amenity rank of 3076J=100). The coefficient suggests a positive correlation between the amenity index and the perceived walkability score. It has to be noted that the amenity index denotes an average rank of distances to amenities. Thus, a higher amenity index denotes a worse ranking (since a better ranking would be a lower number). Therefore, the sign of the coefficient suggests that the PC5-average accessibility of amenities is positively correlated with the perceived walkability scores. However, with an alpha above 0.05, no significant effects can be inferred. A similar conclusion can be drawn for the coefficient of the indexed distance-based amenity rank employed in the index-based hedonic regression.

(5) Association with land-use mix

As a measure to investigate the presence of any effect of a higher diversity in land use on the local perceived walkabilities, the base model looks into the statistical association between the land-use mix entropy index and the perceived walkability. Results indicate no significant statistical relationship. Thereby, the issue of endogeneity has to be considered: Although not interpretable in the correlation matrixes and variance inflation factor computations, geographical representations suggest a spatial correlation between land-use mix and other variables related to the degree of urbanity. Again, a similar conclusion can be drawn for the index-based model. Although a widely cited predictor of walkability, no significant statistical relationship has been found with perceived walkability. This can be drawn to the fact that past research has looked into what *makes* people walk in specific places, rather than what makes them enjoy walking at these places. As such walking rates might be high, in areas exhibiting high land-use mixes because a lot of destinations can be found in the area and are thus walked to, whereas they are not exactly located at walkable spots.

Obs.	F(17, 174)	Prob > F	R^2	Adj. R^2
192	5.28	0.0000	0.3402	0.2758

PWS	Coefficient	Std. err.	t	P>|t|	– 95% CI	–
Distance to train rails (m)	0.006	0.004	1.44	0.152	-0.002	0.015
Distance to arterial road (m)	0.013	0.005	2.66	0.009	0.003	0.022
Distance to other rails (m)	0.000	0.003	-0.05	0.959	-0.006	0.006
Distance to sea (m)	0.000	0.001	0.7	0.484	-0.001	0.002
Distance to watercourse (m)	-0.073	0.016	-4.58	0	-0.104	-0.041
Distance to lake (m)	-0.003	0.004	-0.77	0.445	-0.012	0.005
Density of green (m^2/m^2)	25.212	6.224	4.05	0	12.928	37.496
Average building height (m)	39.980	62.105	0.64	0.521	-82.597	162.556
-1872 (%)	1.642	10.208	0.16	0.872	-18.506	21.790
1872-1940 (%)	-9.797	10.368	-0.94	0.346	-30.259	10.665
1940-1984 (%)	0.334	0.344	0.97	0.332	-0.344	1.013
Address density (#/km2)	0.000	0.003	-0.1	0.921	-0.005	0.005
Rank-based amenity index (0 to 1)	-0.004	0.017	-0.21	0.831	-0.037	0.030
Entropy index (0 to 1)	0.001	64.155	-1.19	0.237	-202.721	50.522
Men (%)	50.737	36.361	1.4	0.165	-21.029	122.502
No migrant background (%)	20.462	15.389	1.33	0.185	-9.910	50.834
Households with 1+ children (%)	18.173	19.529	0.93	0.353	-20.371	56.716
Constant	-9.934	34.840	-0.29	0.776	-78.698	58.830

Figure 13: Results of the multiple linear regression of physical environment variables on PC5 perceived walkability score (base model).

Obs.	F(9,182)	Prob > F	R^2	Adj. R^2
192	7.62	0	0.2788	0.2431

PWS	Coefficient	Std. err.	t	P>|t|	– 95% CI	–
Traffic environment index (tindex)	0.765	0.212	3.6	0	0.346	1.184
Water environment index (windex)	-0.384	0.129	-2.98	0.003	-0.638	-0.130
Green environment index (gindex)	0.227	0.074	3.07	0.002	0.081	0.372
Compactness environment index (cindex)	0.033	0.018	1.85	0.065	-0.002	0.068
Amenity environment index (amindex)	-0.001	0.002	-0.34	0.736	-0.004	0.003
Land-use meta-index (lindex)	-0.047	0.089	-0.53	0.593	-0.223	0.128
Men (%)	47.893	35.730	1.34	0.182	-22.605	118.391
No migrant background (%)	23.899	13.539	1.77	0.079	-2.814	50.612
Households with 1+ children (%)	24.759	19.241	1.29	0.2	-13.206	62.724
Constant	-35.442	37.194	-0.95	0.342	-108.830	37.945

Figure 14: Results of the multiple linear regression of physical environment indexes on PC5 perceived walkability scores (index model)

H. Discussion & limitations

The methodology applied in this paper does not come without its limitations. Since hedonic regression models do not allow for an identification of causality, an assumption of exogeneity was made. In other words, an absence of endogeneity is assumed. Meanwhile, in this paper, endogeneity potentially occurred in the form of reverse causality. An potential case of reverse causality can be retraced to the dynamics between walkability and real estate: As stated in the literature review, Tinessa et al. (2020) have found that walkable neighborhoods attract higher rates of retailers, whereas (by capturing the proximity of retail) in the amenity rank) this paper seeks to explain perceived walkability as a result of the high density of retail.

Considering the wide range of attributes that have been associated with variations in walkability in past literature, many physical attributes of the environment have not been captured in the cross-sectional dataset, exposing the models to OVB. For example, variables which are particularly valued by specific populations segments have been omitted: These include the presence of white canes and traffic lights equipped with noise signals (relevant for visually impaired pedestrians), as well as the width of sidewalks and the quality of the sidewalk tiles (relevant for wheelchair users).

Moreover, as hinted by the low R^2 in both models, merely a low proportion of the variation in the dependent variable (perceived walkability) is predicted by the variables seeking to define the physical environment. The low explained variation can of course be drawn back to the limitations in the used methodology. However, regressions seeking to predict behavior generally exhibit low explain variations, often to a large extent due to the heterogeneity of the studied populations (Fennell et al., 2019). To better understand the drivers of walking and the hedonic motives behind such physical activity, the study would be better structured if it examined micro-level urban design features which can enhance walking preferences. Empirical approaches as the one used in this paper oversimplify the motives behind perceived walkability by omitting context sensitivities, which can have an impact, among others, on the affective experience gained from walking (Johansson, Sternudd & Kärrholm, 2016).

The limited representativeness of the perceived walkability data also has to be taken into consideration: While the team behind the crowdsourced map, which this paper is based on, managed to gather a fair number of responses on which places respondents perceived as walkable or unwalkable, the location of a large share of “hearts” (walkable places) and “broken hearts” (unwalkable places) might be imprecise. This is partly due to the fact that respondents could pin the chosen spots while being elsewhere, e.g. from home. Although more costly, a better approach to gather the localized responses would be if respondents were at the spot themselves and shared their GPS location while answering the survey. Furthermore, the dataset did not include specifications on the socio-demographic characteristics of the respondents associated with specific hearts and broken hearts, making it difficult to truly isolate the effect of the physical environment from the impact of socio-demographic characteristics on walking preferences. Instead, only the overall statistics of the respondent panel has been provided: With a quasi-absence of minors in the sample and overrepresentation of above-average educated people, for example, the responses might not be as representative for specific segments of society in Rotterdam. As such, any upcoming survey-based walkability research should strive to consider a wider inclusion of underrepresented population segments in the collection of responses.

I. Conclusion

Throughout the paper, the role of a wide range of physical characteristics in perceived walkability has been considered. Considered variables included the distance or density of traffic components, natural features, built environment characteristics, amenities, and the diversity of land use. Two hedonic utility models have been constructed, aiming to estimate the effect of the physical environment areas on the perceived walkability across PC5 areas in Rotterdam. While perceived walkability was captured by means of a PC5-based "perceived walkability score" (PWS), spatial data on the physical characteristics have been transformed by using GIS-based geoprocessing and analysis tools. Both models were destined at answering the five hypotheses and tackle the following central research question:

RQ: "How does the physical environment explain perceived walkability?"

The first hypothesis has been formulated as follows: *"A higher proximity to non-pedestrian traffic infrastructure has a negative effect on perceived walkability".* Upon transforming the data to obtain average PC5 distances to arterial roads, train rails and other rails, a base model found merely the effect of arterial roads significant. Meanwhile, when incorporating these distance metrics (including train rails and other rails) into the index-based hedonic model, an overall significant negative relationship between proximity of important non-pedestrian infrastructure and perceived walkability is identified. Therefore, while motorized traffic has been found to negatively impact perceived walkability in both models, the former model does not identify any significant relationship between rail traffic and perceived walkability. The hypothesis is therefore partly rejected.

The second hypothesis was: *"A higher presence of natural features has a positive effect on perceived walkability."* Having investigated the effect of both green spaces and water bodies, the effects of green spaces show clearer trends than, at least some kinds of, water bodies. The base model identifies a positive effect of the density of green space on perceived walkability. The proximity to freshwater courses has also been found to have a positive effect. However, proximities to other water bodies, such as lakes or seawater, have not been found significant in explaining variations in perceived walkability. However, the index-based model shows a significantly positive effect of rich natural environments on perceived walkability. Still, with insufficient revealed perception of walkability in the proximity of seawater, the water environment index is possibly subject to a bias. As a result, the hypothesis is partly rejected.

The third hypothesis was: *"A higher density of built structures has a positive effect on perceived walkability".* After applying GIS-based landscape metrics to describe the average building height per PC5 area and the distribution of four architectural periods across Rotterdam's PC5 areas, unexpected correlations arise such as a negative correlation (yet insignificant statistical relationship) between buildings built in 1872 and 1940 and perceived walkabilities. Further looking into the association with the compactness of the built environment, no statistical significance has been identified in these relationships. The index-based model led to a similar conclusion.

The fourth hypothesis was: *"A higher presence of practical amenities has a positive effect on perceived walkability".* To measure the effect of the accessibility of amenities on perceived walkability, a rank-based index has been, denoting the average rank of each PC5 area in terms of accessibility to amenities. No statistically significant association has been found between the PC5 perceived walkability scores (PWS) and the accessibility of amenities.

The fifth hypothesis was: *"A higher land-use mix has a positive effect on perceived walkability".* Although a widely cited predictor of walkability, no significant statistical relationship has been found with perceived walkability. The lack of any relationship could be drawn to the fact that the papers research walking preferences rather than walking rates. Areas with high land-use mix rates have a compact built structure and can thus capture a higher number of walking willingness radii of people – however, this does not per se infer a more enjoyable walking experience.

This brings us to a final answer to the central research question ("How does the physical environment explain perceived walkability?"). While walkability studies widely point at the walkable character of compact city structures, this study reveals that perceived walkability, i.e. the extent to which people *enjoy* walking, is predominantly a result of the rich natural environment and the absence or lower presence of arterial roads for motorized traffic. Meanwhile, variables measuring the compactness of a neighborhood (such as the density of architecture, the accessibility of amenities and the diversity of land use) do not exhibit any significant effect.

Bibliography

Adams, W. (2006). The Future of Sustainability: Re-thinking Environment and Development in the Twenty-First Century. *Report of the International Union for Conservation of Nature Renowned Thinkers Meeting*. Zurich.

Banai, R. (1998). The New Urbanism: An Assessment of the Core Commercial Areas, with Perspectives from (Retail) Location and Land-Use Theories, and the Conventional Wisdom. *Environment and Planning B: Urban Analytics and City Science 3.619 Impact Factor 5-Year Impact Factor 3.889 Journal Indexing & Metrics »* , 169-85.

Bateman, I. (2007). Contingent valuation and hedonic pricing: problems and possibilities. *Landscape Research*, 30-32.

Bauman, A., Smith, B., Stoker, L., Bellew, B., & Booth, M. (2008). Geographical influences upon physical activity participation: evidence of a 'coastal effect'. *Australian and New Zealand Journal of Public Health*, 322-324.

Blecic, I., Canu, D., Cecchini, A., Congiu, T., & Fancello, G. (2016). Factors of Perceived Walkability: A Pilot Empirical Study. *Computational Science and Its Applications ICCSA*, (pp. 125-137). Beijing.

Brownson, R., Hoehner, C., Day, K., Forsyth, A., & Sallis, J. (2009). Measuring the Built Environment for Physical Activity: State of the Science. *American Journal of Preventive Medicine*, 99-123.

Canepa, B. (2007). Bursting the Bubble: Determining the Transit-Oriented Development's Walkable Limits. *Transportation Research Record: Journal of the Transportation Research Board*, 28-34.

Cao, X. J., Mokhtarian, P., & Handy, S. (2009). Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings. *Transport Reviews*, 359-395.

CBS. (2018). *Onderzoek Verplaatsingen in Nederland (OViN)*. From CBS: https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/onderzoek-verplaatsingen-in-nederland--ovin--

CBS. (2019, June 3). *Huishoudens naar inkomen en autobezit per PC5-gebied*. From CBS: https://www.cbs.nl/nl-nl/maatwerk/2019/23/huishoudens-naar-inkomen-en-autobezit-per-pc5-gebied

CBS. (2020, January 1). *Toelichting Wijk- en Buurtkaart 2018, 2019 en 2020*. From CBS: https://www.cbs.nl/nl-nl/longread/aanvullende-statistische-diensten/2020/toelichting-wijk-en-buurtkaart-2018-2019-en-2020

Cerin, E., Saelens, B., Sallis, J., & Frank, L. (2006). Neighborhood Environment Walkability Scale: Validity and Development of a Short Form. *Medicine & Science in Sports & Exercise*, 1682-1691.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 199-219.

College van Rijksadviseurs. (2020). *Naar een gezonde stad te voet.* Den Haag: Rijksoverheid.

De Haas, M., Hamersma, M., & Faber, R. (2020). *Mobility and the coronavirus crisis.* Den Haag: Netherlands Institute for Transport Policy Analysis.

Decisio. (2020). *Waarde van actieve mobiliteit voor Rotterdam.* Rotterdam: Gemeente Rotterdam.

European Environmental Agency. (2016). *Urban Sprawl in Europe: Joint EEA-FOEN Report.* Luxembourg: EEA.

Ewing, R., & Cervero, R. (2001). Travel and the Built Environment: A Synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, 87-114.

Ewing, R., & Handy, S. (2009). Measuring the Unmeasurable: Urban Design Qualities Related to Walkability. *Journal of Urban Design*, 65-84.

Foraster, M., Deltelle, A., Basagaña, X., Medina-Ramón, M., Inmaculada, A., Bouso, L., . . . Künzli, N. (2011). Local determinants of road traffic noise levels versus determinants of air pollution levels in a Mediterranean city. *Environmental Research*, 177-183.

Frank, L., Sallis, J., Conway, T., Chapman, J., Saelens, B., & Bachman, W. (2006). Many Pathways from Land Use to Health: Associations between Neighborhood Walkability and Active Transportation, Body Mass Index, and Air Quality. *Journal of the American Planning Association*, 75-87.

Frank, L., Sallis, J., Saelens, B., Leary, L., & Cain, K. (2009). The development of a walkability index: application to the Neighborhood Quality of Life Study. *British Journal of Sports Medicine*, 13-44.

Frumkin, H., Frank, L., & Jackson, R. (2004). *Urban Sprawl and Public Health: Designing, Planning, and Building for Healthy Communities.* Washington DC: Island Press.

Gemeente Rotterdam. (2019). *Rotterdam Veilig Vooruit Proactieve en risicogestuurde aanpak voor verkeersveiligheid.* Rotterdam: Gemeente Rotterdam.

Gemeente Rotterdam. (2020). *Lopen in Rotterdam: Meting via Digitaal Stadspanel Rotterdam.* Rotterdam: Gemeente Rotterdam.

Gemeente Rotterdam. (2020). *Rotterdam Walks 2025: The pedestrian on a pedestal.* Rotterdam: City of Rotterdam.

Glazier, R., Weyman, J., & Creatore, M. (n.d.). Development and validation of an urban walkability index for Toronto. *Toronto Community Heal.*

Grasser, G., Van Dyck, D., Titze, S., & Stronegger, W. (2013). Objectively measured walkability and active transport and weight-related outcomes in adults: a systematic review. *International Journal of Public Health volume*, 615-625.

Grasser, G., Van Dyck, D., Titze, S., & Stronegger, W. (2016). A European perspective on GIS-based walkability and active modes of transport. *The European Journal of Public Health*, 145-151.

Habibian, M., & Hosseinzadeh, A. (2018). Walkability index across trip purposes. *Sustainable Cities and Society*, 216-225.

Health Effects Institute. (2021). *Annual Report 2021: Trusted Science for a Changing World.* Boston.

Humpel, N., Owen, N., & Leslie, E. (2002). Environmental factors associated with adults' participation in physical activity: A review. *American Journal of Preventive Medicine*, 188-199.

Johansson, M., Sternudd, C., & Kärrholm, M. (2015). Perceived urban design qualities and affective experiences of walking. *Journal of Urban Design*, 256-275.

Jones, L. (2010). *Investigating neighborhood walkability and its association with physical activity levels and body composition of a sample of Maryland adolescent girls.* University of Maryland, College Park.

Kleiber, A., & Abbott, J. (2011). Contingent valuation and hedonic pricing: problems and possibilities. *The Review of Economics and Statistics* , 1331-1342.

Kwon, M., Lee, C., & Xiao, Y. (2017). Exploring the role of neighborhood walkability on community currency activities: A case study of the crooked river alliance of TimeBanks. *Landscape and Urban Planning*, 302-314.

Lancaster, K. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 132-157.

Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., & Hugo, G. (2007). Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. *Health Place*, 111-122.

Liao, B., Van Den Berg, P., Van Wesemael, P., & Arentze, T. (2020). Empirical analysis of walkability using data from the Netherlands. *Transportation Research Part D*.

Litman, T. A. (2003). Economic Value of Walkability. *Transportation Research Record: Journal of the Transportation Research Board*, 3-11.

Lizárraga, C., Martín-Blanco, C., Castillo-Perez, I., & Chica-Olmo, J. (2022). Do University Students' Security Perceptions Influence Their Walking Preferences and Their Walking Activity? A Case Study of Granada (Spain). *Sustainability*, 14.

Lo, A., & Houston, D. (2018). How do compact, accessible, and walkable communities promote gender equality in spatial behavior? *Journal of Transport Geography*, 42-54.

Mackett, R., & Brown, B. (2011). Transport, Physical Activity and Health: Present knowledge and the way ahead. *Centre for Transport Studies, UCL*.

Mingardo, G. (2019). Slow Traffic Urban Retail – A comparative analysis of 8 EU cities. *Walk21 Conference.* Rotterdam.

Nieuwenhuijsen, M., & Khreis, H. (2016). Car free cities: Pathway to healthy urban living. *Environment International*, 251-262.

Nguyen, D. (2010). Evidence of the Impacts of Urban Sprawl on Social Capital. *Environment and Planning B: Urban Analytics and City Science*, 610-627.

PDOK (2022). *Dataset BGT:* From PDOK: https://app.pdok.nl/lv/bgt/download-viewer/

Rogers, S., Gardner, K., & Carlson, C. (2013). Social Capital and Walkability as Social Aspects of Sustainability. *Sustainability*, 3473-3483.

Rietveld, P. (2000). The accessibility of railway stations: the role of the bicycle in The Netherlands. *Transportation Research Part D: Transport and Environment*, 71-75.

Saelens, B., Sallis, J., Black, J., & Chen, D. (2013). Neighborhood-Based Differences in Physical Activity: An Environment Scale Evaluation. *American Journal of Public Health*, 1552-1558.

Santos, G., Behrendt, H., Maconi, L., Shirvani, T., & Teytelboym, A. (2010). Part I: Externalities and economic policies in road transport. *Research in Transportation Economics*, 2-45.

Song, Y., Merlin, L., & Rodriguez, D. (2013). Comparing measures of urban land use mix. *Computers, Environment and Urban Systems*, 1-13.

Speck, J. (2012). *Walkable City: How Downtown Can Save America, One Step at a Time.* New York: North Point Press.

Sugiyama, T., Francis, J., Middleton, N., Owen, N., & Giles-Corti, B. (2010). Associations between recreational walking and attractiveness, size, and proximity of neighborhood open spaces. *American Journal of Public Health*, 1752-1757.

Tinessa, F., Pagliara, F., Biggiero, L., & Delli Veneri, G. (2021). Walkability, accessibility to metro stations and retail location choice: Some evidence from the case study of Naples. *Research in Transportation Business & Management*.

Tu, C., & Eppli, M. (2003). Valuing New Urbanism: The Case of Kentlands. *Real Estate Economics*, 425-451.

Wagtendonk, A., & Lakerveld, J. (2019). *Walkability score Netherlands version 1.0.* Amsterdam: Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam UMC.

WHO. (2020). *Supporting healthy urban transport and mobility in the context of COVID-19.* Geneva.

WHO. (2022, March 30). *Environment and Health*. From euro.who.int: https://www.euro.who.int/en/health-topics/environment-and-health/Transport-and-health/data-and-statistics/physical-activity2

Yaffe, K., Barnes, D., & Nevitt, M. (2001). A Prospective Study of Physical Activity and Cognitive Decline in Elderly Women. *Archives of Internal Medicine*, 1703-1708.

Young, R. (1996). *Measuring Economic Benefits for Water Investments and Policies.* World Bank Books.

Appendix

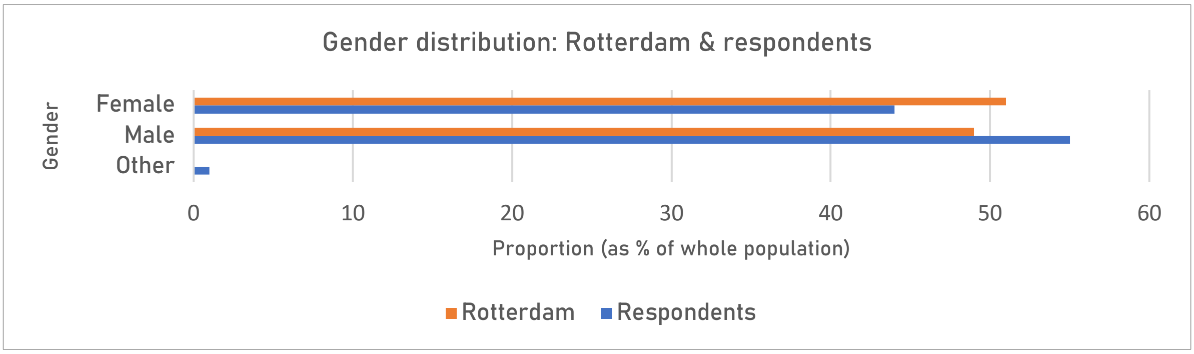

Figure A1: Gender distribution – comparing the respondent sample with Rotterdam's overall population. Source: Gemeente Rotterdam (2020) & CBS (2022)

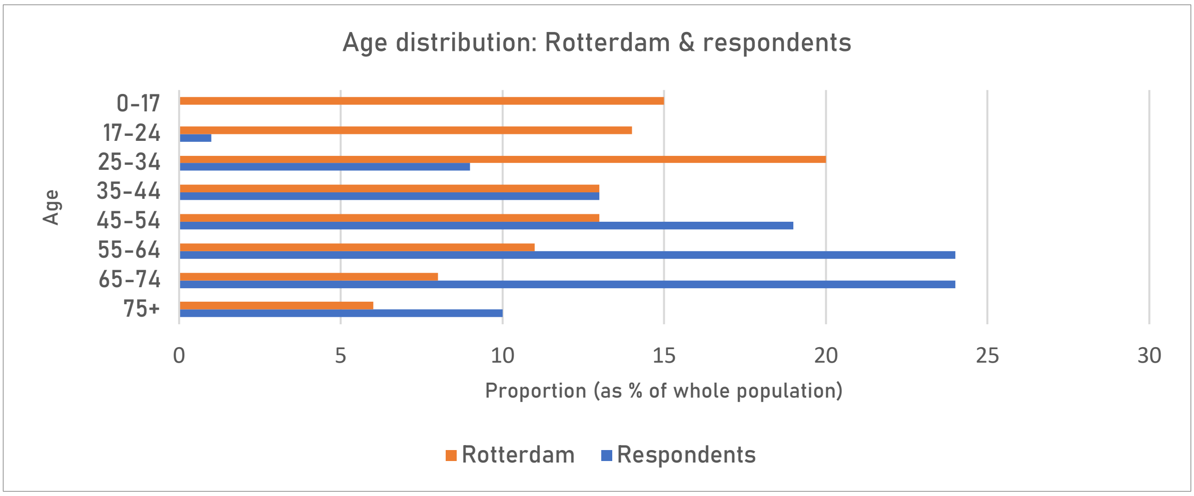

Figure A2: Age distribution – comparing the respondent sample with Rotterdam's overall population Source: Gemeente Rotterdam (2020) & CBS (2022)

Figure A3: Map with selected physical environment features in 3076J (PC5 area serving as index base for empirical analysis). Window 1 highlights the terraced housing, window 2 highlights the presence of rail and arterial road infrastructure, window 3 highlights the presence of a green/blue lung.

	PWS	Distance to train rails (m)	Distance to arterial road (m)	Distance to other rails (m)	Distance to sea (m)	Distance to watercourse (m)	Distance to lake (m)	Density of green (m^2/m^2)

PWS	1							
Distance to train rails (m)	0.2254	1						
Distance to arterial road (m)	0.0254	0.1243	1					
Distance to other rails (m)	0.1094	0.0804	-0.2211	1				
Distance to sea (m)	-0.1309	0.0902	0.1333	-0.5178	1			
Distance to watercourse (m)	-0.3587	0.0509	0.102	-0.1641	0.111	1		
Distance to inland water area (m)	-0.0538	0.1203	-0.0127	0.4708	-0.2795	0	1	
Density of green (m^2/m^2)	0.376	0.1888	-0.4322	0.3393	-0.4063	-0.174	0.0573	1
-1872 (%)	0.0583	0.0586	0.1195	-0.0696	0.0648	-0.0029	-0.1598	-0.0477
1872-1940 (%)	-0.0941	-0.1499	-0.0099	-0.2566	0.1132	0.1417	-0.1013	-0.2449
1940-1984 (%)	0.0257	0.0983	0.0908	0.2873	-0.0865	-0.1049	0.0875	0.1856
Average building height (m)	-0.0969	-0.2391	0.3165	-0.1953	0.1312	-0.0333	-0.0126	-0.4322
Address density (#/km2)	-0.2611	-0.2784	0.5271	-0.5275	0.3623	0.219	-0.1441	-0.6355
Rank-based amenity index (0 to 1)	-0.2183	-0.2489	0.5853	-0.4645	0.2438	0.2283	-0.2744	-0.5856
Entropy index (0 to 1)	-0.1253	-0.0992	-0.0517	-0.1234	0.2125	-0.0691	-0.0762	-0.132
Men (%)	0.0709	0.0225	0.0229	-0.0236	0.0065	-0.0244	-0.0826	-0.0458
No migrant background (%)	0.314	0.267	-0.1851	0.3711	-0.4373	-0.202	-0.0213	0.5008
Households with 1+ children (%)	0.122	0.0599	-0.3038	0.1873	-0.1147	-0.0704	0.0842	0.3097

Figure A4/a: Correlation matrix of the multiple linear regression of physical environment variables on PC5 perceived walkability score (base model) (Part 1/2)

	Average building height (m)	C: -1872 (%)	C: 1872-1940 (%)	C: 1940-1984 (%)	Address density (#/km2)	Rank-based amenity index (0 to 1)	Entropy index (0 to 1)	Men (%)	No migrant background (%)	Households with 1+ children (%)

Average building height (m)	1									
Construction: -1872 (%)	0.0295	1								
Construction: 1872-1940 (%)	-0.045	-0.6605	1							
Construction: 1940-1984 (%)	0.0471	-0.1023	-0.0529	1						
Address density (#/km2)	0.0281	0.3746	-0.1695	0.3493	1					
Rank-based amenity index (0 to 1)	0.124	0.344	-0.1313	0.3199	0.8839	1				
Entropy index (0 to 1)	0.0999	-0.0571	0.0438	0.1608	0.0249	0.0636	1			
Men (%)	-0.0643	0.1093	-0.0902	0.0661	0.0242	0.0538	0.0743	1		
No migrant background (%)	0.0353	-0.111	0.1412	-0.3133	-0.5205	-0.3711	-0.1497	-0.1305	1	
Households with 1+ children (%)	-0.1329	-0.0658	-0.075	-0.372	-0.3681	-0.4261	-0.1531	-0.2002	0.062	1

Figure A4/b: Correlation matrix of the multiple linear regression of physical environment variables on PC5 perceived walkability score (base model) (Part 2/2)

Variable	**VIF**	**1/VIF**
Address density (#/km2)	7.82	0.127936
Amenity index	7.14	0.140051
Year of construction: 1872-1940 (%)	2.53	0.395018
Distance to other rail (m)	2.29	0.436092
Density of green areas (m2/m2)	2.22	0.450636
Year of construction: 1940-1984 (%)	2.19	0.456694
Distance to arterial road (m)	2.12	0.471391
No migrant background (%)	2.12	0.472679
Distance to sea (m)	1.99	0.502611
Distance to inland water area (m)	1.8	0.555783
Average building height (m)	1.69	0.59256
Distance to train rails (m)	1.57	0.638858
Households with 1+ children (%)	1.56	0.642387
Entropy index (0 to 1)	1.19	0.841734
Men (%)	1.17	0.854241
Distance to watercourse (m)	1.15	0.869023
Year of construction: Before 1872 (%)	1.11	0.899787

Mean VIF	2.45

Figure A5: Variance inflation factors (VIFs) of the physical environment and control variables (base model).

	Traffic index	Water index	Green index	Compact index	Amenity index	Land-use index	Men (%)	No migr. backgr. (%)	H. w/ 1+ children (%)

Perceived walkability scores (PWS)	1								
Traffic environment index (tindex)	0.321	1							
Water environment index (windex)	-0.2888	-0.0524	1						
Green environment index (gindex)	0.376	0.2105	-0.2033	1					
Compactness environment index (cindex)	-0.1747	-0.2368	0.0533	-0.5975	1				
Amenity environment index (amindex)	-0.2304	-0.1238	0.1658	-0.5317	0.5633	1			
Land-use meta-index (lindex)	-0.1253	-0.1674	0.0161	-0.132	0.1213	0.0791	1		
Men (%, indexed)	0.0709	0.072	-0.0274	-0.0458	0.095	0.0966	0.0743	1	
No migrant background (%, indexed)	0.314	0.2467	-0.2031	0.5008	-0.4227	-0.393	-0.1497	-0.1305	1
Households with 1+ children (%, indexed)	0.122	0.0204	0.0103	0.3097	-0.4229	-0.4217	-0.1531	-0.2002	0.062

Figure A6: Correlation matrix of the multiple linear regression of physical environment indexes and indexed control variables on PC5 perceived walkability score (index model, PC5/3076J = 100)

Variable	VIF	1/VIF
Compactness environment index (cindex)	2.01	0.498302
Green environment index (gindex)	1.93	0.517554
Amenity environment index (amindex)	1.77	0.56434
No migrant background (%)	1.57	0.638193
Households with 1+ children (%)	1.45	0.691536
Traffic environment index (tindex)	1.13	0.882968
Water environment index (windex)	1.09	0.920561
Men (%)	1.08	0.924555
Land use diversity index (lindex)	1.07	0.933885

Mean VIF	1.46

Figure A7: Variance inflation factors (VIFs) of the physical environment indexes (index model)

Assumption: Normal error terms Variable: Fitted values of PWS Ho: Constant variance

Chi2(1) = 0.43
Prob > Chi2 = 0.5128

Figure A8: Breusch-Pagan/Cook-Weisberg test for heteroskedasticity (base model)

Assumption: Normal error terms Variable: Fitted values of PWS Ho: Constant variance
Chi2(1) = 0.53
Prob > Chi2 = 0.4651

Figure A9: Breusch-Pagan/Cook-Weisberg test for heteroskedasticity (index model)

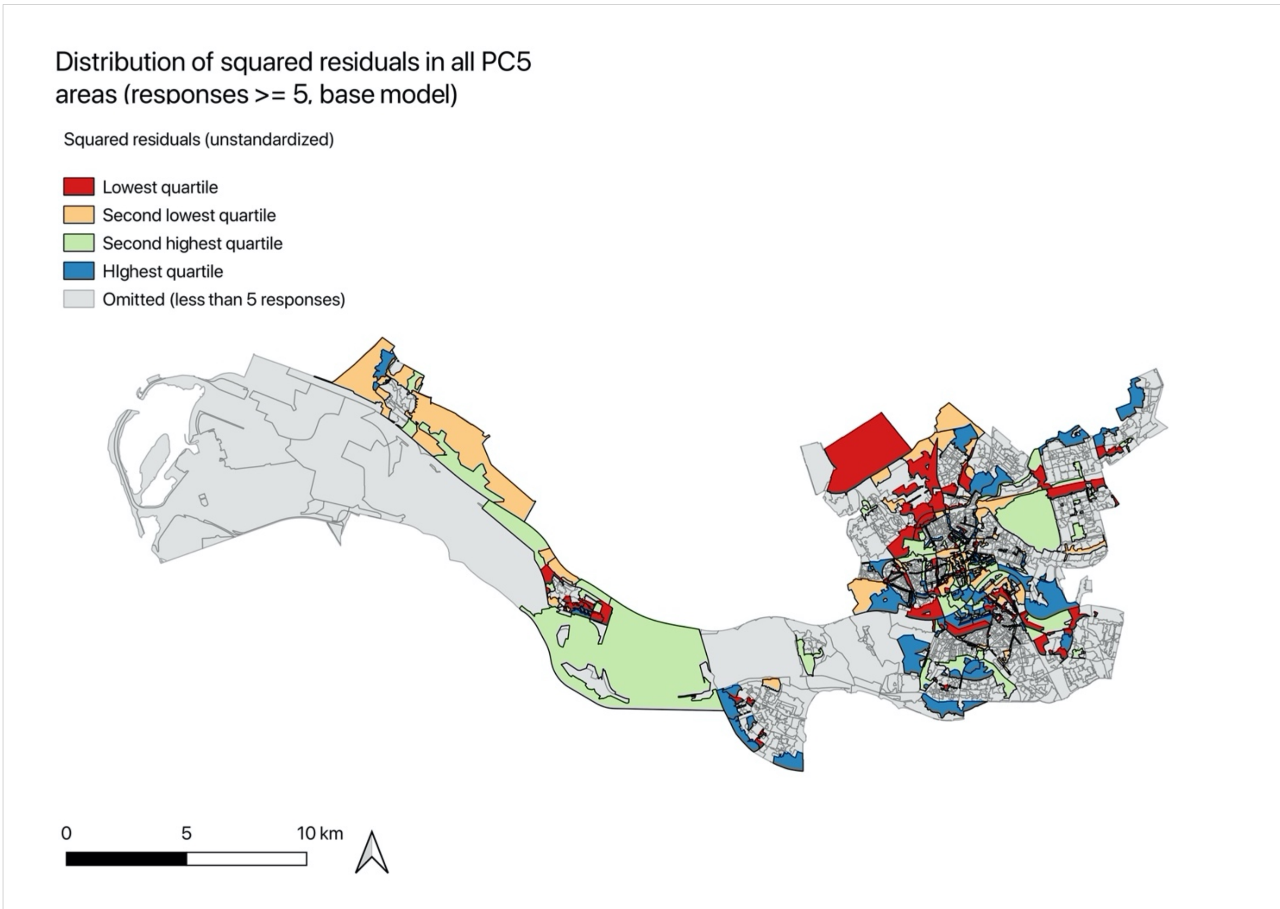

Figure A10: Spatial distribution of squared residuals in all PC5 areas (responses >=5, base model)

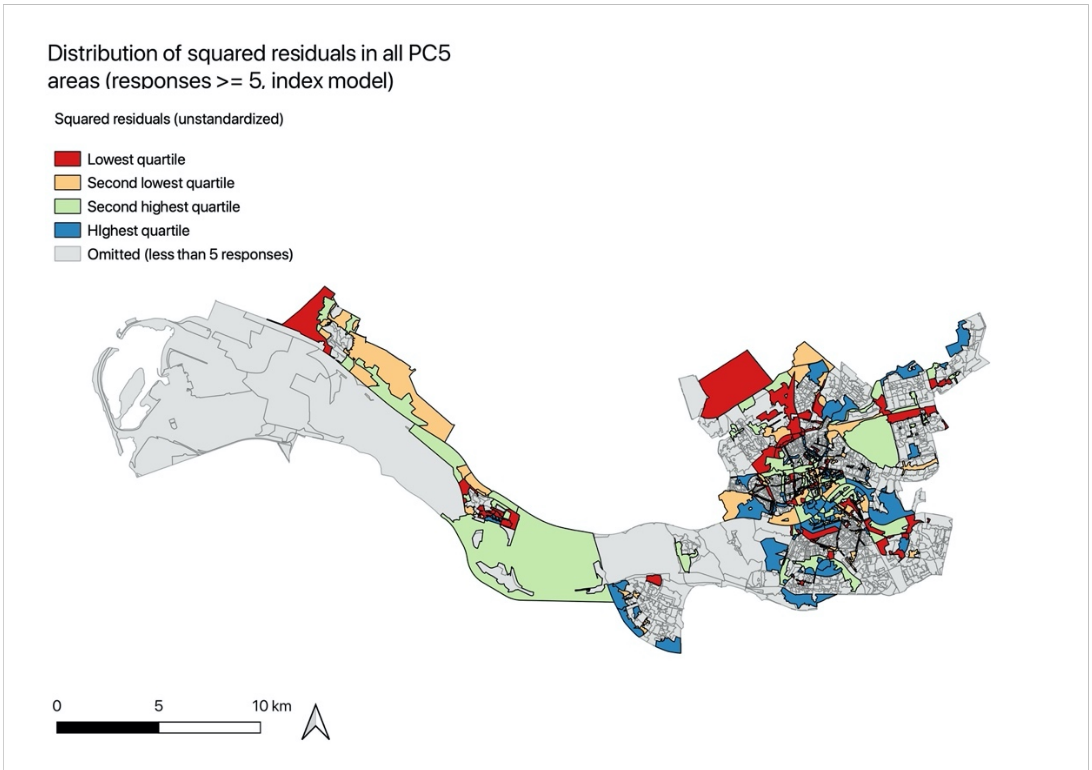

Figure A11: Spatial distribution of squared residuals in all PC5 areas (responses >=5, index model)

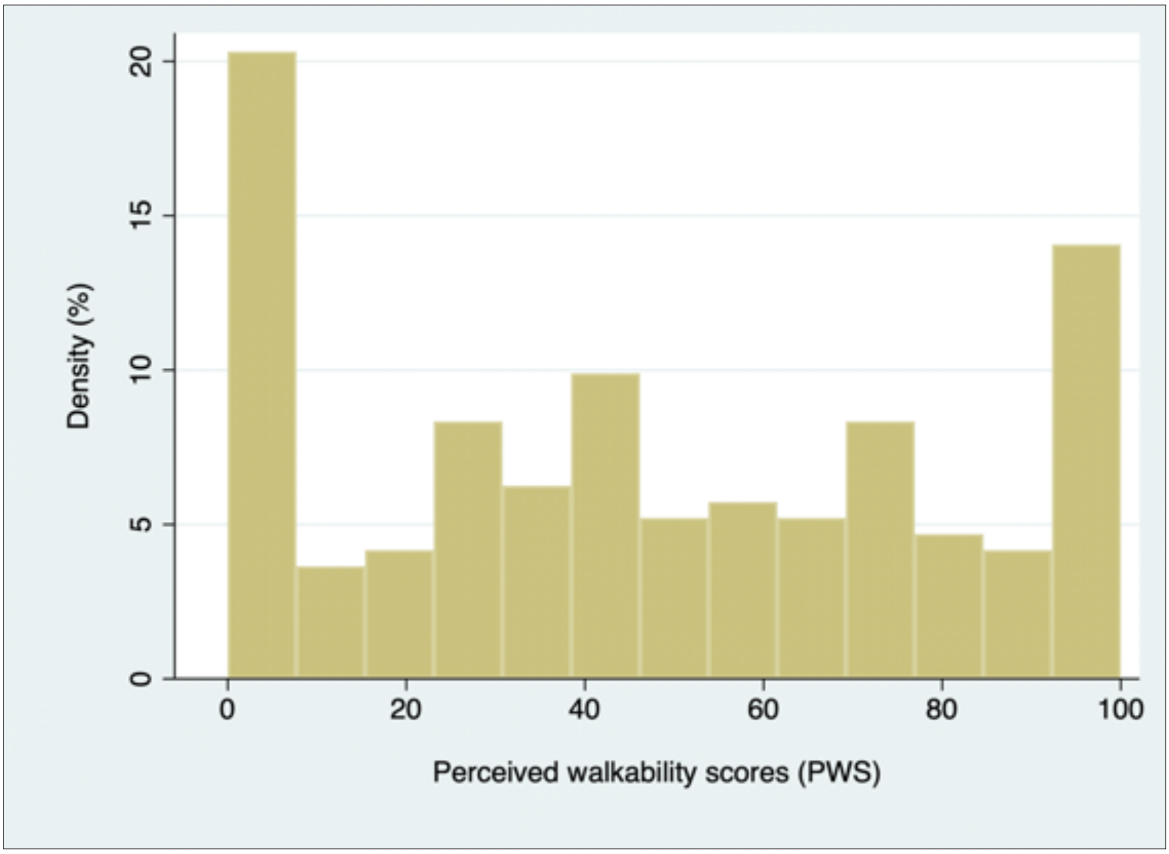

Figure A12: Histogram / Density of perceived walkability scores (%)

Variable	Obs	Mean	Std. Deviation	Minimum	Maximum
Perceived walkability scores (1 to 100)	192	47	34	**0**	100
Distance to train rails (m)	192	746.06	618.60	11.43	3776.68
Distance to arterial road (m)	192	1087.57	634.55	7.22	2424.21
Distance to other rails (m)	192	587.70	1009.06	0.20	6552.06
Distance to sea (m)	192	24893.84	5211.43	842.39	31963.08
Distance to watercourse (m)	192	143.57	142.17	0.00	611.27
Distance to lake (m)	192	1118.58	634.05	0.00	2866.11
Density of green (m2/m2)	192	0.51	0.50	0.00	2.07
Average building height (m)	192	0.01	0.04	0.00	0.29
Year of construction (%): Before 1872	192	0.33	0.33	0.00	1.00
Year of construction (%): 1872-1940	192	0.35	0.30	0.00	1.00
Year of construction (%): 1940-1984	192	0.30	0.26	0.00	1.00
Year of construction (%): 1984-2019	192	9.67	7.95	-2.98	62.83
Address density (#/km2)	192	4245.14	2262.32	56.00	8512.00
Rank-based amenity index (average rank)	192	473.13	332.40	6.29	1016.71
Entropy index (0 to 1)	192	0.13	0.04	0.02	0.22
Men (%)	192	0.51	0.06	0.33	1.00
No migrant background (%)	192	0.55	0.20	0.10	0.90
Households with 1+ children (%)	192	0.23	0.13	0.00	0.58

Figure A13: Summary statistics of perceived walkability scores (PWS), physical environment characteristics and control variables (base model)

Variable	Obs	Mean	Std. dev.	Min	Max
Perceived walkability scores (PWS)	192	47	34	**0**	100
Traffic environment index (tindex)	192	96.99	10.79	53.71	121.06
Water environment index (windex)	192	129.74	17.39	63.72	160.72
Green environment index (gindex)	192	41.09	40.55	0.00	167.02
Compactness environment index (cindex)	192	310.33	170.48	7.33	1201.92
Amenity environment index (amindex)	192	1625.09	1647.01	1.14	5806.33
Land-use meta-index (lindex)	192	89.63	25.07	12.12	152.35
Men (%)	192	1.00	0.12	0.65	1.94
No migrant background (%)	192	0.91	0.33	0.17	1.50
Households with 1+ children (%)	192	0.45	0.27	0.00	1.17

Figure A14: Summary statistics of perceived walkability scores (PWS), physical environment indexes and control variables (index model)